The Physical Educator Vol. 72 • pp. 359–383 • 2015

TEACHER EDUCATION

Relationship Between Teacher Fidelity and Physical Education Student Outcomes

Jerry W. Loflin

Abstract

To evaluate the influence of scientifically based curricula intended for the physical education (PE) classroom, researchers should consider fidelity, or the extent to which teachers who are implementing a PE curricular intervention adhere to the model of curriculum and instruction inherent in the research design. The purpose of this study was to assess PE teachers' fidelity of implementation as they implemented a research-based PE curricular intervention and to examine the relationship between teachers' fidelity to the curricular intervention and student outcomes. Six teachers from three middle schools taught one 20-lesson unit from a PE intervention to their sixth grade students. Teachers' fidelity of implementation was documented using nonparticipatory observations. Students' knowledge acquisition was measured on a standardized knowledge test prior to and after the 20-lesson unit. Students' physical activity intensity levels were measured on lesson observation days using accelerometers. Multiple regression with fidelity score as the predictor

Jerry W. Loflin is an associate professor, School of General Studies and Academic Support, Davidson County Community College. Please send correspondence to Jerry W. Loflin, School of General Studies and Academic Support, Davidson County Community College, P.O. Box 1287, Lexington, NC 27293. E-mail: jwloflin@davidsonccc.edu

The Science of Healthful Living SEPA project was funded by the National Institutes of Health through the National Center for Research Resources Grant R25 RR032163 and is currently supported by the NIH Office of Research Infrastructure Programs / OD R25OD011063. Content is solely the responsibility of the author and does not necessarily represent the official views of ORIP / OD or NIH.

and knowledge acquisition as the criterion variable indicated that teachers' fidelity scores accounted for a large portion of variance in student knowledge growth ($R^2 = .79$, adj $R^2 = .74$, $p < .05$). Approximately 48% of all observed lesson time involved moderate to vigorous physical activities. Based on the findings from this research, it appears the more faithful teachers are to teaching research-based curricula as designed, the greater the impact the curricula can have on student achievement.

During a time of student, teacher, and school accountability, state legislators are taking control of schools through legislative actions across the United States. Educational mandates and policies related to school vouchers, school performance pay, teacher licensure, and testing legislation have been popular talking points among many legislators despite resistance and criticism from members of the education community. These reforms, along with legislation directed toward specific content areas such as PE, have redefined the way teachers teach and students learn.

As educational practices continue to receive strong criticisms, researchers are increasingly concentrating on the influence of scientifically based curricula and teaching methods on student outcomes. The need to attend to research-based practices also extends to the field of PE. Research should be directed toward understanding and measuring classroom environments in which PE students move and learn, as well as the educational practices to which students best respond based on their learning characteristics (Sallis et al., 2012).

To evaluate the influence of scientifically based curricula intended for the PE classroom, researchers should consider fidelity, or the extent to which teachers who are implementing a PE curricular intervention adhere to the model of curriculum and instruction inherent in the research design (Lillehoj, Griffin, & Spoth, 2004; Lynch & O'Donnell, 2005). Logically, if teachers do not implement research-based curricular interventions as designed, measured outcomes may not be attributable to the effectiveness of the intervention. However, in few studies have researchers measured fidelity of implementation (FOI), or "the extent to which delivery of an intervention adheres to the protocol or program model originally developed" (Mowbray, Holter, Teague, & Bybee, 2003, p. 315).

The purpose of this study was twofold. First, I assessed PE teachers' FOI as they implemented a research-based PE curricular intervention. Second, I examined the relationship between teach-

ers' fidelity to the curricular intervention and student outcomes. I was guided by two research questions: (a) To what extent do teachers implement lessons from a sixth grade PE curricular intervention with fidelity, and (b) what is the relationship between FOI and student outcomes (i.e., knowledge gain and physical activity intensity levels)?

Assessing FOI increases researchers' and practitioners' understanding of the fit between the intervention and the context. However, the complexity of the implementation context can affect the efficiency and generalizability of the innovation to other situations. Researchers have rarely used fidelity variables to explain intervention outcomes. Appropriate and accurate fidelity data can inform research and practice to better explain intervention outcomes.

During a time of teacher and student accountability, high-stakes testing, and No Child Left Behind mandates, schools and students must have effective curricula to increase achievement outcomes. Furthermore, given the calls by public health organizations to increase physical activity and decrease overweight and obesity levels in children and adolescents, school-based PE programs need evidence-based interventions proven to increase educational achievement and/or physical activity outcomes. Measuring FOI provides one method of assessing moderators of curricular effectiveness. According to O'Donnell (2008), there is need to conduct research on FOI and to understand why and how teachers impact intervention implementation. Studying FOI as part of curricular effectiveness research is critical to understanding intervention success or failure. Valid measures of FOI also can enhance statistical power and explain variance in outcome studies. In this research, I examined variables that enhanced or constrained the effectiveness of a large-scale PE intervention and the influence of FOI on student outcomes.

Fidelity of Implementation

Educators are increasingly expected to use an assortment of instructional and curricular approaches to address the diverse needs of students and improve student achievement. The No Child Left Behind Act of 2001 (NCLB, 2003) states that educators must use scientifically based teaching methods and strategies proven to be effective. Therefore, researchers and educators should be concerned about the degree to which the educational interventions they design, implement, and adopt are evidence based (Lynch & O'Donnell, 2005; Mowbray et al., 2003; Slavin, 2003).

To determine the effectiveness and impact of a research-based curricular intervention, it is important to assess the degree to which its implementation adheres to the model of curriculum and instruction inherent in the research design (O'Donnell, 2008). Although there are many competing definitions, I broadly define FOI "as the extent to which delivery of an intervention adheres to the protocol or program model originally developed" (Mowbray et al., 2003, p. 315). Assessing FOI in the context of any intervention study is essential because "failure to establish fidelity can severely limit the conclusions that can be drawn from any outcome evaluation" (Dumas, Lynch, Laughlin, Smith, & Prinz, 2001, p. 39).

In studies assessing FOI, empirical evidence indicates significant correlations between the extent to which intervention studies are implemented with fidelity and the level of treatment outcomes (Blakely et al., 1987; Dane & Schneider, 1998; Ruiz-Primo, 2006). Although most of these studies have taken place in the context of public health and counseling research (O'Donnell, 2008), the number of FOI studies in which the effectiveness of K–12 curricular interventions has been evaluated is increasing (Lillehoj et al., 2004; Lynch & O'Donnell, 2005; O'Donnell, 2008). For evaluation of teacher fidelity in the context of such research-based curricula, a five-component FOI framework has been proposed to include adherence, coverage, program differentiation, delivery quality, and participant responsiveness (Lynch & O'Donnell, 2005; O'Donnell, 2008). Given the "…compelling need to better understand how curricula, instruction, and student diversity affect student achievement in K–12 classrooms" (Lynch & O'Donnell, 2005, p. 2), fidelity studies in which K–12 curricular intervention effectiveness has been investigated deserve consideration and have been recommended (O'Donnell, 2008).

Notably, to assess FOI, it is not necessary to evaluate all five components. Researchers should attend to components that are of interest to their study (Lillehoj et al., 2004; Lynch & O'Donnell, 2005), and few researchers have reported measuring all five. For example, Lynch and O'Donnell (2005) monitored adherence, delivery quality, and participant responsiveness in evaluating teacher fidelity during the implementation of a research-based middle school science curriculum. Similarly, Lillehoj et al. (2004) focused on adherence and delivery quality while assessing fidelity in the context of a middle school problem behaviors study. For purposes of this study, the fidelity components of adherence, coverage, and delivery quality were of primary interest.

Science of Healthful Living Project

I implemented this study as a part of and during the second year of the larger 5-year Science of Healthful Living (SHL) project conducted by the University of North Carolina at Greensboro with support provided by the National Institutes of Health (NIH). The goal of the SHL project was to design and evaluate a science-based middle school *SHL* curriculum to increase students' cognitive knowledge and interest in health-related science, increase their intention to pursue a life science–related career, and improve their family members' understanding of NIH-funded clinical and basic research results. For the SHL project, a health-related science-based curriculum was designed for middle school PE. The curriculum consists of two 20-lesson units, Cardio Fitness Club and Healthy Lifestyles, in which students engage in the scientific inquiry process in PE. Students engage in physically active lessons to examine the effects of exercise and healthy nutrition on their bodies. They explore topics such as the physiology of exercise, food-fueled energy systems, caloric balance, nutrition/portion sizes, and the role of physical activity and nutrition in stress management. The scientific inquiry process is embedded in each lesson using the 5Es (i.e., Engagement, Exploration, Explanation, Elaboration, and Evaluation) learning cycle strategy (Bybee et al., 1989) to help physical educators and students engage in the scientific inquiry process in a hands-on approach to science. Students are afforded opportunities to make predictions about physical activity concepts, test their predictions, and draw conclusions about their predictions in physically active environments. Additional information about the larger SHL project can be found at http://www.uncg.edu/hhs/science-of-healthful-living.

During Year 2, the SHL project was set in seven local education agencies in the Piedmont region of North Carolina. Twenty-five middle schools, including five new middle schools recruited in Year 2 of the study, participated in the project, including over 16,000 diverse students and 70 middle school PE teachers. A randomized controlled experimental design was used to stratify the middle schools during Year 1 into matched pairs according to socioeconomic status (derived from free and reduced-price meals school-level data) and scores on end-of-grade tests in mathematics and reading. One school in each pair was randomly assigned to teach the *SHL* curriculum (experimental school) and one school taught the traditional sport-based or multiactivity PE (control school).

Method

Research Design

I used a mixed methods design to collect and analyze quantitative and qualitative data examining FOI in the SHL project (Creswell, 2013; Creswell & Plano Clark, 2011; Tashakkori & Teddlie, 2010). Using multiple data collection strategies provided a more complete analysis than any single method alone could accomplish when addressing the problem (Creswell, 2013; Creswell & Plano Clark, 2011; Tashakkori & Teddlie, 1998, 2010). A mixed methods design was appropriate because FOI measurement and analysis required quantitative and qualitative data collection and analysis during various study stages. Mixed methods approaches offered complementary insights into understanding the faithful implementation of the *SHL* curriculum and the impact FOI had on student outcomes.

Research Sampling

A selective sample of schools was chosen from the 15 schools randomly assigned to implement the *SHL* curriculum. Any experimental school (n = 15) that had at least one new SHL teacher during the 2012–2013 school year was considered for the study. Of the 15 schools with teachers new to the project, three schools located in a cluster were selected to facilitate data collection. FOI in each participating sixth grade PE class (n = 6) at the three schools was examined in this study.

Research Sites and Participants

Participants were middle school PE teachers and their sixth grade students from Albany, New River, and St. Anthony Middle Schools (pseudonyms). These middle schools were located in the Piedmont region of North Carolina and served students in Grades 6–8, aged 11–14. Each school was selected because the teachers were voluntarily participating in the SHL project effectiveness study for the first time during the 2012–2013 school year. The three schools selected for this research had been randomly assigned to the intervention group. PE teachers taught the *SHL* curriculum consisting of 20 structured lessons from the sixth grade Cardio Fitness Club unit.

Albany Middle School had over 220 sixth grade students enrolled during the 2012–2013 academic year. Approximately 52% of the students were male, 14% were White, 69% were Black, 7% were Hispanic, 6% were Asian, and 73% received free and reduced-price

meals (FARM; Education First NC School Report Cards, 2011). Three middle school PE teachers (one male and two females) were employed at Albany Middle School during the 2012–2013 academic year. Students received gender-segregated PE instruction for the school year on a 1-week PE and 1-week health education rotation. The two female teachers and their three sixth grade classes participated in this study. The two sixth grade PE classes assigned to the same teacher were combined in this study to form one sixth grade PE class.

New River Middle School had over 190 sixth grade students enrolled during the 2012–2013 academic year. Approximately 50% of the students were male, 35% were White, 12% were Black, 48% were Hispanic, and 77% received FARM (Education First NC School Report Cards, 2011). Two middle school PE teachers (one male and one female) were employed at New River Middle School during the 2012–2013 academic year. Students received coeducational PE instruction for one semester using a 4-week PE and 2-week health education rotation. Only the sixth grade students enrolled in three PE classes during the first semester participated in this study. Two of the three PE classes occurred at the same time and place and were team taught. These two classes were combined to form one class and assigned to the lead teacher.

St. Anthony Middle School had over 180 sixth grade students enrolled during the 2012–2013 academic year. Approximately 48% of the students were male, 52% were White, 17% were Black, 25% were Hispanic, and 60% received FARM (Education First NC School Report Cards, 2011). Two middle school PE teachers (one male and one female) were employed at St. Anthony Middle School during the 2012–2013 academic year. Students received coeducational PE instruction for one semester using a 4-week PE and 2-week health education rotation. Only the sixth grade students enrolled in three PE classes during the first semester participated in this study. Two of the three PE classes occurred at the same time and place and were team taught. These two classes were combined to form one class and assigned to the lead teacher.

Data Sources

Lesson observations. I collected detailed field notes for each observed lesson and systematically coded the field notes to examine adherence based on coverage and delivery quality per sixth grade PE class.

Standardized knowledge tests. Student knowledge test gain scores served as one of two intervention outcomes in this study. Sixth grade students receiving the intervention completed a 10-item pre- and postinstruction knowledge test. The two assessments were multiple-choice (four choices) and administered through Qualtrics (Qualtrics Labs, Inc., Provo, UT). The test items for the Cardio Fitness Club unit were created to match the content covered in the unit to increase the content validity of the tests. The pretests were administered during the lesson immediately preceding Lesson 1, and the posttests were administered immediately after Lesson 20.

In-class student physical activity intensity levels. The second outcome variable, student physical activity intensity levels, was measured with Actigraph GT3X+ accelerometers (Actigraph, Pensacola, FL). The GT3X+ accelerometer is a small (4.6 cm × 3.3 cm × 1.5 cm), light (19 g) device contained in a plastic case that participants attach using an adjustable elastic waist strap. The solid state accelerometer can be used to measure and record physical activity vector magnitude counts in three physical axes. In the prefiltered raw mode, accelerometer output represents the actual G-force signal untransformed by a band-pass filter and sampled at a frequency of 30 Hz (John & Freedson, 2012). The G-force signal is then converted by a 12-bit analog-to-digital converter with the output band-pass filtered across a frequency range of 0.25 Hz to 2.5 Hz (John & Freedson, 2012). The filtered signal is then rectified and integrated over a user-specified time frame (i.e., epoch). At the end of each epoch, a vector magnitude count is calculated and stored in memory, the integrator is reset, and the process starts over. The sampling epoch length was set at 10 s for this study.

Intensity levels were categorized as sedentary, light, moderate, vigorous, and very vigorous. Intensity level category cutoff points (sedentary: vector magnitude count < 150; light: 150–499; moderate: 500–3,999; vigorous: 4,000–7,599; very vigorous: > 7,599) categorized by Freedson, Pober, and Janz (2005) determined student physical activity intensity levels (i.e., sedentary, light, moderate, vigorous, or very vigorous) during SHL lessons. The device has demonstrated high interinstrument reliability (r = .86–.89) and intrainstrument reliability (r = .86; Melanson & Freedson, 1995).

Data Collection

Lesson observations. Once knowledge pretesting was complete at each participating school, participating teachers began teaching

the sixth grade Cardio Fitness Club unit (Lessons 1–20) to their sixth grade students. I began visiting each participating school beginning with Lesson 6 to collect detailed field notes on Lessons 6–20. Observation field notes for Lessons 1–5 were not collected to allow the teachers to become familiar with the lesson format, pacing, and content prior to data collection. Observation field notes were recorded on the standard qualitatively oriented SHL observation form. The SHL observation form was used because it aligned with the general structure of each lesson and facilitated field notes collection. It contained sections for recording the standard school, teacher, grade, period, and date of observation and categories for describing in detail events that occurred during each of the 5Es within the lessons. The categories for the 5Es contained subcategories related to lesson timing, delivery quality, student enjoyment, and student journal use during that section. I observed, in a nonparticipatory capacity, at least 40% of Lessons 6–20 implemented in each sixth grade class. Participating teachers were not informed about when or which lessons I would observe.

Standardized knowledge growth. Depending on school academic calendars and school computer lab availability, participating teachers administered the sixth grade, 10-item standardized knowledge pretest to their sixth grade PE students prior to teaching Cardio Fitness Club Lesson 1. I sent a Qualtrics pretest hyperlink to teachers via e-mail communication. Teachers administered the pretest to their students in the schools' computer lab or via individual student laptop computers or tablets. After completing the 20 lessons, participating teachers readministered the 10-item test used for pretesting via Qualtrics. Class completion rates were tracked in Qualtrics during the pre- and posttests, and teachers were reminded which classes had not completed the testing.

In-class student physical activity intensity levels. After pretesting was completed at each school, each participating teacher assisted me in selecting six students per sixth grade class (n = 36 students). Students were selected purposefully to generate a balanced sample based on gender and body mass index to wear physical activity accelerometers.

During Lessons 6–20 of the sixth grade Cardio Fitness Club unit, I collected physical activity intensity level data each time I observed a lesson. The six students selected per class were introduced to and assigned a uniquely numbered accelerometer. Each student wore the assigned accelerometer at least once during Lessons 1–5 to learn

how to put it on and to minimize the reactivity effect of wearing the accelerometer for the first time. Students put on the accelerometer as they entered the gymnasium from the locker room and returned it just prior to reentering the locker room at the end of the class period. I observed each student attach the accelerometer to his or her waist to ensure the proper positioning of the accelerometer. Students were asked to wear the device on their waist above their right knee using a supplied elastic waist band. During physical activity intensity level data collection, I recorded lesson unit and number and lesson start and end time.

Data Reduction

Lesson observations. I transcribed the lesson observation field notes into individual Microsoft Word files organized by school and teacher observation dates. The transcription process allowed me to become more acquainted with the data.

Standardized knowledge growth. Student standardized knowledge test item responses were downloaded from Qualtrics to the Statistical Package for the Social Sciences (SPSS; IBM SPSS Statistics Version 20, Release 20.0.0, IBM, Inc., Armonk, NY) and organized by teacher and class. On the pre- and posttests, individual item responses were converted from answer selected (i.e., 1 = A, 2 = B, 3 = C, and 4 = D) to answer correct (coded as 1) or incorrect (coded as 0). Pre- and posttest total raw scores were determined for each student by summing their individual item responses. Each student's pretest raw score was then subtracted from his or her corresponding posttest raw score to produce a change score. The students' change scores were averaged in each class to produce an average class change score (see Table 1).

In-class student physical activity intensity levels. In-class student physical activity intensity levels were recorded as vector magnitudes by the Actigraph GT3X+ accelerometers. Vector magnitudes were downloaded from ActiLife 6 (Actigraph, Pensacola, FL) to SPSS. Total vector magnitude for each measured student was converted to vector magnitude per minute to reflect in-class physical activity intensity level. Vector magnitude per minute was determined by dividing the accelerometer physical activity count sum by the total minutes (lesson end time minus lesson start time) in the respective SHL lesson. Each student's in-class physical activity intensity level per observed lesson per class was averaged to produce an average class physical activity intensity level (see Table 1).

Table 1
Class-Level Student Outcomes

School and teacher	Number of students	Knowledge change (*SD*)	PA intensity level (*SD*)
Albany Middle			
Teacher 1	23	13.9% (21.3)	2,260.03 CPM (225.75)
Teacher 2[a]	55	12.0% (20.9)	2,069.33 CPM (85.47)
New River Middle			
Teacher 3[b]	45	22.0% (19.4)	2,186.58 CPM (351.16)
Teacher 4	28	15.7% (19.9)	1,734.3 CPM (207.44)
St. Anthony Middle			
Teacher 5[b]	55	17.8% (20.4)	2,520.47 CPM (426.65)
Teacher 6	23	20.9% (19.1)	2,468.28 CPM (117.57)

Note. CPM = vector magnitude count per minute.
[a]Outcome data were combined from two physical education classes assigned to this teacher. [b]Teacher team-taught and absorbed the students assigned to the other teacher assisting with the lesson.

Data Analysis

Inductive analysis was used to analyze the transcribed lesson observations immediately following each school visit. "Inductive analysis involves discovering patterns, themes, and categories in one's data. Findings emerge out of the data, through the analyst's interactions with the data…" (Patton, 2002, p. 453). Using the constant comparative method (Glaser, 1978), I looked for codes and categories in the data. As described by Corbin and Strauss (2008), I first employed open coding of the data to determine if there were recurring topics that could be classified into categories and labeled. I elected to open code the observation field notes by hand to facilitate constant comparison and become more acquainted with the data. The generated open codes described and referred to actual events in the PE classes. Open coding allowed me to divide the data into distinct categories, look for differences and similarities within the observation field notes, and ask questions about phenomena revealed in the data (Corbin & Strauss, 2008). The open coding procedure allowed me to concentrate on the events observed in the PE classes and guided my observations as I explored cases of teacher fidelity.

Once categories were developed and labeled, I looked for relationships among the categories using axial coding. During the axial coding process, I developed subcategories focusing on phenomena, antecedents, contexts, and any intervening conditions. As suggested by Corbin and Strauss (2008), I ended my analysis with axial coding because I was only interested in using some of the tools of grounded theory and was not focused on developing theory.

Using the open and axial codes, I developed a rubric that represented the patterns reflective of the teachers' adherence, delivery quality, and coverage. A 42-item dichotomous rubric was developed from the qualitative analysis to measure teacher adherence, coverage, and delivery quality (see Figure 1). I then used the rubric to quantify teachers' FOI per observed lesson.

Once the qualitative data analysis was complete, I used the teachers' FOI scores generated by the rubric to further inform the quantitative analysis. Multiple regression was used to measure the contribution of overall lesson FOI and FOI to each of the 5Es to the prediction of class knowledge change and class physical activity intensity levels.

Results

During the coding process, four major categories emerged from the data: (a) high adherence and good delivery quality, (b) high adherence but poor delivery quality, (c) low adherence but good delivery quality, and (d) low adherence and poor delivery quality. The four categories emerged throughout and across lesson observations. No one category represented any single lesson as teachers fluctuated between the four categories as lessons progressed through their critical components. The developed rubric, therefore, contains six sections related to the critical components of the lessons. Within the six sections are statements to prompt the observer to ascertain the level of teacher delivery quality, coverage, and adherence to the lesson as designed. Table 2 displays the participating teachers' fidelity scores determined by the 42-item dichotomous rubric.

Multiple regression was used to assess the contribution of fidelity to the prediction of class knowledge change and class physical activity intensity levels over the course of the 20-lesson Cardio Fitness Club unit. The outcomes of paired sample *t* tests indicated the changes between each class's average pretest score and average posttest scores were statistically significant (see Table 3). In addition, the correlation between each class's pretest scores and posttest scores were statistically significant (see Table 3).

Science of Healthful Living (*SHL*) Fidelity of Implementation Observation Rubric (Target Schools)

Date of Observation_____ Raw Score = _____/42 = _____
Date Observation Scored_____
School ID_____
Teacher ID_____
Class Period Start and End Time_____
Grade_____
SHL Module and Lesson_____

Please check Yes (1 point) or No (0 points) for each item.

	NO	YES	COMMENTS
1. Teacher and student materials (task cards, journals, equipment) ready before class			
2. Teacher organized and familiar with lesson			
3. Teacher stated the lesson essential question			
4. Teacher referred to the lesson evaluation rubric			
5. Teacher followed the timing of the lesson (or made lesson longer)			
6. Teacher emphasized the intended science concepts throughout the lesson			
Engagement			
7. Engagement occurred as designed or occurred with appropriately substituted activities			
8. Journal used as intended			
9. Teacher models movement tasks			
10. Teacher corrects immature movement patterns			
11. Teacher checks for student understanding before moving on to the next step			
12. Teacher encourages student participation			
13. Students are engaged in the lesson segment			
14. Students are successful during activities			
Exploration			
15. Exploration occurred as designed or occurred with appropriately substituted activities			
16. Journal used as intended			
17. Teacher models movement tasks			
18. Teacher corrects immature movement patterns			
19. Teacher checks for student understanding before moving on to the next step			
20. Teacher supplies students adequate think time			
21. Teacher encourages student participation			
22. Students are engaged in the lesson segment			
23. Students are successful during activities			
Explanation			
24. Explanation occurred as designed or occurred with appropriately substituted activities			
25. Journal used as intended			
26. Teacher checks for student understanding before moving on to the next step			
27. Teacher supplies students adequate think time			
28. Teacher provides students ample opportunities to talk with their peers			
29. Teacher provides students individual turns to speak			
30. Teacher encourages student participation			
31. Students are engaged in the lesson segment			
Elaboration			
32. Elaboration occurred as designed or occurred with appropriately substituted activities			
33. Journal used as intended			
34. Teacher checks for student understanding before moving on to the next step			
35. Teacher supplies students adequate think time			
36. Teacher provides students ample opportunities to talk with their peers			
37. Teacher provides students individual turns to speak			
38. Teacher encourages student participation			
39. Students are engaged in the lesson segment			
Evaluation			
40. Evaluation occurred as designed or occurred with appropriately substituted activities			
41. Journal used as intended			
42. Physical activity homework assigned			

Figure 1. Fidelity of implementation observation rubric.

Table 2
Teachers' Average Fidelity of Implementation Scores

School and teacher	Number of observations	Total fidelity score	Engagement	Exploration	Explanation	Elaboration	Evaluation
Albany Middle							
Teacher 1	6	.48	.84	.81	.55	.06	.17
Teacher 2	6	.42	.77	1.0	.38	.1	0
New River Middle							
Teacher 3	7	.83	1.0	.82	.95	.77	.47
Teacher 4	7	.75	.83	1.0	.71	.84	.71
St. Anthony Middle							
Teacher 5	7	.8	.83	.8	1.0	.8	.66
Teacher 6	7	.82	1.0	.87	.91	.95	.74

Table 3
Differences and Correlations Between Pre- and Posttest Scores

School and teacher	Pretest (%)	Posttest (%)	Differences			Correlations	
	M (SD)	*M (SD)*	*t*	*df*	*p**	*r*	*p***
Albany Middle							
Teacher 1	31.7 (15.3)	45.7 (21.1)	3.14	22	.005	.35	.04
Teacher 2	33.1 (15.7)	45.1 (20.5)	4.27	54	< .001	.36	.01
New River Middle							
Teacher 3	27.8 (14.3)	49.8 (20.3)	7.62	44	< .001	.41	.01
Teacher 4	35.4 (16.2)	51.1 (18.5)	4.18	27	< .001	.35	.04
St. Anthony Middle							
Teacher 5	32.9 (14.9)	50.7 (17.8)	6.47	54	< .001	.23	.03
Teacher 6	30.9 (10.0)	51.7 (19.0)	5.25	22	< .001	.26	.04
All Teachers	31.9 (14.8)	48.8 (19.5)	12.58	228	< .001	.32	< .001

*$p < .01$. **$p < .05$.

Regression diagnostic statistics were calculated with the multiple regression analyses to check data for cases that exerted undue influence over the parameters of the model and/or were extreme outliers (Field, 2009). Deviant cases, either individually or in combination with other cases, can significantly influence regression statistics. The following regression diagnostics were examined to check for influential data points: (a) Cook's distance (a measure of the overall influence of a case on the model), (b) leverage or hat values (gauges the influence of the observed value of the outcome variable over the predicted values), and (c) standardized DFBETA (detects cases that influence the regression coefficient). Additionally, collinearity diagnostics (i.e., tolerance and variance inflation factor) were conducted to ensure that no variables were closely related to one another. The obtained variance inflation factor in all cases but one was less than 11.5. Elaboration fidelity score had a strong collinearity with the other predictors and was excluded from all multiple regression models.

Class-level knowledge change and class-level physical activity intensity level were the criterion variables in the multiple regression analyses. Overall fidelity score was the sole predictor variable in the first analysis for each criterion variable, and the scores for the four *E*s (i.e., Engagement, Exploration, Explanation, and Evaluation) were entered in the second analysis for each criterion variable.

Overall fidelity score accounted for 79%, adj $R^2 = .74$, $F(1, 4) = 15.4$, $p < .05$, of the variance in knowledge change. The overall fidelity beta coefficient was statistically significant (see Table 4). When the four *E*s were entered as predictors, Engagement, Exploration, Explanation, and Evaluation contributed to the overall model fit for knowledge change, $R^2 = .99$, adj $R^2 = .99$, $F(4, 1) = 416.99$, $p < .05$. The beta coefficients for Engagement and Explanation were statistically significant (see Table 4).

Table 4
Regression Output for Knowledge Change

Fidelity component	Knowledge Change						
	M	***SD***	***B***	***SE B***	**β**	***t***	***p***
Overall	.68	.19	0.19	0.05	.89	3.92	.02
Engagement	.88	.1	0.23	0.01	.58	17.35	.04
Exploration	.88	.09	0.08	0.02	.2	4.5	.14
Explanation	.75	.25	0.13	0.01	.82	10.04	.04

Table 4 (cont.)

Fidelity component	*M*	*SD*	Knowledge Change *B*	*SE B*	β	*t*	*p*
Evaluation	.46	.31	−0.03	0.01	−.24	−3.87	.16
Elaboration	.59	.4	Excluded from model				

Therefore, PE students taught by PE teachers with higher fidelity levels displayed more knowledge growth. The specific *E*s that significantly contributed to knowledge gain included Engagement and Explanation.

Multiple regression targeting class-level physical activity intensity levels did not produce significant findings using overall fidelity score or the scores of the four *E*s (see Table 5). The mean proportion of time spent in moderate to vigorous physical activities and sedentary activities over 40 Cardio Fitness Club lesson observations was 47.9% (*SD* = 2.4%) and 41.1% (*SD* = 2.7%), respectively (see Table 6). Twenty-one of the 40 (52.5%) lesson observations met the Centers for Disease Control and Prevention's recommendation (U.S. Department of Health and Human Services, 2010) of spending at least 50% of PE class time participating in moderate to vigorous physical activities.

Table 5
Regression Output for Physical Activity Intensity Levels

Fidelity component	*M*	*SD*	Physical activity intensity levels *B*	*SE B*	β	*t*	*p*
Overall	.68	.19	514.92	437.69	.51	1.18	.31
Engagement	.88	.1	−434.57	1,463.11	−.23	−0.3	.82
Exploration	.88	.09	−1,537.38	1,966.03	−.77	−0.78	.58
Explanation	.75	.25	−131.28	1,388.38	−.17	−0.1	.94
Evaluation	.46	.31	414.15	863.57	.68	0.48	.72
Elaboration	.59	.4	Excluded from model				

Table 6
Mean Percentage of Lesson Time Spent at Different Physical Activity Intensities

School and teacher	Sedentary[a] (*SD*)	Light[a] (*SD*)	Moderate[a] (*SD*)	Vigorous[a] (*SD*)	Very vigorous[a] (*SD*)
Albany Middle					
Teacher 1[b]	37.9 (9.4)	14.2 (3.4)	41.5 (7.2)	4.2 (1.9)	2.1 (2.1)
Teacher 2[b]	42.4 (3.6)	12.5 (0.4)	38.6 (3.6)	5.5 (0.6)	1.0 (0.7)
New River Middle					
Teacher 3[c]	37.9 (7.4)	10.8 (2.2)	41.5 (7.1)	7.8 (3.4)	1.9 (1.0)
Teacher 4[c]	44.5 (16.8)	9.4 (2.7)	39.5 (11.6)	5.4 (4.1)	1.2 (0.7)
St. Anthony Middle					
Teacher 5[c]	41.0 (4.0)	8.9 (3.0)	36.8 (4.8)	9.8 (4.9)	3.5 (2.1)
Teacher 6[c]	42.7 (0.1)	10.2 (1.6)	35.1 (1.6)	9.0 (1.8)	3.0 (0.7)
All Teachers	41.1 (2.7)	11.0 (2.0)	38.8 (2.6)	7.0 (2.3)	2.1 (1.0)

[a]Cut-off points suggested by Freedson, Pober, and Janz (2005). [b]Physical activity intensity level data collected on six lessons. [c]Physical activity intensity level data collected on seven lessons.

Discussion

Few researchers have examined the impact FOI has on student outcomes (O'Donnell, 2008). They have suggested that high-fidelity implementation enhances intervention outcomes. The purpose of this study was to assess PE teachers' FOI as they implemented a research-based PE curricular intervention and to examine the relationship between teachers' fidelity to the curricular intervention and student outcomes. Several findings deserve discussion.

First, students receiving the SHL curricular intervention made statistically significant improvements in their fitness-based knowledge. The mean gain for the six PE classes was 16.9% (*SD* = 20.3%). However, after inspecting the mean pretest score (*M* = 31.9%, *SD* = 14.8%) and the mean posttest score (*M* = 48.8%, *SD* = 19.5%), students performed below what most educators would consider proficient at the end of the intervention. The majority of the teachers and students in this study were participating in multiactivity PE pro-

grams before agreeing to participate in this study. The mean pretest score suggests that students had limited prior knowledge of fitness-based concepts. A large body of findings shows that learning proceeds primarily from relevant prior knowledge, and only secondarily from the taught information (Dochy, Segers, & Buehl, 1999). The limited relevant prior knowledge of the sixth grade students in the present investigation could have constrained their knowledge gain. It is unclear how much fitness education the students in this study received prior to enrolling in sixth grade, but it is clear that the North Carolina Essential Standards for Healthful Living for Grades 3–5 includes fitness education as a part of its standard course of study (Public Schools of North Carolina, 2011). Additionally, if the students in this study received fitness education in fifth grade, the dose and coverage would be questionable given the current trends in instructional time in elementary school PE.

Second, multiple regression shows that overall FOI accounted for 79% of the variance in fitness-based knowledge growth in sixth grade students participating in the curricular intervention. Overall, FOI made a statistically significant contribution to the prediction of student knowledge gain. Specific *E*s (i.e., Engagement and Explanation) had a statistically significant effect on knowledge gain. These findings reinforce earlier research demonstrating that FOI has statistically significant implications on student outcomes (Allinder, Bolling, Oats, & Gagnon, 2000; Butler-Songer & Gotwals, 2005; Hall & Loucks, 1977; Penuel & Means, 2004; Ysseldyke, Spicuzza, Kosciolek, & Boys, 2003). Therefore, the more faithful teachers are to teaching research-based curricula as designed, the greater the impact the curricula can have on student achievement. During a time of student and teacher accountability, it is critical that educators use curricula proven to be effective.

Although no statistically significant relationships were found between teacher fidelity levels and student physical activity intensity levels, it is worth noting that approximately 48% of all observed lesson time involved moderate to vigorous physical activities. This proportion is slightly higher compared with other studies (e.g., Chow, McKenzie, & Louie, 2009; McKenzie et al., 2006) reporting physical activity intensity levels during PE lessons. I also found that in over 52% of the observed lessons, students spent at least 50% of the PE lesson time participating in moderate to vigorous physical activities. Data like these are limited in other studies, but it is clear that approximately 48% of the observed lessons did not meet the

Centers for Disease Control and Prevention's recommendation for moderate to vigorous physical activity.

As with most research studies, this investigation had a few limitations deserving discussion. First, I relied on six fidelity checks at one school and seven fidelity checks at the other two schools. A more thorough account of teacher FOI and its impact of student outcomes would have been attained with more frequent fidelity checks (e.g., daily). Second, although a representative matched sample with control group was used in the NIH randomized control trial, only a subset of this population was examined in this research. Specifically, the first-year (in the intervention) middle school teachers participating in this study were not demographically representative of the sample of teachers participating in the larger NIH-sponsored study. Therefore, the findings of the study have limited generalizability. Additional research should be implemented using a larger, demographically heterogeneous sample size to satisfy the statistical assumptions associated with multiple regression.

The findings have potential to address the challenge of implementing effective approaches in practice. Cook, Landrum, Tankersley, and Kauffman (2003) highlighted that approaches may be rendered counterproductive or ineffective when implemented with inadequate fidelity or if not used with appropriate dosage (amount of treatment). Considering this problem in the context of the current investigation, teachers implementing the *SHL* curriculum (a research-based PE intervention) with lower FOI experienced smaller increases in knowledge gain compared with teachers implementing the intervention with higher FOI. Failing to consider FOI is a major threat to internal validity. The role of the intervention in enhancing or constraining outcomes cannot be fully determined without measuring FOI.

Using procedures integrated into the research design for maximizing internal validity enhances the integrity of study findings. The findings of this study highlighted that the extent to which the intervention was taught as prescribed (quality), the extent to which the content was delivered (coverage), and the extent to which the internal structure of the lessons were maintained during delivery (structure) can significantly affect student outcomes. Given that few researchers to date have assessed the role of FOI on PE outcomes, future fidelity research in PE is justified. Monitoring FOI may modify outcomes of existing and future program effectiveness, efficacy, and evaluation studies of PE curricula.

Additionally, practicing teachers need continual professional development and support to implement and maintain evidence-based practices in their classrooms. To build and maintain teachers' capacities to deliver effective programs, they need frequent and consistent support for a minimum of 1 year in the form of instruction, feedback, and motivators (National Institute of Child Health and Human Development, 2000). Furthermore, researchers (e.g., Gersten & Brengelman, 1996) have suggested that successful teacher training for evidenced-based practices follows when content and approaches are similar to the values, beliefs, and goals of important school staff (i.e., lead teachers and administrators). It is also important that coaches and principals use valid and reliable fidelity measures to facilitate continuing mentoring and classroom observations. Structuring observations to identify crucial teaching behaviors offers a method to examine teacher ability/effectiveness to implement evidence-based practices. In this research study, adherence to lesson delivery as intended (i.e., following the lesson structure and script) appeared to increase the contribution of the PE curriculum to student knowledge gain. Teachers and administrators may consider incorporating fidelity into existing documents used for mentoring or administrative observations.

In summary, statistically significant growth was evident in the fitness-based knowledge of sixth grade PE students, but results varied as PE teachers taught the *SHL* curriculum with different fidelity levels. High fidelity requires time and effort. In spite of years of teaching experience, it requires numerous professional development opportunities to educate teachers and staff about how to implement evidenced-based educational interventions. Increasing fitness-based knowledge at the middle school level is possible when scientifically based PE instruction is conveyed with fidelity.

References

Actigraph. (2012, June). ActiLife 6 data analysis software (Version 6.2) [Software]. Pensacola, FL: Actigraph. Available from http://www.actigraphcorp.com/

Allinder, R. M., Bolling, R. M., Oats, R. G., & Gagnon, W. A. (2000). Effects of teacher self-monitoring on implementation of curriculum-based measurement and mathematics computation achievement of students with disabilities. *Remedial and Special Education, 21*, 219–226. doi:10.1177/074193250002100403

Blakely, C., Mayer, J., Gottschalk, R., Schmitt, N., Davidson, W., Roitman, D., & Emshoff, J. (1987). The fidelity-adaptation debate: Implications for the implementation of public sector social programs. *American Journal of Community Psychology, 15*, 253–268.

Butler-Songer, N., & Gotwals, A. W. (2005, April 14). *Fidelity of implementation in three sequential curricular units.* Paper presented at Fidelity of Implementation Symposium at AERA, Montreal, Canada. Retrieved from http://www.gwu.edu/~scale-up/fidelity.html

Bybee, R. W., Buchwald, C. E., Crissman, S., Heil, D. R., Kuerbis, P. J., Matsumoto, C., & McInerney, J. D. (1989). *Science and technology education for the elementary years: Frameworks for curriculum and instruction.* Washington, DC: The National Center for Improving Science Education. Retrieved from ERIC database. (ED314237)

Chow, B. C., McKenzie, T. L., & Louie, L. (2009). Physical activity and environmental influences during secondary school physical education. *Journal of Teaching in Physical Education, 28,* 21–37.

Cook, B. G., Landrum, T. J., Tankersley, M., & Kauffman, J. M. (2003). Bringing research to bear on practice: Effecting evidence-based instruction for students with emotional or behavioral disorders. *Education and Treatment of Children, 26*, 345–361.

Corbin, J., & Strauss, A. (2008). *Basics of qualitative research* (3rd ed.). Thousand Oaks, CA: Sage.

Creswell, J. W. (2013). *Qualitative inquiry and research design: Choosing among five approaches* (3rd ed.). Thousand Oaks, CA: Sage.

Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research* (2nd ed.). Thousand Oaks, CA: Sage.

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

Dochy, F., Segers, M., & Buehl, M. M. (1999). The relation between assessment practices and outcomes of studies: The case of research on prior knowledge. *Review of Educational Research, 69*, 145–186. doi:10.3102/00346543069002145

Dumas, J. E., Lynch, A. M., Laughlin, J. E., Smith, E. P., & Prinz, R. J. (2001). Promoting intervention fidelity: Conceptual issues, methods, and preliminary results from the EARLY ALLIANCE prevention trial. *American Journal of Preventive Medicine, 20*(1, Suppl. 1), 38–47. doi:10.1016/S0749-3797(00)00272-5

Education First NC School Report Cards. (2011, October). *NC school report cards: 2010–2011 school year.* Retrieved from www.ncreportcards.org

Field, A. (2009). *Discovering statistics using IBM SPSS Statistics* (3rd ed.). Thousand Oaks, CA: Sage.

Freedson, P., Pober, D., & Janz, K. F. (2005). Calibration of accelerometer output for children. *Medicine & Science in Sports & Exercise, 37*(Suppl. 11), S523–S530. doi:10.1249/01.mss.0000185658.28284.ba

Gersten, R., & Brengelman, S. U. (1996). The quest to translate research into classroom practice: The emerging knowledge base. *Remedial and Special Education, 17*, 67–74. doi:10.1177/074193259601700202

Glaser, B. (1978). *Theoretical sensitivity: Advances in the methodology of ground theory.* Mill Valley, CA: Sociology Press.

Hall, G., & Loucks, S. (1977). A developmental model for determining whether the treatment is actually implemented. *American Educational Research Journal, 14*, 263–276. doi:10.3102/00028312014003263

John, D., & Freedson, P. (2012). Actigraph and actical physical activity monitors: A peek under the hood. *Medicine & Science in Sports & Exercise, 44*(Suppl. 1), S86–S89. doi:10.1249/MSS.0b013e3182399f5e

Lillehoj, C. J., Griffin, K. W., & Spoth, R. (2004). Program provider and observer ratings of school-based preventive intervention implementation: Agreement and relation to youth outcomes. *Health Education & Behavior, 31*, 242–257. doi:10.1177/1090198103260514

Lynch, S., & O'Donnell, C. (2005, April 14). *The evolving definition, measurement, and conceptualization of fidelity of implementation in scale-up of highly rated science curriculum units in diverse middle schools.* Paper presented at Fidelity of Implementation Symposium at AERA, Montreal, Canada. Retrieved from http://www.gwu.edu/~scale-up/fidelity.html

McKenzie, T. L., Catellier, D. J., Conway, T., Lytle, L. A., Grieser, M., Webber, L. A., . . . Elder, J. P. (2006). Girls' activity levels and lesson contexts in middle school PE: TAAG baseline. *Medicine & Science in Sports & Exercise, 38*, 1229–1235. doi:10.1249/01.mss.0000227307.34149.f3

Melanson, E. L., Jr., & Freedson, P. S. (1995). Validity of the computer science and applications, Inc. (CSA) activity monitor. *Medicine & Science in Sports & Exercise, 27*, 934–940.

Mowbray, C. T., Holter, M. C., Teague, G. B., & Bybee, D. (2003). Fidelity criteria: Development, measurement, and validation. *American Journal of Evaluation, 24*, 315–340. doi:10.1177/109821400302400303

National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel: Teaching children to read—An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups* (NIH Publication No. 00-4754). Retrieved from http://www.nichd.nih.gov/publications/pubs/nrp/Documents/report.pdf

No Child Left Behind (NCLB) Act of 2001, 20 U.S.C.A. § 6301 *et seq.* (West 2003).

O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K–12 curriculum intervention research. *Review of Educational Research, 78*, 33–84. doi:10.3102/0034654307313793

Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3rd ed.). Thousand Oaks, CA: Sage.

Penuel, W. R., & Means, B. (2004). Implementation variation and fidelity in an inquiry science program: An analysis of GLOBE data reporting patterns. *Journal of Research in Science Teaching, 41*, 294–315. doi:10.1002/tea.20002

Public Schools of North Carolina, State Board of Education, Department of Public Instruction. (2011). *North Carolina essential standards: Healthful living*. Retrieved from http://www.ncpublicschools.org/acre/standards/new-standards/#healthful

Ruiz-Primo, M. A. (2006). *A multi-method and multi-source approach for studying fidelity of implementation* (CSE Report 677). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.

Sallis, J. F., McKenzie, T. L., Beets, M. W., Beighle, A., Erwin, H., & Lee, S. (2012). Physical education's role in public health: Steps forward and backward over 20 years and HOPE for the future. *Research Quarterly for Exercise and Sport, 83*, 125–135. doi:10.1080/02701367.2012.10599842

Slavin, R. (2003). A reader's guide to scientifically based research: Learning how to assess the validity of education research is vital for creating effective, sustained reform. *Educational Leadership, 60*(5), 12–16.

Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology: Combining qualitative and quantitative approaches.* Thousand Oaks, CA: Sage.

Tashakkori, A., & Teddlie, C. (2010). *Handbook of mixed methods in social and behavioral research* (2nd ed.). Thousand Oaks, CA: Sage.

U.S. Department of Health and Human Services. (2010). *Strategies to improve the quality of physical education.* Retrieved from http://www.cdc.gov/healthyyouth/physicalactivity/pdf/quality_pe.pdf

Ysseldyke, J. R., Spicuzza, S., Kosciolek, S., & Boys, C. (2003). Effects of a learning information system on mathematics achievement and classroom structure. *Journal of Educational Research, 96*, 163–173. doi:10.1080/00220670309598804