Results of the 2004–05 AAPT/PTRA Rural Project Student Impact Study

Eric R. Banilower
Sherri L. Fulp

September 2005

Submitted to:

George Amann
193 Primrose Hill Road
Rhinebeck, NY 12572

Warren Hein
American Association of Physics Teachers
One Physics Ellipse
College Park, MD 20740-3845

Jan Mader
2223 5th Avenue, SW
Great Falls, MT 59404

Karen Jo Matsler
3743 Hollow Creek
Arlington, TX 76001

Jim Nelson
14722 Gainesborough Court
Orlando, FL 32826

Submitted by:

Horizon Research, Inc.
326 Cloister Court
Chapel Hill, NC 27514-2296

Results of the 2004-05 AAPT/PTRA Rural Project
Student Impact Study

Introduction

In the summer of 2004, the American Association of Physics Teachers/Physics Teaching Resource Agent (AAPT/PTRA) program conducted 25, week-long professional development institutes for teachers of physics in rural schools. The professional development was intended to increase the participants' knowledge of physics content and pedagogy and to provide participants with activities they could implement in their own classrooms, with the end goal of improving student learning of physics. The logic model underlying the program's efforts is summarized in Figure 1.

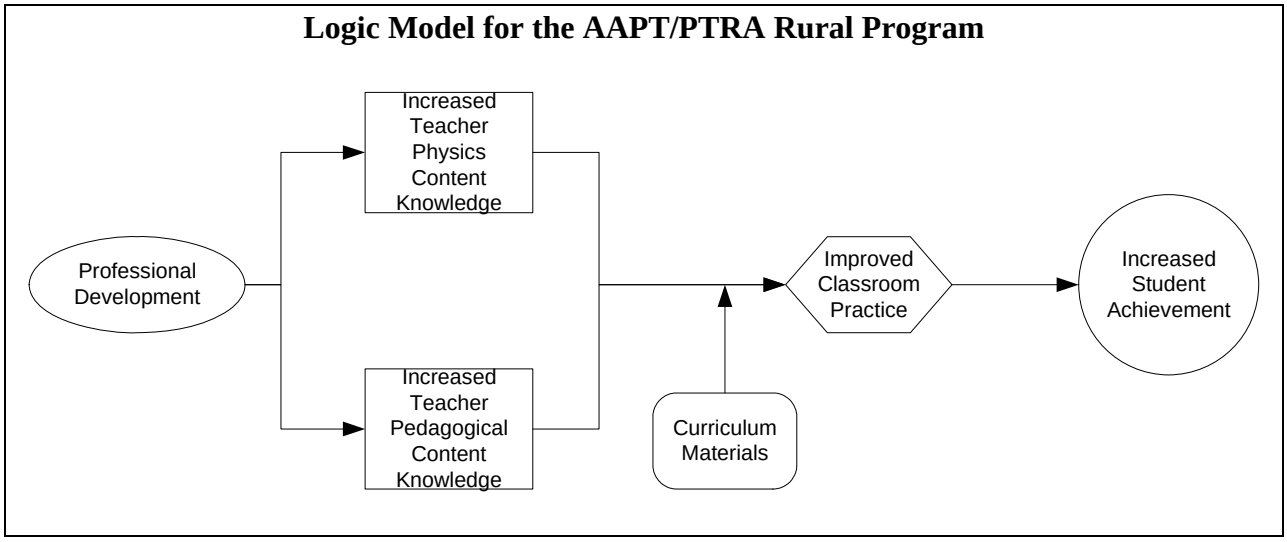

Figure 1

Horizon Research, Inc. (HRI) conducted a study of the impact of teachers' participation in 23 of these institutes on the physics content knowledge of their students. Of the 23 institutes, 14 focused on kinematics and dynamics and 9 focused on momentum and energy. (Students of teachers from the two institutes that focused on different topics were not included in this study.) This study examines the relationship between teachers' participation in PTRA professional development and student achievement. The analyses use the results of a pre-test/post-test content assessment administered to students of participating teachers at the beginning and end of their instruction in the study topics. This study is based on data from 49 teachers, including 1,589 students in 105 classes.

These analyses seek to answer the question, "Do students of teachers who participate in an AAPT/PTRA rural institute focused on a specific topic exhibit greater content knowledge after receiving instruction on that content area than students of teachers who participated in an AAPT/PTRA rural institute focused on a different area?" The study takes advantage of the two foci of the institutes to create comparison groups. In addition, this study examines whether changes in student test scores vary by gender, ethnicity, and class-level. Unless otherwise noted, only differences that are statistically significant at the 0.05 level are discussed in the text of this report.

Instrumentation

This study employed a 50-item assessment composed of selected-response items compiled primarily by the AAPT/PTRA leadership with assistance from HRI. The items were selected based upon the content goals of the rural institutes and reviewed by the PTRAs who had authored the workshop materials used in these institutes. The assessment targeted common concepts in kinematics, dynamics, impulse and momentum, and energy. A copy of the assessment can be found in Appendix A.

Scores on two scales were examined. The first scale was comprised of kinematics and dynamics items; the second scale measured concepts in momentum and energy. Each scale score is computed as the percent of items correct. Table 1 shows the number of items and reliability (Cronbach's alpha) for the assessment scales; each scale has at least an acceptable reliability on the post-test.[1] The low reliabilities on the pre-test may be a result of students guessing on many of the items due to lack of familiarity with the content.

Table 1
Assessment Scale Reliabilities

	Number of Items	Reliability (Cronbach's Alpha)	
		Pre	Post
Kinematics and Dynamics	28	0.6	0.7
Momentum and Energy	22	0.5	0.6

Race/ethnicity and gender data were also collected from students. Teachers provided information about their classes, including the class title and instructional time spent on each of the targeted topics. Classes were classified as being introductory (e.g., "Physical Science"), college preparatory (e.g., "Physics I"), or advanced (e.g., "AP Physics).

[1] Typically, a Cronbach's alpha ≥ 0.60 is considered acceptable, ≥ 0.70 is fair, ≥ 0.80 is good, and ≥ 0.90 is excellent.

The Sample

Each rural institute that focused on kinematics and dynamics or momentum and energy was asked to recruit volunteers for this study; 182 out of the roughly 450 teachers that participated in these 23 Rural Institutes volunteered to participate in the study. Because a disproportionate number of teachers volunteered from some institutes, stratified random sampling was used to ensure equal weighting of each rural institute (i.e., so the results would be representative of the AAPT/PTRA Rural Project as a whole and not biased by the institutes that recruited more teachers for the study). From the volunteers, a sample of 90 teachers was drawn. It is important to note that although random sampling was used to select the sample, the sample was drawn from those teachers volunteering for the study. Thus, the potential for selection bias in the sample was not eliminated (e.g., it may be that only those teachers that were most enthusiastic about the project volunteered), and any findings of this study may not generalize to the typical teacher participating in the project.

Of the 90 teachers who were sampled and agreed to participate in the study, only 49 returned usable data. Some teachers dropped out of the study; others provided only portions of the data needed. Although some attrition was predicted, more teachers withdrew than was expected. Thus, the statistical power of the study was lower than predicted (i.e., the probability of detecting a significant difference among the groups if a difference truly existed was smaller than expected).

Of the 49 teachers that returned complete data, 27 had participated in a 2004 kinematics and dynamics institute, 16 had participated in a 2004 energy and momentum institute and a 2003 kinematics and dynamics institute, and 6 had participated in only a 2004 momentum and energy institute. These 49 teachers administered the assessment to 105 classes, and 1,589 students completed both a pre- and post-test. Tables 2–5 provide descriptive statistics on the teachers, classes, and students participating in the study.

Table 2
Descriptive Statistics: Teacher Attendance at AAPT/PTRA Rural Institutes

	Percent of Teachers (N = 49)
Kinematics and Dynamics Only	55
Momentum and Energy Only	12
Kinematics and Dynamics and Momentum and Energy	33

Table 3
Descriptive Statistics: Class Level

	Percent of Classes (N = 105)
Introductory	29
College Preparatory	66
Advanced	6

Table 4
Descriptive Statistics: Other Class Data

	N	Min	Max	Mean	Standard Deviation
Class Size	105	2	35	17.74	6.35
Instructional Time: Kinematics and Dynamics (hundreds of minutes)	105	0	41.25	16.64	8.66
Instructional Time: Momentum and Energy (hundreds of minutes)	105	0	24.00	9.38	5.15

Table 5
Descriptive Statistics: Students

	Percent of Students (N = 1,589)
Gender	
Female	46
Male	54
Race/Ethnicity[†]	
American Indian/Alaskan Native	1
Asian	2
Black/African-American	2
Hispanic/Latino	7
Native Hawaiian/Other Pacific Islander	1
White	90

[†] The total percent adds to more than 100 as students could select more than one category.

Analysis and Results

The main research question of this study was whether teacher participation in an AAPT/PTRA rural institute on a particular topic leads to greater student achievement in that topic. A secondary question was whether students of teachers participating in two rural institutes outperformed students of teachers participating in only one institute. Thus, the following groups were examined in this study:

1. Teachers who had participated in a kinematics and dynamics institute only;
2. Teachers who had participated in momentum and energy institute only; and
3. Teachers who had participated in a kinematics and dynamics institute and a momentum and energy institute.

Descriptive statistics for the pre- and post-test scores are shown in Table 6. Overall, students tended to score higher on the post-test than on the pre-test, though scores at both administrations were rather low.

Table 6
Descriptive Statistics for Student Scores on the Assessment Scales, by Pattern of Teacher Participation in AAPT/PTRA Rural Institutes

	Kinematics and Dynamics Only		Momentum and Energy Only		Kinematics and Dynamics and Momentum and Energy	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
Kinematics and Dynamics						
Pre-test	45.85	12.73	44.25	12.36	45.95	12.53
Post-test	54.82	16.85	55.31	15.90	56.34	14.87
Momentum and Energy						
Pre-test	44.30	12.27	44.03	11.66	43.22	12.44
Post-test	53.27	15.84	50.17	16.02	54.32	14.36

The student assessment data have a nested structure, with students nested within classes, and classes nested within teachers. Statistical techniques that do not account for potential grouping effects (e.g., characteristics of a particular class of students) in nested data structures can lead to incorrect estimates of the relationship between independent factors and the outcome. Hierarchical (multilevel) regression modeling is an appropriate technique for nested data[2] and was used to examine student assessment scores. An advantage of this approach is that is allows one to appropriately model characteristics of all levels of data (i.e., characteristics of students, characteristics of classes, and characteristics of teachers).

Two models were examined, one for each outcome: kinematics/dynamics and momentum/ energy post-test scores. The main independent variable of interest was teacher participation in AAPT/PTRA rural institutes. Pre-test scores were included to control for initial status. Student gender and race/ethnicity (collapsed into two categories: white/Asian and non-Asian minority) were included to examine whether performance was consistent across different groups of students. Class size, course level (i.e., introductory, college preparatory, or advanced), and amount of instructional time on the outcome topic were included at the class level. Teacher participation in AAPT/PTRA rural institutes was included at the teacher level.

Regression coefficients and standard errors for each model are presented in Table 7; the main independent variables of interest in this study are shaded. For continuous independent variables (e.g., pre-test score), a positive regression coefficient indicates a positive correlation between the independent variable and the outcome (i.e., higher values of the independent variable are associated with higher values of the outcome) and a negative regression coefficient indicates a negative correlation (i.e., higher values of the independent variable are associated with lower values of the outcome). For categorical independent variables (e.g., female), the regression coefficient indicates the added effect of being a member of that group relative to the comparison category (e.g., a positive coefficient for the female variable indicates that females, on average, scored higher than males, a negative coefficient would mean that females tended to score lower than males). Following Table 7 is an interpretation of these regression results for each outcome.

[2] Bryk, A.S. & Raudenbush, S.W. (1992). *Hierarchical Linear Models: Applications and data analysis methods.* Newbury Park, CA: Sage Publications.

Table 7
Regression Coefficients and Standard Errors, by Outcome Scale

	Kinematics and Dynamics	Momentum and Energy
Intercept	54.75 (0.66)	52.47 (0.82)
Teacher Variables		
Institute Attended (compared to kinematics & dynamics only)		
Momentum & energy only	-1.11 (2.38)	-4.04 (2.79)
Kinematics & dynamics and momentum & energy	0.38 (1.37)	0.19 (1.77)
Class Variables		
Class size	-0.03 (0.08)	0.01 (0.09)
Instructional time on topic (hundreds of minutes)	0.08 (0.08)	0.11 (0.14)
Class level (compared to college preparatory)		
Introductory	-10.69*** (1.46)	-10.46*** (1.53)
Advanced	8.65** (2.33)	6.52* (2.47)
Student Variables		
Pre-test score	0.47*** (0.04)	0.35*** (0.03)
Female	-2.10** (0.72)	-3.03*** (0.70)
Non-Asian minority	-3.69** (1.09)	-2.04 (1.08)

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Kinematics and Dynamics

Controlling for class level, size, instructional time, student demographics, and pre-test scores, no significant differences were found among students on the kinematics and dynamics outcome by the pattern of their teacher's participation in AAPT/PTRA rural institutes. Figure 2 shows the adjusted mean post-test scores for students in each group (means are adjusted for student demographics and pre-test scores and represent the "typical" student across all of the teachers' classes).

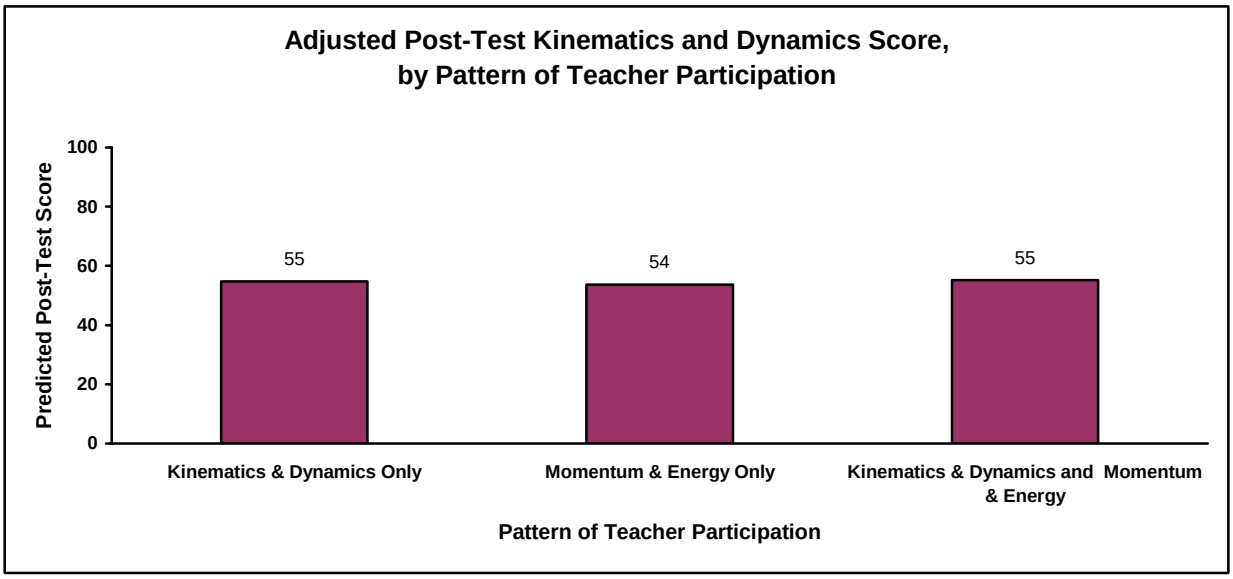

Figure 2

In addition, the amount of instructional time on kinematics and dynamics was not related to student scores. This finding could be interpreted in a number of ways. One interpretation could be that the assessment items were not well aligned with teachers' objectives for their courses. Another interpretation could be that the items were aligned, but not sensitive to instruction (in other words, the items may not have been very good). A third interpretation could be that only a little instruction is needed to improve student performance in these areas, and that there are diminishing returns for further instruction. Alternatively, it could be that the teachers' instruction was just not very effective.

Significant differences in assessment results were detected between certain sub-groups of students. Females tended to score lower on this scale than did males by about 2 points; non-Asian minority students scored lower than whites/Asians by about 4 points. Students in introductory classes scored about 11 points lower and students in advanced classes scored about 9 points higher than students in college preparatory classes. These differences in student scores were not related to teacher participation in AAPT/PTRA rural institute (i.e., teacher participation did not widen or narrow these achievement gaps).

Momentum and Energy

The results for the momentum and energy scale are similar to those for the kinematics and dynamics scale. Controlling for class level, size, instructional time, student demographics, and pre-test scores, no significant differences were found among students on the momentum and energy outcome by their teacher's pattern of participation in AAPT/PTRA rural institutes. As was the case with the kinematics and dynamics outcome, the amount of instructional time on momentum and energy was not related to student scores. Figure 3 shows the adjusted mean post-test scores for students in each group.

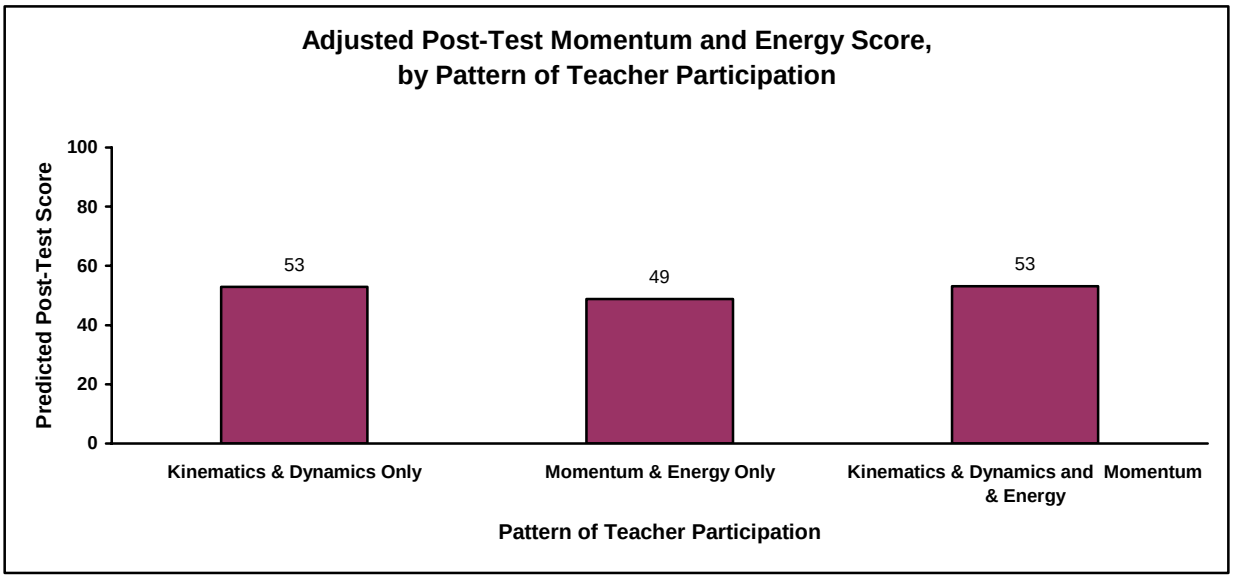

Figure 3

There was no significant difference on this outcome among students by race/ethnicity, though a gender gap was detected. Males tended to score slightly higher (about three points) on the momentum and energy scale than females. Teacher participation in an AAPT/PTRA rural institute was not related to this achievement gap. Students in introductory classes scored about 10 points lower and students in advanced classes scored about 7 points higher than students in college preparatory classes.

Table 8 shows individual item statistics, by pattern of teacher participation in AAPT/PTRA rural institutes, for both the pre- and post-tests. Although the differences between the percentage of students answering individual items correctly on the pre- and post-tests were not tested statistically,[3] these data may be informative for the project leadership in identifying which areas of outreach participants' teaching are relative strengths and which are relative weaknesses. However, these data should be examined with caution as differences (or lack thereof) in individual items may not be reliable and may be partially due to the quality of the assessment item rather than impact of AAPT/PTRA rural institutes.

[3] Using the typical convention of statistical significance at the 0.05 level, there is a 5 percent chance of a statistical test yielding a false positive (i.e., indicating that a difference is statistically significant when it really is not). Statistically testing 50 items individually would likely result in multiple false positives. Although techniques exist for controlling this error rate when examining multiple outcomes, their use decreases the statistical power of a test (i.e., the ability of a test to detect a difference that really exists), making it more likely that actual differences will not be identified as statistically significant. Thus, HRI restricted the use of statistical tests to the scale scores, which are more reliable estimates of knowledge than are individual items.

Table 8
Proportion of Students Responding Correctly, by Pattern of Teacher Participation

Item	Scale	Percent of Students								
		Kinematics & Dynamics Only			Momentum & Energy Only			Kinematics & Dynamics and Momentum & Energy		
		Pre-Test	Post-Test	Difference	Pre-Test	Post-Test	Difference	Pre-Test	Post-Test	Difference
4	Kinematics	0.32	0.34	0.02	0.29	0.35	0.07	0.37	0.39	0.03
12	Kinematics	0.69	0.74	0.04	0.54	0.75	0.21	0.67	0.76	0.09
13	Kinematics	0.33	0.49	0.16	0.23	0.55	0.32	0.38	0.46	0.09
16	Kinematics	0.19	0.29	0.10	0.18	0.28	0.10	0.17	0.30	0.13
25	Kinematics	0.59	0.70	0.10	0.62	0.66	0.04	0.55	0.66	0.11
26	Kinematics	0.88	0.91	0.02	0.86	0.90	0.04	0.89	0.92	0.04
27	Kinematics	0.92	0.92	0.00	0.92	0.89	-0.04	0.90	0.94	0.04
28	Kinematics	0.94	0.94	0.00	0.96	0.95	-0.01	0.93	0.94	0.01
31	Kinematics	0.34	0.42	0.07	0.30	0.34	0.05	0.37	0.49	0.11
32	Kinematics	0.34	0.40	0.05	0.37	0.37	0.00	0.38	0.48	0.11
33	Kinematics	0.59	0.66	0.06	0.62	0.72	0.10	0.59	0.69	0.10
34	Kinematics	0.31	0.38	0.07	0.30	0.41	0.11	0.36	0.44	0.08
38	Kinematics	0.38	0.47	0.10	0.36	0.54	0.18	0.38	0.54	0.15
39	Kinematics	0.72	0.74	0.02	0.64	0.74	0.10	0.68	0.74	0.07
40	Kinematics	0.41	0.50	0.09	0.36	0.51	0.15	0.40	0.53	0.12
41	Kinematics	0.74	0.78	0.03	0.77	0.80	0.03	0.73	0.84	0.11
3	Dynamics	0.29	0.54	0.25	0.48	0.58	0.10	0.33	0.58	0.25
5	Dynamics	0.39	0.75	0.36	0.29	0.56	0.28	0.34	0.73	0.39
6	Dynamics	0.53	0.65	0.12	0.57	0.69	0.11	0.54	0.64	0.10
10	Dynamics	0.59	0.80	0.21	0.52	0.78	0.26	0.61	0.84	0.23
18	Dynamics	0.40	0.45	0.05	0.40	0.40	0.00	0.36	0.44	0.08
20	Dynamics	0.24	0.31	0.06	0.21	0.42	0.21	0.21	0.34	0.12
23	Dynamics	0.32	0.34	0.02	0.32	0.37	0.05	0.29	0.36	0.06
35	Dynamics	0.18	0.20	0.02	0.19	0.23	0.04	0.22	0.20	-0.02
37	Dynamics	0.55	0.61	0.06	0.56	0.61	0.05	0.54	0.62	0.08
47	Dynamics	0.37	0.53	0.16	0.38	0.51	0.13	0.40	0.54	0.13
48	Dynamics	0.10	0.14	0.04	0.08	0.15	0.08	0.10	0.10	0.00
50	Dynamics	0.17	0.37	0.21	0.08	0.41	0.33	0.19	0.29	0.10
7	Momentum	0.22	0.37	0.15	0.20	0.35	0.15	0.21	0.35	0.14
8	Momentum	0.70	0.74	0.04	0.70	0.71	0.02	0.69	0.82	0.12
11	Momentum	0.13	0.18	0.05	0.15	0.10	-0.05	0.13	0.16	0.03
14	Momentum	0.64	0.76	0.12	0.61	0.70	0.09	0.63	0.72	0.10
15	Momentum	0.09	0.12	0.03	0.05	0.05	0.00	0.08	0.15	0.08
19	Momentum	0.73	0.69	-0.04	0.74	0.56	-0.18	0.74	0.74	0.00
22	Momentum	0.42	0.41	-0.01	0.42	0.34	-0.08	0.41	0.45	0.04
24	Momentum	0.36	0.47	0.11	0.36	0.50	0.13	0.38	0.50	0.12
46	Momentum	0.46	0.60	0.14	0.51	0.53	0.02	0.44	0.57	0.12
49	Momentum	0.12	0.14	0.03	0.14	0.13	-0.01	0.12	0.13	0.01
1	Energy	0.82	0.88	0.06	0.75	0.90	0.15	0.79	0.91	0.11
2	Energy	0.85	0.88	0.03	0.88	0.92	0.05	0.82	0.89	0.07
9	Energy	0.09	0.50	0.41	0.10	0.48	0.38	0.12	0.43	0.32
17	Energy	0.77	0.79	0.02	0.71	0.81	0.10	0.76	0.85	0.09
21	Energy	0.56	0.53	-0.02	0.67	0.54	-0.12	0.55	0.59	0.03
29	Energy	0.62	0.72	0.10	0.52	0.70	0.18	0.59	0.76	0.18
30	Energy	0.37	0.58	0.21	0.36	0.57	0.21	0.39	0.61	0.22
36	Energy	0.29	0.53	0.23	0.30	0.46	0.16	0.27	0.53	0.26
42	Energy	0.51	0.63	0.11	0.49	0.59	0.10	0.49	0.63	0.14
43	Energy	0.56	0.65	0.08	0.58	0.61	0.03	0.52	0.66	0.14
44	Energy	0.21	0.26	0.04	0.24	0.22	-0.02	0.19	0.20	0.01
45	Energy	0.23	0.31	0.07	0.21	0.25	0.04	0.19	0.29	0.11

Summary and Conclusions

This study utilized a pre-test/post-test comparison group design to examine the impact of the teachers' participation in the AAPT/PTRA rural institutes on their students' physics content knowledge in two areas: (1) kinematics and dynamics and (2) momentum and energy. Teachers were classified into one of three categories: participation in a kinematics and dynamics institute only, participation in a momentum and energy institute only, or participation in both institutes. The analyses controlled for pre-test scores, student gender, student race/ethnicity (categorized as white/Asian or non-Asian minority), amount of instruction on the topic, class level (introductory, college preparatory, or advanced), and class size.

The results of this study do not provide evidence that the AAPT/PTRA rural institutes have had an impact on student achievement. There were no significant differences in student scores among the three groups of teachers on the kinematics and dynamics outcome or the momentum and energy outcome. It is important to note that although this study does not provide evidence of the project's impact on student achievement, the limitations of the study prohibit concluding that the project is not having such an impact. In other words, this study should not be seen as providing definitive evidence for either conclusion.

For both outcomes, the amount of instructional time devoted to the topic was not related to student scores. This finding, along with the overall results could be interpreted in a number of ways. One interpretation could be that the assessment items are not well aligned with teachers' objectives for their courses. Another interpretation could be that the items are aligned, but not sensitive to instruction. A third interpretation could be that only a little instruction is needed to improve student performance in these areas, and that there are diminishing returns for further instruction. Alternatively, it could be that the teachers' instruction was just not very effective (the average kinematics and dynamics pre-test score was about 43, the post was about 53; the average momentum and energy pre-test score was about 41, the post was about 48).

Additionally, the analyses examined if performance on the assessment was consistent across different types of students and classes. Gender was a significant predictor of achievement in both content areas, with males scoring slightly higher than females on the post-tests (controlling for pre-test score). Non-Asian minority students scored lower than white/Asian students on the kinematics and dynamics scale, but not on the momentum and energy scale. Not surprisingly, students in introductory classes scored lower, and students in advanced classes scored higher, than students in college preparatory classes on both scales.

In regards to the overall study findings, a number of interpretations as to the lack of measurable impact on student content knowledge are plausible. First, because of the smaller than expected number of teachers returning complete data, the probability of the study finding a significant difference, if one really existed, was rather modest. A second possibility is that even though the assessment was aligned with the AAPT/PTRA institute topics, it was not well aligned with teachers' instructional goals. Another possible explanation is that although the institutes are having a positive impact on teachers' understanding of the content (as evidenced by the teacher

impact study[4]), the institutes are having less of an impact on teachers' understanding of student learning (the common misconceptions, why people have them, and how to confront the misconceptions) and/or their ability to provide high quality instruction.

In order for the AAPT/PTRA program to maximize its potential for impacting student learning, the project leadership may want to consider ways it can better support teachers as they attempt to implement what they have learned. For example, teachers may need more assistance in identifying how/where the AAPT/PTRA activities fit into their curriculum or how to obtain the equipment needed to implement the activities. Teachers may also need more time and support during the institutes and at follow-up meetings to practice implementing the newly acquired activities and pedagogies.

[4] Banilower, E.R., & Fulp, S.L. (2005). Results of the 2004 AAPT/PTRA Rural Institute Teacher Impact Study. Chapel Hill, NC: Horizon Research, Inc.

Appendix A

AAPT/PTRA Rural Institute Student Assessment