

Running Head: FIDELITY TO STRUCTURE

Considering the Role of Fidelity of Implementation in Science Education Research:

Fidelity as Teacher and Student Adherence to Structure

Joelle R. Lastica & Carol L. O'Donnell

The George Washington University

Paper presented at the Annual Meeting of the

American Educational Research Association

Chicago, IL

April 10-14, 2007

This work was based on the work conducted by SCALE-uP: A collaboration between George Washington University and Montgomery County Public Schools (MD); Sharon Lynch, Joel Kuipers, Curtis Pyke, and Michael Szesze serve as principal investigators of SCALE-uP.

Funding for SCALE-uP was provided by the National Science Foundation, the U.S. Department of Education, and the National Institute of Health (REC-0228447). Any opinions, findings, conclusions, or recommendations are those of the author and do not necessarily reflect the position or policy of endorsement of the funding agencies.

Correspondence to: Joelle Lastica; E-mail: jlastica@gwu.edu; Work Phone: 202-994-1176 and FAX: 202-994-0692

Considering the Role of Fidelity of Implementation in Science Education Research:

Fidelity as Teacher and Student Adherence to Structure

Background

Fidelity of implementation has been defined as the extent to which the delivery of an intervention adheres to the program model originally developed, and confirms that the implementation of the independent variable in outcome research occurred as planned (Mowbray, Holter, Teague & Bybee, 2003). Fidelity of implementation is a relatively new construct in education research; however, its use in program evaluation dates back 25-30 years (Sechrest, West, Phillips, Redner, & Yeaton, 1979, as cited by Mowbray, Holter, Teague, & Bybee, 2003), most of which comes from mental health (McGrew, Bond, Dietzen & Salyers, 1994; Teague, Bond, & Drake, 1998) and public health intervention research (Dumas et al, 2001; Dusenbury et al, 2003; Emshoff, Blakely, Gottshalk, Mayer, Davidson, & Erickson, 1987; Vincent et al, 2001).

The current emphasis in education on “evidence-based” research has created renewed interest in the importance of *fidelity of implementation*. In their 2006 Request for Applications, the Institute of Education Sciences (IES)—the research arm of the Department of Education established by the Education Sciences Reform Act of 2002—required intervention researchers to describe how treatment fidelity will be measured, how often it will be assessed, and the degree of acceptable variation during an intervention study (U. S. Department of Education, 2006).

The applicant should specify how the implementation of the intervention will be documented and measured. In strong applications, investigators will make clear how the fidelity measures capture the critical features of the intervention. The proposal should either indicate how the intervention will be maintained consistently across multiple groups (e.g., classrooms and schools) over time or describe the parameters under which

variations in the implementation may occur. Investigators should propose research designs that permit the identification and assessment of factors impacting the fidelity of implementation (p. 73).

Including fidelity measures in educational research is critical not only to determine if the intervention is implemented (as a validity check), but, ultimately, to determine how fidelity relates to student outcomes. Without reliable and valid fidelity measures during implementations of classroom interventions, the outcomes could be mistakenly attributed to the treatment or the influence of unknown contaminants to the treatment (Moncher & Prinz, 1991) and researchers may have insufficient evidence to support internal validity (Dumas, Lynch, Laughlin, Smith, & Prinz, 2001) and construct validity (Reichardt, 2006).

Fidelity of Implementation as a Means of Controlling Threats to Validity

Studying fidelity of implementation to educational interventions is important for a variety of reasons. As decisions regarding curricula are being made by school, district, or local education agency personnel on a regular basis, fidelity of implementation studies can help these decision-makers by explaining why innovations succeed and fail, helping them to understand how curricular changes can impact student outcomes, as well as providing information on how likely an intervention can and will be implemented with fidelity (Dusenbury et al, 2003). Songer and Gotwals (2005) also emphasize the importance of fidelity of implementation research in the classroom when they relate, “how much and in what way teachers implement curricular units plays a large role in the learning that takes place.” In addition, studies of fidelity of implementation can help to establish validity in research designs.

In any research study, prior to discussing results of a causal relationship between *Treatment A* and *Outcome B*, it is critical to first establish that *Treatment A* was indeed

implemented as the researchers expected. Educational researchers refer to fidelity of implementation as one way to establish *internal validity* (Dumas, Lynch, Laughlin, Smith, & Prinz, 2001; Moncher & Prinz, 1991). Cook & Campbell (1979), for example, defined internal validity as “the approximate validity with which we infer that a relationship between two variables is causal or that the absence of a relationship implies the absence of cause” (p. 37). Mowbray et al. (2003) cited Chen, who defined internal validity in saying “without documentation and/or measurement of a program’s adherence to an intended model, there is no way to determine whether unsuccessful outcomes reflect a failure of the model or failure to implement the model as intended.” Teague, Bond, and Drake (1998) conveyed a similar meaning of internal validity when they stated that without providing detailed descriptions of interventions, conclusions about the presence or absence of effects would be questionable. Other researchers (DiLoreto, 1971; Heimberg and Becker, 1984; Moncher & Prinz, 1991) referred to the tendency of non-adherence to protocols to increase as treatment progresses as a threat to internal validity. However, Moncher and Prinz added that treatment fidelity is also related to construct fidelity.

Some researchers refer to the measure of fidelity of implementation as a means of establishing *construct validity* (Farrington, 2003; Moncher & Prinz, 1991; Reichardt, 2006). Moncher and Prinz (1991), for example, citing Heimberg and Becker’s work, said that by demonstrating fidelity, one establishes the construct validity of a treatment, an essential step in outcome research. Reichardt (2006) described a threat to construct validity as “if a program were judged to be ineffective when in fact it had not been implemented with reasonable strength and fidelity” (p. 82). Farrington (2003) defined construct validity as “the adequacy of the operational definition and measurement of the theoretical constructs that underlie the intervention and the outcome” (p. 54). He also included as a threat to construct validity “the extent to which the

intervention succeeded in changing what it was intended to change (p. 54),” giving as an example, the degree of treatment fidelity or implementation failure. Cherryholmes (1998) purported that construct validity, which, he explains is involved whenever a test or research measurement “is to be interpreted as a measure of some attribute or quality which is not ‘operationally defined,’” can be argued to be prior in importance to content, internal, statistical conclusion and external validity.

Answering the question of whether the treatment was actually delivered, and if so, to what degree it was delivered, becomes crucial in any experimental design. Establishing fidelity of implementation to the treatment supports both the internal and construct validity of an efficacy or effectiveness study because it provides evidence that the intervention was unequivocally implemented in the treatment classrooms.

Context

This paper explores fidelity of implementation within the context of a large-scale quasi-experiment *Scaling up Curriculum for Achievement, Learning, and Equity Project* (SCALE-uP), a five-year, \$5.7 million grant, funded by the Interagency Education Research Initiative (IERI) and administered by the National Science Foundation (NSF). SCALE-uP partners researchers from The George Washington University (GWU) Graduate School of Education and Human Development (Department of Teacher Preparation and Special Education) and the Columbian College of Arts of Sciences (Department of Anthropology) with Montgomery County Public Schools (MCPS).

Building on the work of Mowbray et al. (2003), SCALE-up divides its fidelity of implementation study into two components: *Fidelity to Process* and *Fidelity to Structure*. Fidelity to Process examines the way in which services are delivered and fidelity to structure

encompasses the framework for service delivery (Mowbray et al., 2003). This paper reports on SCALE-uP's study of Fidelity as Adherence to Structure, defined by SCALE-uP as the extent to which teachers implement lesson and unit components as designed and students complete written and procedural work as outlined by the program developer. Although Mowbray and colleagues (2003) refer to this as fidelity to structure, others refer to it as fidelity to the "surface" of the actual program materials (Gersten et al., 2005).

The purpose of SCALE-uP is to study the "scale-up"—or the transition of curriculum materials from idiosyncratic adoption to broad, effective implementation across a large and diverse school system. The interventions of SCALE-uP include three middle school science curriculum units rated excellent or satisfactory according to the AAAS Project 2061 Curriculum Analysis (AAAS, 2003; Kesidou & Roseman, 2002). These units include *Chemistry That Applies* (Michigan Department of Education, University of Michigan, 1993); *GEMS: The Real Reasons for Seasons: Sun-Earth Connections*, (Lawrence Hall of Science [LHS], University of California at Berkeley, 2000); and, *ARIES: Exploring Motion and Forces: Speed, Acceleration, and Friction* (Harvard-Smithsonian Center for Astrophysics, 2001).

During SCALE-uP's fourth year quasi-experiment (Year 4: 2005-2006) the study had grown to involve 35 middle schools and 85,000 students. In addition to measures of effectiveness, SCALE-uP also focused upon the implementation effects of *fidelity*, *scale*, and school science departments' increasing *experience* with highly rated science curriculum units. Upon the grant's completion in 2007, the No Child Left Behind mandates for accountability require assessing K-12 students in science. By that time, over 200,000 students and 250 teachers will have been involved in this study. SCALE-uP researchers hope to be able to report on the effectiveness of the three units as they go to scale and hope to identify and explain factors that

affect outcomes for diverse groups of students. Ultimately, the research intends to identify the conditions under which effective evidence-based interventions improve student learning and achievement when applied on a large scale. As such, before sharing results from SCALE-uP's Year 4 investigation of its Treatment unit *ARIES: Exploring Motion and Forces: Speed, Acceleration, and Friction* with a sample of sixth grade students in MCPS, a diverse, Washington, D.C. metropolitan area school district, it became critical to address this question of whether the treatment was actually implemented, and, if so, to what degree it was implemented in the SCALE-uP Treatment classrooms.

Research Questions

Given the conceptual framework developed by SCALE-uP to study fidelity of implementation and reported in Lynch (2007), the collection and analysis of the Year 4 Fidelity as Adherence to Structure data were guided by the following research questions:

1. To what extent did teachers and students adhere to the *ARIES: Motion and Forces* Treatment unit?
2. Is there a relationship between Fidelity as Teacher Adherence to Structure and classroom mean achievement in the *ARIES: Motion and Forces* classrooms?
3. Is there a relationship between Fidelity as Student Adherence to Structure and classroom mean achievement in the *ARIES: Motion and Forces* classrooms?
4. If statistically significant relationships exist, how much of the variation in classroom mean achievement is predicted by each Fidelity as Adherence to Structure variable?

Research Design

Population

The population under investigation was middle school students from large, diverse school

systems. The study population was sixth grade students in MCPS in the 2005-2006 school year. MCPS is a large Maryland school district (approximately 136,000 students total, 29,000 in grades 6-8 in 2005) located in the Washington, DC, metropolitan area. The study population has no ethnic majority and is among the top-performing in the state of Maryland. MCPS is an ideal laboratory for this study, because while it is recognized as one of the best large school districts in the country, it has also become one of the most diverse. Comprised in 2006 of almost 140,000 students representing 163 countries and speaking 123 languages, MCPS faces the challenges brought by an increasingly diverse student population, including poverty, limited English proficiency, and troubling achievement disparities among groups of students (MCPS, 2006).

Study Sample

The unit of analysis for this fidelity of implementation was the classroom. The study sample consisted of 30 Treatment classrooms randomly selected from the total number of 44 Treatment classrooms in the 5 MCPS schools in the study population. These classrooms did not include “self-contained” SPED and ESOL classrooms because there was no guarantee that the teachers had attended the professional development workshops associated with the intervention unit and could deliver the intervention with reasonable fidelity. The sample classrooms were picked randomly using SPSS version 12. Six classrooms from each Treatment school were selected from each School Profile Category (a process used to stratify the 5 schools by socio-economic status and other demographic characteristics) in order to obtain teacher and student variability. (See the Implementation study of *Exploring Motion and Forces* [2004-2005]: SCALE-Up Report No. 8 for more information on the selection of School Profile Categories.)

Curriculum Unit

ARIES: Exploring Motion and Forces: Speed, Acceleration, and Friction, hereafter referred to as *ARIES: M&F*, was developed for the ARIES curriculum program by the Science Education Department at the Harvard-Smithsonian Center for Astrophysics. It is a six-week physical science curriculum unit designed for grades 5-8. The curriculum materials are inquiry-centered and activity-based, with an emphasis on students' direct experience with phenomena. A substantial amount of materials is required to implement *ARIES: M&F*. These materials (sliding disks, ramps, marbles, bells, rolling carts, etc.) are often constructed and then used by the students themselves as they explore physical phenomena. MCPS provided these curriculum materials to all of the teachers selected for the Treatment condition.

Instrumentation and data collection

SCALE-uP researchers used a combination of classroom observations, teacher interviews, and student journals to ascertain how the treatment unit—*ARIES: M&F*—was being implemented in the 30 classrooms in the 5 Treatment schools.

ACOP. The first phase of measuring a teacher's Fidelity as Adherence to Structure involved a series of classroom observations using the Adherence Classroom Observation Protocol (ACOP), which was used to observe each teacher's adherence to the following six structural components of an *ARIES: M&F* lesson: *Materials* (1 item), *Recording Ideas* (2 items), *Procedures in the Student Journal* (4 items), *Interpreting Ideas* (2 items), *Procedures in the Teacher's Guide* (1 item), and *Sequencing* (1 item) as outlined in the *ARIES: M&F* student *Science Journal* and *Teacher Guide*. In *ARIES: M&F*, each of the 18 connected, but separate, lessons was centered on a laboratory activity. The student *Science Journal* activates students' prior knowledge before the activity begins, guides students during the activity, scaffolds their data collection, and provides queries to assist students' reasoning from lab-based evidence. The

Teacher Guide provides rationale, background content information, and suggestions for how to conduct each lesson. Trained observers viewed one *ARIES: M&F* lesson in its entirety to describe the extent to which these lesson components were implemented during a lesson. Each ACOP item was scored on a four-point Likert-like scale, with 3 representing close adherence to the written lesson and 0 indicating no adherence or that the teacher did not conduct this component of the lesson. Because each of the 11 ACOP items represented a distinct lesson component, ACOP item scores were not combined or aggregated, but kept separate.

To develop the observation protocol, SCALE-uP researchers met over a course of approximately 14 weeks to discuss and provide feedback about face validity, content (did the protocol reflect the items in the Student *Science Journal* and *Teacher Guide*), and language. In January 2006, SCALE-uP trained five observers obtained from MCPS' Department of Shared Accountability (DSA) in the use of the ACOP by using video data enactments of *ARIES: M&F* recorded from a former treatment classroom. The inter-rater reliability was conducted 10 times. Syntax was written using SPSS version 12 in order to conduct the analysis (If rater 1 = rater 2, then $IRR = 1$. If rater 1 \neq rater 2, then $IRR = 0$). The inter-rater reliability scores ranged from .45 to .90, with an average score of .65. The IRR was conducted again after removing two items that were confusing to the raters, after which, the IRR scores ranged from .44 to 1, with an average score of .75. The pairs with low inter-rater reliability of the scale were retrained. Later that month, the five raters independently observed the same classroom using another *ARIES: M&F* video enactment after reconciliation and retraining. The inter-rater reliability was again conducted ten times, resulting in a range of inter-rater reliability scores of .63 to 1.0, with an average score of .82. The scores were reconciled again, and the results improved tremendously after re-training.

SCALE-uP trained observers were sent into each classroom selected for observation. ACOP data were collected from 30 classrooms of 11 teachers in 5 schools. These observations were coordinated by MCPS DSA personnel, who, in January 2006, made initial contact with the classroom teachers randomly selected for observation regarding their teaching schedules for the *ARIES: M&F* unit. Once communication was established with classroom teachers, DSA staff coordinated with the SCALE-uP research team to ensure observers would be available to collect classroom observation data for these scheduled observations. Therefore, lesson observations were not conducted at random, but were constrained by schedules of the observed classroom teachers and the availability of DSA and SCALE-uP staff. Observers were required to capture an entire lesson, and some lessons extended beyond one class period on a given day. Observations for each class period lasted 45-100 minutes, depending on the schedule of “regular” versus block class periods. Classroom observations began in the beginning of February 2006 and were completed at the end of March 2006. Perhaps due to the scheduling limitations, out of 17 observable *ARIES: M&F* lessons (or explorations, as they are referred to in the *ARIES: M&F* curriculum unit), all 30 classroom observations were of explorations 2-11, 60% of them being of explorations 2-5. Table 1 shows how the observations were spread out over the two months.

Table 1

Schedule of Completed ARIES: M&F Classroom Observations

Exploration number	Teacher A	Teacher B	Teacher C	Teacher D	Teacher E	Teacher F	Teacher G	Teacher H	Teacher I	Teacher J	Teacher K
2	Feb 7	Feb 7	Feb 9	Feb 6			Mar 22		Mar 23		
3			Feb 21	Feb 7	Feb 10	Feb 15		Mar 24	Mar 24		
4			Feb 23				Mar 27	Mar 28			
5	Feb 14	Feb 14								Feb 6	
6				Feb 21						Feb 7	Feb 7
7											Feb 10
8						Mar 22				Feb 17	Feb 16
9						Mar 30					
10					Mar 24						
11	Mar 3	Mar 3			Mar 24						
12											
13											
14											
15											
16											
17											
18											

AIP. Because the ACOP only measures adherence to structure to one lesson, it was important to also capture each teacher's fidelity to the entire unit. To do this, the second phase for understanding teachers' Fidelity as Adherence to Structure involved interviewing the Treatment teachers about their adherence to the entire *ARIES: M&F* curriculum unit.

The Adherence Interview Protocol (AIP) was used to guide our understanding of the extent to which teachers reported they adhered to the following six structural components of the *ARIES: M&F* unit: *Materials* (1 item), *Recording Ideas* (2 items), *Procedures in the Student Journal* (3 items), *Interpreting Ideas* (2 items), *Procedures in the Teacher's Guide* (1 item), and *Sequencing* (4 items), over the course of the entire unit. The AIP was aligned with 8 of the 11 ACOP items. Each of these was recorded on a four point Likert-like scale, with 3 representing teachers' self-reports of full adherence and 0 representing no adherence. Interviews also included demographic and background data to determine the Treatment teachers' experience, education, formal science knowledge, and experience teaching motion and forces, which are summarized in Table 2.

Table 2

Teachers' Educational Background Data

Teacher (ID #)	Teacher certification?	Science degree?	Total # science credits taken in school?	Total # physics or physical science credits taken in school?	Training or CEU's in physics or physical science?	Total # CE credit hours for physics or physical science?	Years teaching secondary science	Years teaching Motion and forces topics?	Teacher highest degree?	Has teacher worked in the science profession (outside of teaching?)
1 (1001)	Sec. Ed	Yes	62	9	No	0	17	5	MS	Yes
2 (1002)	Elem. Ed	No	15	3	No	0	10	10	BA	No
3 (2003)	Sec. Ed	No	90	15	No	0	7	3	BA	Yes
4 (2004)	Sec. Ed	No	36	6	No	0	16	1	BA	No
5 (3001)	Elem. Ed	No	24	3	No	6	9	4	BA	No
6 (3002)	Elem. Ed	No	18	6	No	0	13	13	MA	No
7 (4001)	Elem. Ed	No	9	3	No	0	14	11	MA	No
8 (4002)	Elem. Ed	No	20	0	Yes	0	2	2	MA	No
9 (4003)	Sec. Ed	No	12	3	No	0	1	1	BA	No
10 (5001)	Elem. Ed	No	39	9	No	0	11	10	MA	No
11 (5002)	Elem. Ed	No	34	3	No	0	10	3	MA	No

IRB approval for teacher interviews was obtained in December 2005. During May 2006, SCALE-uP research staff contacted Year 4 Treatment teachers to schedule interviews to be conducted after the completion of the *ARIES: M&F* unit. Scheduling was done primarily through emails with follow up phone calls as needed. A SCALE-uP research assistant went to the teachers' schools for the interview. At the start of the interview, teachers were given copies of the interview protocol, an information sheet, and an informed consent sheet to review. Teacher participants were told the interview would take about thirty minutes and issues surrounding confidentiality were discussed. During the interview, the protocol was read to the teacher and the

interviewer took descriptive notes. Although it was not requested, some teachers provided photocopies of extra curriculum materials (i.e., handouts, planning notes, etc.) they had used to teach the *ARIES: M&F* unit.

Student Science Journal. Finally, to assess Fidelity as Student Adherence to Structure, for the first time during our grant, SCALE-uP collected completed *ARIES: M&F* student *Science Journals* from all observed classrooms at the end of the school year. Using the work of Songer and Gotwals (2005) as a foundation, we sought to measure students' adherence to the *ARIES: M&F* unit. Songer and Gotwals had used the number of worksheets that students had completed as an indicator of levels of fidelity to their science curriculum materials, and found that this simple measure predicted student outcomes well in a regression analysis. These consumable 115-page student *Science Journals* contain 415 open-ended items in which students record their prior ideas, observations, and interpretations/conclusions. Trained raters recorded the number of completed questions in each of the 819 student science journals collected and entered the data by the three lesson components of each of the 18 lessons: *Recording Ideas (RID)*, *Procedures (PROC)*, and *Interpreting Results (INT)*, into an Excel file, which was transferred to SPSS.

Classroom Mean Achievement. Classroom mean achievement (i.e. how well students in the class learned the motion and forces target idea) was assessed using the Motion and Forces Assessment (MFA) (Pyke & Ochsendorf, 2006). The MFA is a curriculum independent assessment composed of ten selected and constructed response items designed for ease of use and diverse classroom settings. It uses language and illustrations that allow it to be read and understood by a maximum number of 6th grade students. SCALE-uP sent teachers a cover letter with directions for administering the MFA. Teachers were instructed to give the MFA to students after they completed the *ARIES: M&F* unit and then return completed assessments to MCPS'

Department of Shared Accountability in an envelope marked with class code. Data from the MFA were collected at the individual student level, but were aggregated to the classroom level for the purposes of this FOI study (see the Implementation study of *Exploring Motion and Forces* [2004-2005]: SCALE-uP Report No. 8 for a complete description of the MFA data collection procedures.).

Results

The results from this Fidelity as Adherence to Structure study are organized by research question. Descriptive results and/or relationships between Fidelity as Adherence to Structure and classroom mean achievement are reported for each question.

Research Question 1: To what extent did teachers and students adhere to the ARIES: Motion and Forces Treatment unit?

Teacher Adherence to Structure. Using the Adherence Classroom Observation Protocol (ACOP), SCALE-uP observed each teacher's (n = 30 classrooms) adherence to 11 different structural components of an *ARIES: M&F* lesson. Descriptive data show that, using a Likert-like scale (3 = fully implemented), teachers adhered most frequently to and differed least in the following lesson components: *Teacher provides opportunities to write responses* (M = 2.90, SD = .31), *...conduct investigation* (M = 2.87, SD = .43), and *...interpret results in writing* (M = 2.87, SD = .35). The lesson component teachers adhered to the least was *Teacher Guide Additions* (M = 0.83, SD = .92). The lesson components to which teachers seemed less likely to adhere and showed greater variability during classroom observations, were *Teacher discusses with students responses to questions* (M = 2.50, SD = 1.08), *...provides opportunity to set up equipment* (M = 2.37, SD = .93, and *Interpreting results verbally* (M = 2.27, SD = 1.08). Complete results are shown in Table 3.

Individual teacher interviews using the AIP revealed that despite some teachers' self-reported resistance to *ARIES: M&F*, teachers adhered to most or all of its lesson components. AIP items teachers reported they adhered to most were *Answer in writing (pre-lab)* ($M = 2.91$, $SD = .30$), *Conduct investigations* ($M = 2.91$, $SD = .30$), and *Answer in writing (post-lab)* ($M = 2.91$, $SD = .30$). All 11 teachers interviewed reported they did not change the unit sequence as outlined in *ARIES: M&F* ($M = 3.00$, $SD = .00$). Results are shown in Table 3.

Table 3

Means, Standard Deviations, Adherence scores, and Correlations with Achievement for the Adherence Classroom Observation Protocol (ACOP) and Adherence Interview Protocol (AIP)

Item	ACOP			AIP	
	<i>M</i> (<i>SD</i>)	Adherence ^a	Correlation with achievement	<i>M</i> (<i>SD</i>)	Adherence ^a
Materials					
Teacher provides students with materials	2.47 (0.51)	100.0	-.25	2.82 (0.41)	100.0
Recording Ideas					
Teacher provides opportunity to write responses	2.90 (0.31)	100.0	-.14	2.91 (0.30)	90.9
Teacher discusses with student responses to questions	2.50 (1.08)	83.3	.21	2.82 (0.41)	81.8
Procedures					
Teacher provides opportunity to set up equipment	2.37 (0.93)	90.0	.07	2.55 (0.52)	54.5
Teacher provides opportunity to record and/or discuss predictions	2.63 (1.06)	87.5	.25	--	--
Teacher provides opportunity to conduct investigation	2.87 (0.43)	96.7	.20	2.91 (0.30)	90.9
Teacher provides opportunity to reflect and write	2.55 (1.10)	85.0	.16	--	--
Interpreting Results					
Interpreting results in writing	2.87 (0.35)	100.0	-.02	2.91 (0.30)	90.9
Interpreting results verbally	2.27 (1.08)	80.0	.61**	2.82 (0.41)	81.8
Teacher Guide Additions	0.83 (0.92)	16.6	.06	--	--
Adhering to Sequence of Activities ^b	2.47 (0.57)	96.7	-.44*	3.00 (0.00)	100.0

* $p < .05$

** $p < .01$

^aFor the ACOP, adherence is calculated as the percent of observations in which teacher adherence to activities was rated mostly implemented (score = 2) or fully implemented (score = 3); for the AIP, adherence is calculated as the percent of teachers' responses in which they provided opportunities for activities most of the time (score = 2) or all of the time (score = 3).

^bIn the ACOP, this item referred to the sequence of activities in the lesson observed; in the AIP, this item referred to the sequence of the explorations in the unit, as a whole.

Student Adherence to Structure. At the end of the school year, SCALE-uP collected 819 completed *ARIES: M&F Student Science Journals* from all 30 Treatment classrooms observed, and trained raters counted the number of completed questions in each Science Journal. We found that students in Treatment classrooms completed, on average, 83.87% of all lesson components. Table 4 shows complete results, including when disaggregated into lesson components.

Table 4

Descriptive statistics and correlations with classroom mean achievement for completed ARIES: M&F Student Science Journals (Spring 2006)

Percent completion of <i>Student Science Journal</i> sections and correlations with classroom mean achievement	Mean (SD)	Min/Max
<i>Recording Ideas</i> $r = .39, p < .05$	87.37 (6.25)	70.19/95.76
<i>Procedures</i> $r = .47, p < .01$	82.93 (9.04)	59.49/94.43
<i>Interpreting Results</i> $r = .34, p > .05$	83.17 (9.44)	56.66/94.61
Total Student Science Journal $r = .48, p < .01$	83.87 (7.59)	67.29/93.98

Research Question 2: Is there a relationship between Fidelity as Teacher Adherence to Structure and classroom mean achievement in the ARIES: Motion and Forces classrooms?

For this research question, we chose to use only instruments with classroom-level data, our unit of analysis, which required that we omit AIP results. On the ACOP, Spearman's rho correlations indicate *Interpreting results and discussion verbally* correlated significantly with classroom mean achievement ($r_s = .61, p < .01$), although teachers were less likely to adhere it ($M = 2.27, SD = 1.08$). *Adhering to sequence of activities* was also shown to negatively correlate with classroom mean achievement ($r_s = -.44, p < .05$). See Table 3 for complete results.

Research Question 3: Is there a relationship between Fidelity as Student Adherence to Structure and classroom mean achievement in the ARIES: Motion and Forces classrooms?

Personal communication with the *ARIES: M&F* developer (B. Ward, personal communication, March 2004) revealed that the completion of the *ARIES: M&F Science Journal* was critical to students' success in the unit. SCALE-uP's conjecture was that a higher frequency of journal completion (which we believed was a measure of high student Fidelity to Structure) would correlate positively with classroom achievement. This seemed to be demonstrated in the data analysis (as shown in Table 4), with all three parts of the lesson, as well as the *Science Journal*, as a whole, showing a statistically significant positive correlation with classroom mean achievement ($r = .48, p < .01$).

Research Question 4: If statistically significant relationships exist, how much of the variation in classroom mean achievement is predicted by each Fidelity as Adherence to Structure variable?

Regression analyses were conducted with both Fidelity as Adherence to Structure variables that correlated significantly with class mean achievement. Controlling for prior science GPA, teacher fidelity to *Interpreting results and discussion verbally* explained 35% of variance in treatment classrooms' mean achievement ($p < .01$). For every one standardized unit that a teacher's adherence to this variable increased, the overall classroom mean on the MFA was predicted to increase 0.593 standardized units, holding prior science GPA constant, $t(27) = 3.815, p < .01$. Similarly, controlling for student science GPA, regression analysis indicated student fidelity to completing the *Science Journal* explained 25% of the variance in treatment classrooms' mean achievement ($p < .01$). For every one standardized unit that students in a classroom completed their *Science Journals*, the overall classroom mean on the Motion and

Forces Assessment was predicted to increase 0.510 standardized units, holding prior science GPA constant, $t(27) = 2.976, p < .01$.

Discussion

Upon analyzing the results from our Fidelity as Adherence to Structure studies, we found that although some more-experienced teachers reported feeling confined by *ARIES: M&F* during their interviews, newer teachers or those without a physical science background welcomed its structure. Results from the ACOP and the AIP showed that the most adhered to lesson components in this unit were those items that required students to write responses. Trained classroom observers and teacher interviews both reported high levels of adherence to structure when teachers provided students the opportunity to write responses to pre-lab questions as well as when they provided students the opportunity to interpret their experimental results in writing. The lesson components of *ARIES: M&F* teachers adhered to least, according to classroom observations and teachers' self-reports, were those items in which teachers permitted students to set-up their own equipment and led a discussion of students' *Science Journal* responses. This was not surprising, given our overall views that teachers seemed most likely to have earlier classes set-up the lab equipment for latter classes and to skip the discussion activities of a lesson, predominantly in the interest of maximizing class time.

Another finding from classroom observations and teacher interviews is the degree to which teachers incorporated *Teacher Guide* additions as they taught the unit. Trained observers reported teachers did not or only somewhat implemented these additions, while teachers reported doing so moderately in their interviews. These additions were included only in the teacher materials; therefore, student science journal data could not be used to corroborate these findings.

Data analyses show that that Fidelity as Adherence to Structure items *Interpreting results and discussion verbally* and completing the *Student Science Journal* were significantly positively correlated with classroom mean achievement. Further regression analyses also showed that together, after controlling for prior science GPA, these variables explained nearly half (45.6%) of the variance in student outcomes ($p < .01$).

Interestingly, data showed greater variability and greater relationships to outcomes for SCALE-uP's measure of Adherence to Student Fidelity to Structure, the *ARIES: M&F Science Journals*, as compared to our measures of Adherence to Teacher Fidelity to Structure, classroom observations and teacher self-reports. This could possibly be attributed to the differences in our methods of data collection. The ACOP was used in only one lesson, and the AIP relied on teachers' self-reports of the entire unit. However, the *Science Journals* provided us with tangible evidence of adherence to the entire unit. Therefore, the *Science Journals*, which also showed greater variability than the ACOP and the AIP, were probably a more accurate measure of adherence, further supporting the observed positive relationship to outcomes: the higher the fidelity to the unit structure, the higher the outcomes.

The results of this study indicate that *Interpreting results and discussion verbally* was a strong predictor of student outcomes. Other findings within SCALE-uP support this conclusion by showing that teachers' fidelity to *Students justifying their ideas* also predicts outcomes (see O'Donnell, Lynch, Watson, & Rethinam, 2007). We posit that since both items would occur during post-lab activity in *M&F*, this suggests that a teacher's ability to bring closure to the lab activity through teacher-guided discussion may be critical to student learning, as measured in this study by classroom observations and post-unit assessment scores.

Implications

In this paper, we offer two important reasons to conduct fidelity studies as part of intervention research. First, in order to establish a causal relationship from *Treatment A* to *Outcome B*, one must be able to ensure that *Treatment A* was implemented as expected. Results from SCALE-uP's Fidelity as Adherence to Structure data provide evidence that *ARIES: M&F* was indeed implemented in SCALE-uP Treatment classrooms with a range of adherence of 80-100%. The research literature is ambiguous as to whether this refers to internal or construct validity, or if, perhaps, construct validity should be subsumed under the scope of internal validity, or vice-versa, as suggested by Cherryholmes (1988); therefore, although he left that determination to researchers, future research should clearly distinguish between these two constructs.

SCALE-uP results also show (as will be discussed in the next symposium paper) that conceptualizing Fidelity to Structure and Fidelity to Process, as well as including the student in the fidelity study, provides a more comprehensive picture of the treatment intervention—in this case, *ARIES: M&F*. With this broader perspective of fidelity, future educational researchers may be able to more fully support claims that a particular treatment intervention, such as curriculum units, instructional strategies, and classroom conditions, are related to desired outcomes, for instance, classroom achievement, higher attendance rates, and decreased attrition rates. In the case of SCALE-uP, as results from this Fidelity as Adherence to Structure study has determined that *ARIES: M&F* was implemented in treatment classrooms, and that a number of Fidelity to Structure variables were related to student outcomes, our next step in this fidelity study is to determine how, if, at all, does Fidelity to Process relate to outcomes and does it account for the variance in the Treatment classrooms' mean achievement?

References

- American Association for the Advancement of Science (AAAS). (1993). *Benchmarks for science literacy*. New York: Oxford University Press.
- Cherryholmes, C. H. (1988). Construct validity and the discourses of research. *American Journal of Education*, *American Journal of Education*, *96*(3), 421-457.
- Cook, T. C., & Campbell, D. T. (1979). *Quasi-Experimentation: Design and analysis issues for field settings*. Chicago: Rand McNally College Publishing Company.
- 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.
- Dumas, J., Lynch, A., Laughlin, J., Smith, E., & Prinz, R. (2001). Promoting intervention fidelity: Conceptual issues, methods and preliminary results from the EARLY ALLIANCE prevention trial. *American Journal of Preventive Medicine*, *20*(Suppl.1), 38–47.
- 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 Theory and Practice*, *18*(2), 237-256.
- Emshoff, J. G., Blakely, C., Gottshalk, R., Mayer, J., Davidson, W. S., & Erickson, S. (1987). Innovation in education and criminal justice: Measuring fidelity of implementation and program effectiveness. *Educational Evaluation and Policy Analysis*, *9*(4), 300-311.
- Farrington, D. F. (2003). Methodological quality standards for evaluation research. *The Annals of the American Academy of Political and Social Science*, *587*, 49-68.

- Gersten, R., Fuchs, L. S., Compton, D., Coyne, M., Greenwood, C., & Innocenti, M. S. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional Children, 71*(2), 149-164.
- Harvard-Smithsonian Center for Astrophysics. (2001). *ARIES: Exploring motion and forces: Speed, acceleration, and friction*. Watertown, MA: Charlesbridge Publishing.
- Heimberg, R. G., & Becker, R. E. (1984). Comparative Outcome Research. In M. Hersen, L. Michelson, & A. S. Beck (Eds.). *Issues in Psychotherapy Research* (pp. 251-283). New York: Plenum Press.
- Kesidou, S., & Roseman, J.E. (2002). How well do middle school science programs measure up? Findings from Project 2061's curriculum review. *Journal of Research in Science Teaching, 39*(6), 522-549.
- Lawrence Hall of Science. (2000). *The real reasons for seasons—sun-earth connections*. Berkeley: The Regents of the University of California.
- Lynch, S. (2007, April). A model for fidelity of implementation in a study of a middle school science curriculum unit: Evaluation based on program theory. In C. O'Donnell (Chair), *Analyzing the relationship between Fidelity of Implementation (FOI) and student outcomes in a quasi-experiment*. Symposium to be conducted at the Annual Meeting of the American Educational Research Association, Chicago, IL.
- McGrew, J. H., Bond, G. R., Dietzen, L., & Salyers, M. (1994). Measuring the fidelity of implementation of a mental health program model. *Journal of Consulting and Clinical Psychology, 62*(4), 670-678.
- Michigan Science Education Resources Project. (1993). *GEMS: Chemistry That Applies*. The State of Michigan.

- Moncher, F. J., & Prinz, R. J. (1991). Treatment Fidelity in outcome studies. *Clinical Psychology Review, 11*, 247-266.
- O'Donnell, C., Lynch, S., Watson, W., & Rethinam, V. (2007, April). Teacher and Student Fidelity of Implementation and Relationships to Classroom Achievement. In C. O'Donnell (Chair), *Analyzing the relationship between Fidelity of Implementation (FOI) and student outcomes in a quasi-experiment*. Symposium to be conducted at the Annual Meeting of the American Educational Research Association, Chicago, IL.
- Pyke, C., & Ochsendorf, R. (2006, April). Concept assessment in curriculum unit evaluation. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA.
- Pyke, C., Lynch, S., Kuipers, J., Szesze, M., & Watson, W. (2005). *Implementation study of Exploring Motion and Forces (2004-2005): SCALE-uP Report No. 8*. Washington, DC: George Washington University, SCALE-uP.
- Reichardt, C. S. (2006). Estimating the Effects of Educational Interventions. In B. Schneider & S. K. McDonald (Eds.). *Scale-up in Education: Ideas in principle. Vol. 1* (pp. 79-99). Lanham, MD: Rowman & Littlefield Publishers, Inc.
- Songer, N. B., & Gotwals, A. W. (2005, April). Fidelity of implementation in three sequential curricular units. In S. Lynch (Chair), *The role of fidelity of implementation in experimental and quasi-experimental research designs: Applications in four studies of innovative science curriculum materials and diverse student populations*. Symposium conducted at the annual meeting of the American Educational Research Association, Montreal, Canada.

Teague, G. B., Bond, G. R., & Drake, R. E. (1998). Program fidelity in assertive community treatment: Development and use of a measure. *American Journal of Orthopsychiatry*, 68(2), 216-232.

U. S. Department of Education. (2006). *Education research request for applications - 84.305*. Retrieved May 15, 2006 from Institute for Education Sciences Web site:
<http://ies.ed.gov/funding/>

Vincent, M. L., Paine-Andrews, A., Fisher, J., Devereaux, R. S., Dolan, H. G., Harris, K. J., et al. (2000). Replication of a community-based multicomponent teen pregnancy prevention model: Realities and challenges. *Family and Community Health*, 23(3), 28-45.