Results of the 2004 AAPT/PTRA Rural Institute Teacher Impact Study

Eric R. Banilower
Sherri L. Fulp

January 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 AAPT/PTRA Teacher 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.

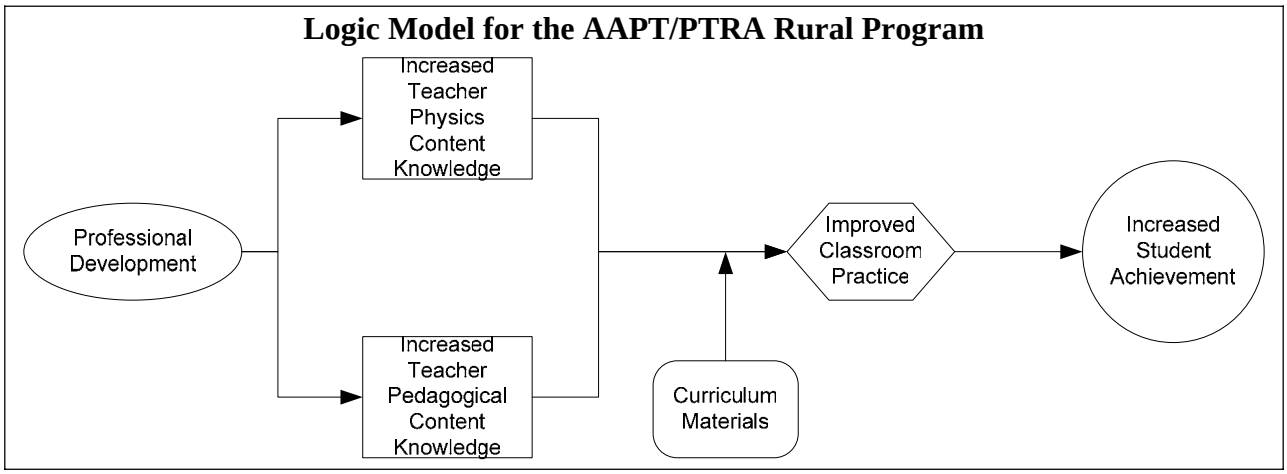

Figure 1

Horizon Research, Inc. (HRI) conducted a study of the impact of 23 of these institutes on participating teachers' content knowledge: the 14 that focused on kinematics and dynamics and the 9 that focused on momentum and energy. (The two institutes that focused on different topics were not included in this study.) Note that this study examines one of the first links in the logic model—the relationship between PTRA professional development and teacher physics content knowledge—using the results of a content assessment administered at the beginning and end of each of the 23 institutes included in this study. Specifically, these analyses seek to answer the question, "Do teachers who participate in an AAPT/PTRA rural institute focused on a specific topic exhibit greater content knowledge after participating in an AAPT/PTRA rural institute on that content area than teachers participating 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 teacher test scores vary by teacher gender and grade-level taught. 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 54-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.

The assessment yields four scale scores: kinematics, dynamics, 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[1]; each scale has at least an acceptable reliability,[2] indicating that the items within each set are well correlated with each other and appear to be measuring the same construct (e.g., kinematics knowledge). In addition, teacher demographic data from a questionnaire completed by the participants at the beginning of the rural institutes were also used in this study.

Table 1
Assessment Scale Reliabilities

	Number of Items	Reliability (Cronbach's Alpha)	
		Pre	Post
Kinematics	12	0.69	0.64
Dynamics	14	0.75	0.75
Momentum	13	0.81	0.81
Energy	14	0.72	0.73

The Sample

The assessment was administered at the beginning and end of each of the 23 rural institutes involved in this study. HRI received pre- and post-test responses from 460 rural institute participants. There was some missing data on the rural outreach questionnaire (the result of participants skipping items); HRI had complete data for these analyses from 424 participants. Table 2 shows the demographic characteristics of these participants. Just over half of the participants were male; nearly all were white. The participants had a wide range of teaching experience, with roughly half having taught for more than 10 years. Eighty percent of the participants taught at the high school level; 20 percent taught in the middle or elementary grades.

The AAPT/PTRA program is intended to serve teachers over a three year time period, with the first year typically focusing on kinematics and dynamics, the second on momentum and energy,

[1] One item, Q52, was dropped from all analyses as it was redundant with an earlier item on the assessment.

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

and the third on either electricity or waves and sound. However, not all teachers are able to participate in all three years, and the project attempts to fill their slots with new teachers. As a result, some teachers (26 percent) in this study had participated in both a kinematics and dynamics institute in 2003 and a momentum and energy institute in 2004.[3] Others teachers were participating in an AAPT/PTRA rural institute for the first time; 59 percent were participating in a kinematics and dynamics institute and 16 percent were participating in a momentum and energy institute. Because participation in a kinematics and dynamics institute in a previous year is expected to affect scores on those two scales, three groups of teachers (based upon their pattern of participation) are compared in those analyses.

Table 2
Demographics of Participants with Complete Data

	Percent of Participants (N = 424)
Gender	
Female	49
Male	51
Race/Ethnicity	
American Indian/Alaskan Native	1
Asian	0
Black or African-American	3
Hispanic or Latino	0
Native Hawaiian or Other Pacific Islander	0
White	95
Prior Teaching Experience	
0–2 Years	14
3–5 Years	17
6–10 Years	22
11–15 Years	18
16–20 Years	12
21–25 Years	7
26 or more Years	11
Grade Level Taught	
Elementary School	3
Middle School	17
High School	80
Participation in RPTRA Institutes	
2004 Kinematics and Dynamics Only	59
2004 Momentum and Energy Only	16
2004 Momentum and Energy and 2003 Kinematics and Dynamics	26

In terms of the number of participants per institute, the smallest institute provided data from 7 teachers, the largest from 46 (see Table 3). The average institute size was 20 participants.

[3] An analysis of project records indicates that nearly 70 percent of the 2003 rural institute participants returned for a 2004 institute. The 26 percent refers to the proportion of participants in this study that attended institutes in both years.

Table 3
Institute Size

Minimum	Maximum	Mean	Standard Deviation
7.00	46.00	20.00	7.93

Analysis and Results

Because a substantial proportion of the momentum and energy institute participants had attended a kinematics and dynamics institute in 2003, treating all 2004 momentum and energy participants as a comparison group for the 2004 kinematics and dynamics participants could lead to an underestimation of the impact of the project. Thus, for the kinematics and dynamics scales, three groups are examined:

1. Teachers who had participated only in a 2004 institute on kinematics and dynamics;
2. Teachers who had participated in a 2003 institute on kinematics and dynamics and a 2004 institute on momentum and energy; and
3. Teachers who had participated only in a 2004 institute on momentum and energy.

As no participants in this study had participated in a 2003 momentum and energy institute, scores on the momentum and energy scales are examined for only two groups: participants in the 2004 momentum and energy institutes and participants in the 2004 kinematics and dynamics institutes.

Descriptive statistics for the pre- and post-test scores are shown in Tables 4 and 5. The lower mean scores for the momentum and energy scales indicated that, overall, the items on these scales were more challenging than the items on the kinematics and dynamics scales. On all four scales, teachers appear to have higher scores on the post-test than on the pre-test, regardless of institute type.

Table 4
Descriptive Statistics for the Kinematics and Dynamics Scales, by Pattern of Participation

	2004 Kinematics and Dynamics Only (N = 249)		2003 Kinematics and Dynamics and 2004 Momentum and Energy (N = 109)		2004 Momentum and Energy Only (N = 66)	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
Pre-Test						
Kinematics	72.42	20.56	81.04	16.85	69.44	20.53
Dynamics	73.69	20.48	77.98	16.18	63.74	20.18
Post-Test						
Kinematics	79.69	17.30	84.25	14.22	69.82	20.87
Dynamics	78.17	18.58	82.11	15.17	66.67	22.35

Table 5
Descriptive Statistics for the Momentum and Energy Scales, by Institute Type

	2004 Kinematics and Dynamics Institute (N = 249)		2004 Momentum and Energy Institute (N = 175)	
	Mean	Standard Deviation	Mean	Standard Deviation
Pre-Test				
Momentum	59.59	25.15	56.57	25.29
Energy	63.77	21.03	60.61	20.57
Post-Test				
Momentum	62.96	25.39	64.70	23.36
Energy	66.32	21.15	70.61	19.66

The teacher assessment data have a nested structure, with teachers nested within rural institutes. Statistical techniques that do not account for potential grouping effects (e.g., participants in one rural institute all had the same workshop experience, while participants in another rural institute all shared a somewhat different workshop experience) 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 apportioning and predicting variance within and across groups in a nested data structure[4] and was used to examine teachers' assessment scores. An advantage of this approach is that is allows one to appropriately model characteristics of both levels of data (i.e., characteristics of teachers and characteristics of rural institutes).

Four models were examined, one for each outcome: kinematics, dynamics, momentum, and energy post-test scores. The main independent variable of interest was whether the teacher participated in an institute focused on the topic of the outcome (either in 2003 or 2004 for the kinematics and dynamics scales). Pre-test scores were included to control for initial status and teacher gender and grade level taught[5] were included to examine whether performance was consistent across different types of participants. Finally, the number of participants at each institute was included in the model to determine whether institute size had an effect on teacher knowledge gains.

Regression coefficients and standard errors for each model are presented in Table 6 (the main independent variables of interest 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

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

[5] Because of the small number of elementary teachers in the sample, teachers were categorized as either "high school" or "elementary/middle school."

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 6 is an interpretation of these regression results for each outcome.

Table 6
Regression Coefficients and Standard Errors, by Outcome Scale

	Kinematics	Dynamics	Momentum	Energy
Intercept	72.96	73.29	63.82	65.71
	(2.20)	(1.58)	(0.64)	(0.85)
Institute Variables				
Institute size	-0.00	-0.10	0.02	-0.01
	(0.13)	(0.08)	(0.05)	(0.07)
Teacher Variables				
Pre-test score	0.59*	0.70*	0.82*	0.73*
	(0.03)	(0.03)	(0.03)	(0.03)
Female	-1.76	-2.46*	0.31	1.39
	(1.27)	(1.16)	(0.99)	(1.18)
High school teacher	0.33	2.22	6.44*	8.69*
	(1.67)	(1.53)	(1.32)	(1.59)

* $p < 0.05$.

Kinematics

Controlling for pre-test score and demographics, teachers who had participated in a 2004 institute on kinematics and dynamics scored, on average, about eight points higher on the post-test kinematics scale than teachers who never participated in an institute on these topics (an effect size[6] of 0.43 standard deviations). In terms of number of items, this difference translates to about one additional kinematics item correct on the 12-item scale.

Similarly, teachers who participated in a 2003 institute on kinematics and dynamics scored over seven points higher on the post-test than teachers who had never participated in an AAPT/PTRA rural institute on these topics (an effect size of 0.42 standard deviations), suggesting that the knowledge gained in the institute is retained over time. This difference equates to about one additional item correct on the 12-item scale. Finally, there were no significant differences in post-test kinematics scores by teacher gender, grade level taught, or institute size.

Figure 2 shows the predicted post-test scores on the kinematics scale (i.e., expected scores based upon the regression equation) for the three groups of participants.

[6] Effect sizes of about 0.20 are typically considered small, 0.50 medium, and 0.80 large. Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences.* Hillsdale, NJ: Lawrence Erlbaum Associates.

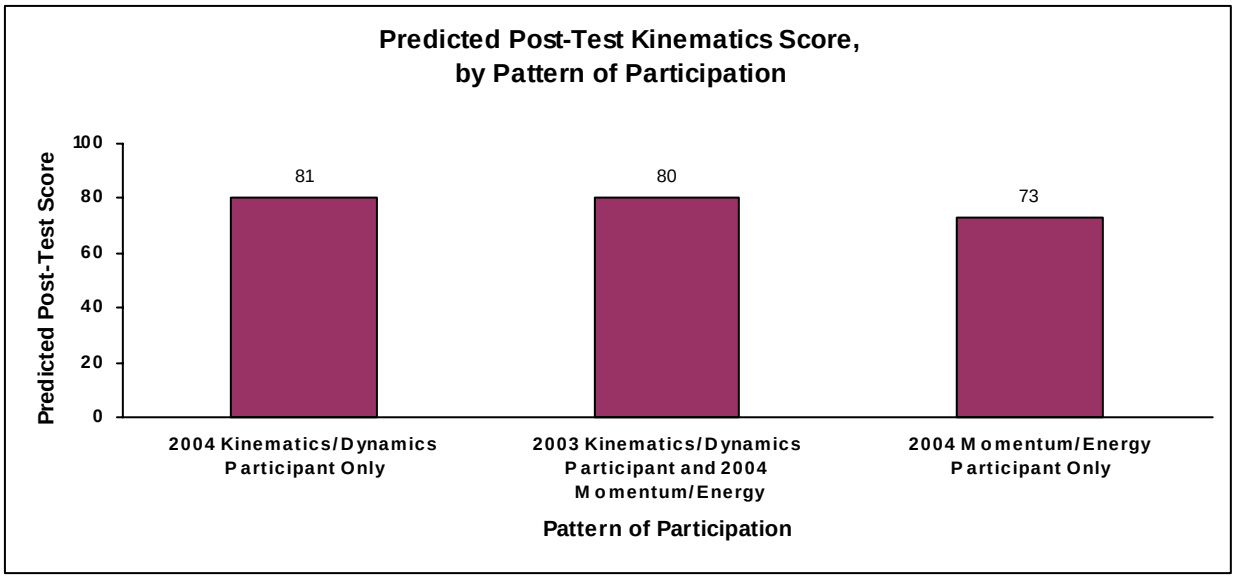

Figure 2

Dynamics

The results for the dynamics scale follow the same pattern as for the kinematics scale. Controlling for pre-test score and demographics, teachers participating in a 2004 institute focused on kinematics and dynamics scored, on average, almost five points higher on the post-test dynamics scale than teachers who had never participated in an AAPT/PTRA rural institute on these topics (an effect size of 0.25 standard deviations). In terms of number of items, this difference translates to an average of two-thirds of an additional dynamics item correct on the 14-item scale.

As was the case on the kinematics scale, teachers who participated in a 2003 institute on kinematics and dynamics scored over six points higher on the post-test than teachers who had never participated in an institute on these topics (and effect size of 0.32 standard deviations). This difference translates to a little less than one additional dynamics item correct on the scale. Figure 3 shows the predicted post-test scores on the dynamics scale for the three groups.

Males tended to score slightly higher (about two points, or about one-third of an item on average) on the dynamics scale than females. This difference in performance did not vary across the three groups of participants. No differences were found for grade level taught or institute size.

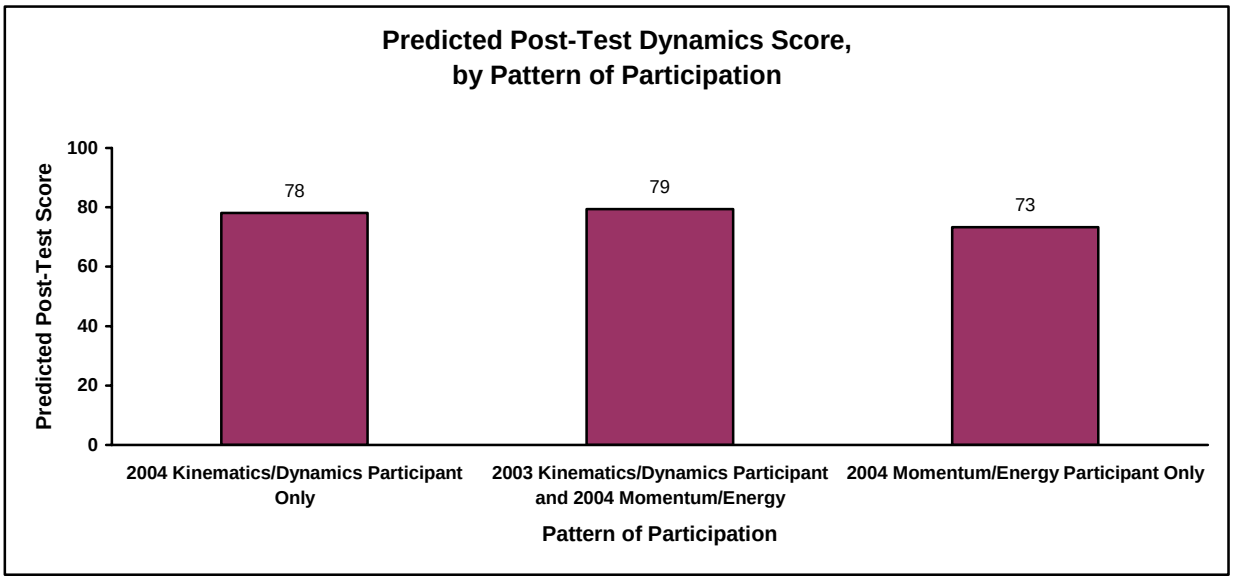

Figure 3

Momentum

Controlling for pre-test score and demographics, teachers participating in a 2004 institute focused on momentum and energy scored, on average, about five points higher on the post-test momentum scale than teachers who did not (an effect size of 0.21 standard deviations). In terms of number of items, this difference translates to an average of two-thirds of an additional momentum item correct on the 13-item momentum scale. Figure 4 shows the predicted post-test scores on the dynamics scale.

High school teachers tended to score over six points higher (nearly one item on average) on the momentum scale than elementary and middle school teachers. This difference in performance between high school and elementary/middle school teachers did not vary by institute type. No differences were found for gender or institute size.

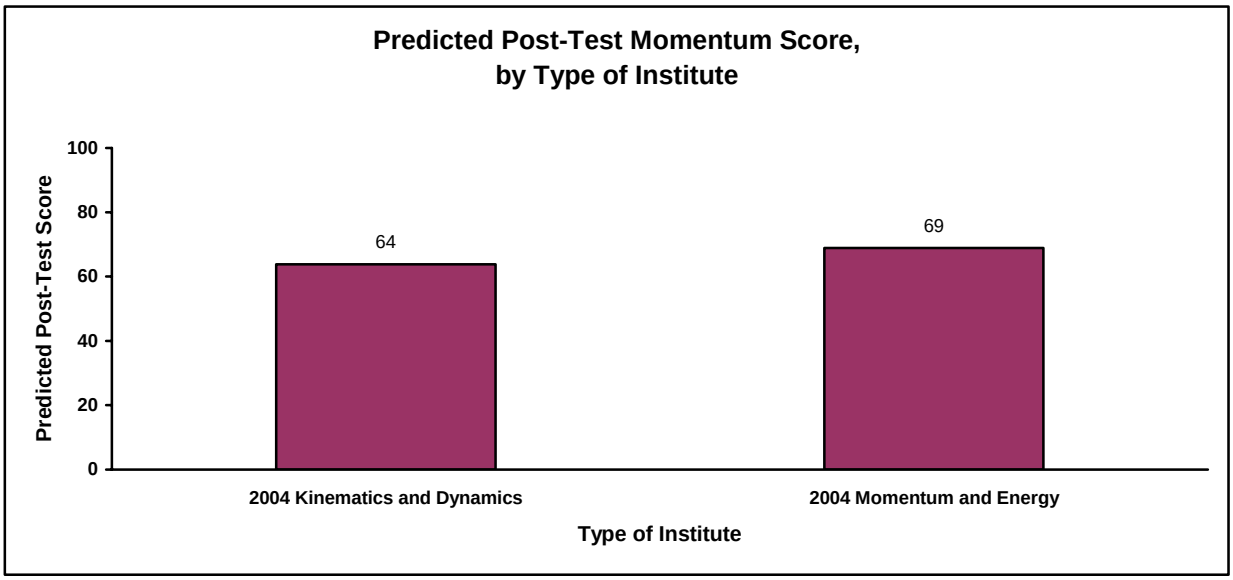

Figure 4

Energy

Controlling for pre-test score and demographics, teachers participating in a 2004 momentum and energy institute scored, on average, almost six points higher on the post-test energy scale than teachers participating in a 2004 institute on kinematics and dynamics (an effect size of 0.28 standard deviations). In terms of number of items, this difference translates to an average of just under one additional energy item correct on the 14-item scale. Figure 5 shows the predicted post-test scores on the energy scale.

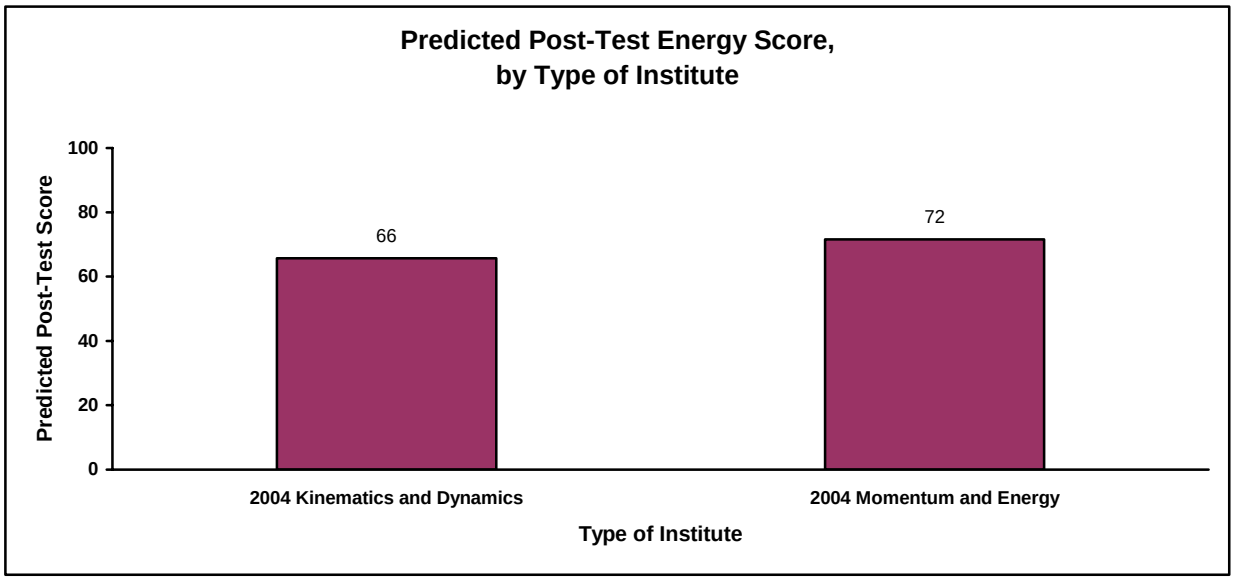

Figure 5

As was the case with the momentum scale, high school teachers outperformed elementary and middle school teachers; on the energy scale the difference was nearly nine points (an average of more than one item). This difference in performance on the energy scale was found in both institute types. No differences were found for gender or institute size.

Tables 7 and 8 show individual item statistics, by content area, for both the pre- and post-tests. (The percentage of participants selecting each response option is shown in Appendix B, by institute type, for both the pre- and post-tests.) Although the differences between the percentage of participants answering individual items correctly on the pre- and post-tests were not tested statistically,[7] these data may be informative for the project leadership in identifying which concepts are being successfully addressed in the rural institutes and which ones are not and the prevailing misunderstandings that remain after instruction. In addition, the data may also be useful in identifying areas in which participants were fairly well prepared prior to the institute. 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 the AAPT/PTRA rural institutes.

[7] 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 52 items individually would 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 7
Outreach Participants Responding Correctly to Kinematics and Dynamics Items, by Participant Type

Item	Scale	Percent of Outreach Participants								
		2004 Kinematics and Dynamics Only			2003 Kinematics and Dynamics and 2004 Momentum and Energy			2004 Momentum and Energy Only		
		Pre-Test	Post-Test	Difference	Pre-Test	Post-Test	Difference	Pre-Test	Post-Test	Difference
1	Kinematics	46	62	16	61	65	5	44	52	8
6	Kinematics	78	84	6	91	94	3	79	70	-9
7	Kinematics	85	84	-2	88	89	1	82	77	-5
12	Kinematics	76	72	-4	83	79	-5	71	65	-6
18	Kinematics	90	94	4	89	88	-1	82	85	3
25	Kinematics	72	82	10	78	80	2	68	73	5
27	Kinematics	59	75	16	74	83	8	62	59	-3
41	Kinematics	81	89	8	89	92	3	80	79	-2
43	Kinematics	56	67	12	70	76	6	44	52	8
48	Kinematics	61	69	8	72	80	8	56	58	2
51	Kinematics	90	95	5	94	94	1	91	92	2
54	Kinematics	76	84	8	84	92	7	74	77	3
2	Dynamics	96	93	-4	94	96	2	95	92	-3
10	Dynamics	39	43	4	40	45	5	24	23	-2
14	Dynamics	79	85	6	83	83	0	65	56	-9
15	Dynamics	67	76	10	70	81	11	50	56	6
20	Dynamics	56	65	8	61	70	8	38	47	9
21	Dynamics	94	93	-1	95	94	-1	91	94	3
23	Dynamics	86	93	6	93	95	3	79	86	8
29	Dynamics	74	78	4	72	80	7	64	73	9
30	Dynamics	79	88	9	89	91	2	68	77	9
31	Dynamics	95	94	-1	96	96	0	91	91	0
34	Dynamics	75	78	2	79	83	4	61	58	-3
35	Dynamics	58	61	2	72	72	0	44	50	6
36	Dynamics	79	86	7	88	91	3	85	83	-2
47	Dynamics	54	63	9	57	72	15	38	47	9

Table 8
Outreach Participants Responding Correctly to Momentum and Energy Items, by Institute Type

Item	Scale	Percent of Outreach Participants					
		2004 Kinematics and Dynamics Institute			2004 Momentum and Energy Institute		
		Pre-Test	Post-Test	Difference	Pre-Test	Post-Test	Difference
3	Momentum	71	78	8	71	83	11
9	Momentum	59	61	2	52	67	15
13	Momentum	79	79	0	74	82	9
19	Momentum	87	88	1	83	86	3
24	Momentum	59	64	5	56	69	13
33	Momentum	35	43	7	34	45	11
38	Momentum	34	36	2	25	37	12
39	Momentum	68	72	4	69	81	13
42	Momentum	69	70	1	67	75	8
44	Momentum	59	62	3	61	63	2
49	Momentum	74	77	2	74	79	5
50	Momentum	56	60	4	54	50	-4
53	Momentum	24	29	4	17	24	7
4	Energy	66	67	1	59	65	6
5	Energy	69	71	2	65	81	16
8	Energy	95	94	0	94	97	3
11	Energy	31	39	8	23	42	19
16	Energy	43	48	5	44	56	12
17	Energy	83	79	-4	75	76	1
22	Energy	77	78	1	73	79	6
26	Energy	54	62	8	49	57	7
28	Energy	80	82	2	77	88	11
32	Energy	81	84	3	79	92	13
37	Energy	36	40	4	38	51	13
40	Energy	43	46	2	42	52	10
45	Energy	61	65	4	65	74	10
46	Energy	74	73	0	66	78	13

Summary and Recommendations

This study utilized a pre-test, post-test comparison group design to examine the impact of the AAPT/PTRA rural institutes on teachers' physics content knowledge in four areas: kinematics, dynamics, momentum, and energy. For kinematics and dynamics, assessment scores for three groups of participants were compared those who had participated in a 2004 institute on kinematics and dynamics, those who had participated in a 2003 institute on these topics, and those who had never participated in an AAPT/PTRA rural institute on these topics. For momentum and energy, assessment scores were compared for those who participated in an AAPT/PTRA rural 2004 momentum and energy workshop and those that did not.

The results of this study provide evidence that the AAPT/PTRA rural institutes have had a positive impact on teachers' physics content knowledge. On each of the four test scales (kinematics, dynamics, momentum, and energy), controlling for pre-test scores and demographics, participants who had taken part in an institute on that topic scored significantly higher than participants who had not. On the kinematics and dynamics scales, those that had participated in an AAPT/PTRA rural kinematics/dynamics institute, either this year or last year, scored higher on the post-test than those who have not participated in an AAPT/PTRA rural kinematics/dynamics institute. Similarly, on the momentum and energy scales, those that had participated in an AAPT/PTRA rural momentum/energy institute performed better than those who had not.

Additionally, the analyses examined if performance on the assessment was consistent across different types of participants. For 3 of the 4 scales, gender was not a significant factor; the exception being on the dynamics scale, where males scored slightly higher on the post-test (controlling for pre-tests score) than females. Grade level taught was a significant factor on both the momentum and energy scales; even after controlling for pre-test scores, high school teachers scored higher on the post-test than elementary and middle school teachers. No differences were found for grade level taught on the kinematics or dynamics scales. Finally, the number of participants in an institute did not significantly impact participants' knowledge gains.

These data provide the project an opportunity to reflect on its efforts and to make adjustments for future rural institutes. To assist the project in this process, HRI offers the following recommendations:

- Even though differences were statistically significant for each of the four scales, teachers receiving professional development on the topic of a scale tended to outperform the comparison teachers by one item or less on a 12- to 14-item scale. In addition, the post-test scores for participants at the momentum/energy and kinematics/ dynamics institutes averaged in the 60 and 70 percent range, respectively. The project leadership may want to consider whether gains of this magnitude, and these levels of mastery on this assessment, meet their goals for the rural institutes. In particular, the leadership may want to consider whether the activities in the institutes are of sufficient quality to achieve the project's content goals. It may be that some activities are overly complex, or require a great deal of technology or equipment

construction, causing the participants to focus on doing the activity/the technology/building the materials rather than the important physics ideas.

- In considering the results of this study, the project leadership may want to examine the individual test item statistics and the patterns of incorrect responses on the post-test. These statistics may help the project leadership identify which concepts the rural institutes are successfully addressing as well as areas in which the rural institutes could be improved.

- The project leadership may want to consider possible causes and potential solutions for the differences in scores on the momentum and energy scales between high school teachers and elementary/middle school teachers. Three possible causes are:

  - *The nature of the content*
    Do the concepts in the momentum/energy institutes require a stronger background in science and/or mathematics to understand than the concepts in the kinematics/dynamics institutes (which high school science teachers are more likely to possess than elementary/middle school teachers)? If so, what additional steps can the project take to help ensure that all participants are successful?

  - *The grade level at which each content topic is typically taught*
    Are the concepts in momentum and energy covered in elementary/middle school? If not, are teachers at these grade levels likely to be sufficiently invested in learning these concepts? How can the project better meet the needs of teachers of different grade levels?

  - *How the content is presented*
    Do the different rural institutes provide equivalent opportunities for all participants to access the targeted concepts? Are ideas in momentum and energy treated more quantitatively and ideas in kinematics and dynamics treated more conceptually?

- Although the gender differences were much less pronounced this year than last year (being a significant factor on only the dynamics scale this year), the project leadership may want to continue to emphasize the need for PTRAs to be sensitive to equity issues in their workshops, to help ensure that all participants have opportunities to master the targeted physics concepts.

Appendix A

2004 AAPT/PTRA Teacher Assessment