Student Outcomes in Physical Science: Exploring the Impact of Newton’s Universe

Materials in the Absence of the Distance Training Course

Bradshaw, T. J., Cunningham, J.D., Osborn, J., & Bradley, K.D.

University of Kentucky
Impact of NU Kit

Abstract

Newton’s Universe (NU) is a unique professional development program utilizing hands-on inquiry based distance learning for Appalachian middle school physical science teachers. This study explores the impact on student learning, if any, of teachers receiving NU kits and no subsequent NU training (kit group) in comparison to receiving the complete NU course and kits (course group) as well as to receiving neither (control group). In an effort to control for the impact of varying teacher knowledge, a Rasch model was fit to calibrate teacher post assessment scores to identify teachers with similar ability for comparison groups. Teachers administered baseline and post assessments on temperature and heat to their middle school science students. Students from the kit group resulted in significantly higher scores on average than students from the course and control groups. Future studies will examine this premise with a larger sample of rural middle school teachers and students.
Student Outcomes in Physical Science: Exploring the Impact of Newton’s Universe

In physical science classrooms across the nation, over half of teachers do not acquire an academic major or minor in any of the physical sciences (Ingersoll, 1999). A lack of content knowledge promotes the practice of fostering common misconceptions or the avoidance of teaching the content altogether (Tilgner, 1990; Wenner, 1993). Not only does the lack of physical content knowledge of teachers negatively impact students, but additionally it has been suggested that teacher perceptions of physical science concepts show striking similarities to K-8 student perceptions (Driver, Guesne, & Tiberghien, 1985; Osborne & Freyberg, 1985; Schoon, 1995; Wandersee, Mintzes, & Novak, 1994). The limited understanding of physical science concepts is particularly troublesome for middle school science teachers in rural Appalachian regions (Atwood, Christopher, & McNall, 2005), where sustained professional development is quite difficult due to limited resources and geographic isolation ultimately hindering the recruitment and retention of highly qualified teachers (Collins, 1999; McClure & Reeves, 2004).

Newton’s Universe (NU), a unique professional development program for rural middle school physical science teachers seeks to resolve these problems. NU utilizes a hands-on inquiry based distance learning format to overcome barriers associated with sustaining professional development in rural Appalachian regions. Teachers have the opportunity to improve their understanding of physical science through exploration of topics on their own time and at their own pace. Upon completion of the online training course, teachers are provided with a NU kit containing laboratory materials with suggestions for inquiry-based labs.
Purpose

Considering the teacher training course and the NU supplemental classroom kit was provided to every teacher in the project, the impact of solely the teacher training course on student learning is confounded by the effect of the NU kit. Previous studies have investigated the impact of the use of kits in the absence of professional development compared to the effect of the combination on student achievement (Young & Lee, 2005). According to Young and Lee (2005), no significant difference existed in the science achievement of students whose teachers had sustained professional development and used the kits and students whose teachers used the kits in the absence of sustained professional development. However, in previous years, both groups of students had been exposed to teachers who may or may not have received sustained professional development in kit usage and reform pedagogy. As Young and Lee note, it is possible that students’ prior exposure to a teacher who received sustained professional development negates the affect of having a teacher who has not participated in extensive professional development (2005). Thus, in order to determine the impact of NU kits on student achievement independent of sustained professional development, students whose teachers do not receive NU professional development should not have prior exposure to teachers who have received NU professional development. This study will explore the impact on student learning, if any, of teachers receiving NU temperature, heat, & energy (THE) kits and no subsequent professional development (kit group), teachers receiving NU THE kits and professional development (course group), and teachers receiving neither the NU THE
kits nor professional development (control group). Neither the kit group nor the control
group participants come from course group school districts.

The research questions are:

1. How can teachers be chosen to represent each comparison group of students
   where teacher ability is taken into account?

2. Do students whose teacher receives only the NU kit perform significantly
different on the temperature and heat assessment in comparison to students whose
   teacher receives NU kits and the NU professional development course?

3. In what ways can these findings be utilized to guide hypotheses for the
   subsequent, large scale study?

First, the course and control teachers are chosen with similar ability as the teacher
representing the kit group, based on a Rasch calibration of raw teacher post assessment
scores. Descriptive statistics provide a general illustration of their student assessment
scores, separating students by whether their teacher received simply the NU kit (kit
group) or NU training and a kit (course group). An analysis of variance is employed to
determine whether a significant difference in gains can be seen between the kit, course,
and control groups. Results will then be used to develop hypotheses for a larger,
subsequent study.

Literature Review

Heat and temperature concepts are closely related to daily life and are the
foundation for the natural sciences of physics, chemistry, and biology (Koh & Paik,
2002). An overwhelming number of middle school students develop misconceptions
about physical science, particularly when discussing temperature and heat concepts, one
of the main focuses of the Newton’s Universe project (Sozbilir, 2003). The development
of these intuitive misconceptions from everyday experiences is particularly resistant to change without effective instruction of these concepts (Driver, 1989; Osborne & Freyberg, 1985).

The lack of proper K-12 physical science instruction is problematic, as research shows that the rigor of physics courses in high school impacts success in college physics courses (Sadler and Tai, 2001). Likewise, Gifford and Harpole (1986) found that taking high school physics, in addition to receiving high grades in math and spending time in the laboratory, were associated with high grades in college physics. The impact of students succeeding in their college physics courses are far-reaching, as physics is likely to be a pre-requisite for engineering, medicine, and related fields. Therefore, a well-developed physical science knowledge foundation from middle and high school can have large impacts on a students’ future.

Aforementioned, a lack of adequate preparation in teacher science content understanding encourages the practice of fostering common misconceptions or the avoidance of teaching certain content areas altogether (Tilgner, 1990; Wenner, 1993). One explanation, therefore, for middle school student misconceptions of temperature and heat concepts is insufficient teacher preparation in these areas. The National Survey of Science and Mathematics Education found that while 77% of K-5 teachers felt well qualified to teach reading/language arts, only 14% felt well qualified to teach physical science (Horizon Research Inc., 2001). Evidence suggests a similar trend for middle school teachers with low percentages responding as well qualified to teach modern physics (10%), light and sound concepts (23%), and force and motion concepts (29%) were also low (Horizon Research Inc., 2001). A professional development project
conducted by Long, Teates, & Zweifel (1992) indicates that background knowledge in major areas of physics needed to be addressed before specialized topics could be discussed, due to a lack of formal physics instruction. A great need emerges from the research for attention to the instruction of physics concepts at the teacher and student levels.

Methods

Participants

The sample consists of sixty-four middle school physical science teachers recruited on a voluntary basis (47 course group; 17 control group) from public schools in rural Kentucky and Virginia in 2006 and 2007 (see Table 1). As a participant in the course group, teachers completed the NU teacher training course and received NU kits and stipends, while control group teachers received only stipends. In addition to these participants, one 9th grade physical science teacher was recruited from an urban school district in Kentucky. This teacher received the NU kit but no professional development (kit group). All groups of teachers were asked to complete surveys and assessments on a voluntary and confidential basis.

Instrumentation

The teacher assessment instrument was developed to determine cognitive gains in conceptual understanding of temperature and heat topics covered in the NU teacher training course. The research team constructed and piloted 25 multiple choice items with four answer choices to measure teacher conceptual understanding of temperature and heat. A Rasch analysis was conducted on the pilot data to guide any necessary revisions to the instrument prior to administration (see Bradley & Sampson, 2006 for a complete
description). The same instrument was used as a baseline and post measure, but only teacher post assessment scores are used in this study to calibrate teacher ability estimates. Administration of the course and control teachers’ post assessments occurred during the course follow-up workshops in August 2006 and August 2007. Administration of kit teacher’s post assessment took place in the teacher’s classroom in September 2007.

Student Instrumentation

To measure student learning of temperature and heat concepts, a student assessment with 39 multiple-choice items was constructed. The items on the assessment were categorized into five content domains: foundations, properties of matter, energy transfer, phase change and thermal energy. A Rasch analysis, similar to Bradley and Sampson (2006), was conducted to revise and refine the instrument prior to administration. Participating teachers taught subsets of the temperature, heat and energy concepts covered in their training; therefore, students were only assessed on the topics their teachers reported as having taught in the classroom. Similar to the teacher instrument, the same student instrument was used as the baseline and post instruction assessment.

Data Collection

Administration of the course and control teachers’ post assessments occurred during the course follow-up workshops in August 2006 (cohort 1 of course group) and August 2007 (cohort 2 of course group and control). Administration of kit teacher’s post assessment took place in the teacher’s classroom in September 2007. The course and control groups’ baseline student assessments (N=1239 and N=193, respectively) occurred before the temperature and heat unit was taught in the fall
semester, followed by the post instruction assessment within two weeks after the unit was taught. Administration of the student assessments (N=134) under the supervision of the kit group teacher followed the same pattern during 2007-2008.

Each teacher received a unique identification number to use when completing all surveys and assessments. Records linking identification number to teachers were kept in a secure location separate from all hard copy and electronic data. All student responses were catalogued by random identification numbers, only identifying students by their teacher, and stored in a secure file. Records linking identification numbers to students were only kept securely by each teacher; therefore, students remain anonymous to researchers in the NU project.

Data Analysis

All teachers, including the kit teacher, were given the multiple-choice assessment developed to measure understanding of temperature and heat covered in the NU teacher training course. All data were stored in an Excel file with unique teacher identification numbers. Minitab statistical software was used for descriptive statistics to provide a general picture of teacher assessment scores, highlighting the kit teacher. A dichotomous Rasch model was utilized to calibrate the raw teacher post assessment scores. Course and control teachers with similar ability measures to the kit teacher are determined using this Rasch calibration. A teacher to represent each group, course and control group, was chosen based on this analysis.

Secondly, descriptive statistics were used to produce a general illustration of student assessment scores, separating students by whether their teacher received simply the NU kit (kit group), NU training and a kit (course group), or neither (control group).
An analysis of variance is employed to determine whether a significant difference in student scores can be seen across groups. To control for the variable of teacher subject matter knowledge, only students having teachers of similar ability estimates are compared in this study.

**Results**

First, in order to choose course and control group teachers with similar understanding of temperature and heat, descriptive statistics for post assessment scores are provided followed by the Rasch calibration of teacher ability estimates. For a general picture of teacher assessment raw score totals, descriptive statistics are provided (see Table 1). The raw score of the kit teacher is italicized for ease of comparison to the descriptive statistics for the group at its entirety and separated by cohorts (Cohort 1 & 2 = course group; cohort 3 = control group). The kit teacher scored more than one standard deviation above the mean for the entire teacher group; however when separated into cohorts, this teacher scored within one standard deviation above the mean for Cohort 1. A graphical illustration of the teacher post assessment raw scores highlighting the teacher receiving the kit alone versus the other three cohorts of teachers is included (see Figure 1).

In an effort to control for the impact of varying teacher knowledge, a dichotomous Rasch model was fit to calibrate teacher post assessment scores to identify teachers with similar ability measures for comparison groups. Prior to discussing these calibrations, attention to the fit of the data to the Rasch model must be visited to investigate stability and accuracy of the estimates. First, teacher reliability and separation estimates were 0.70 and 1.51. These estimates were deemed acceptable for fit of data to the Rasch model. Fit
Impact of NU Kit 11

statistics indicate how well the data fit the expectations of the Rasch model. The expectation is that a more able person should have a higher probability of getting any item correct than a less able person (regardless of item difficulty), and a less difficult item should have a higher probability of being answered correctly than a more difficult item (regardless of the ability of the person responding to the item). An accepted cutoff of mean square fit statistics within one standard deviation of the average infit and outfit (Wright & Stone, 2004), is used to determine person fit in this analysis. Of the 65 teachers measured, six teachers resulted in high fit statistics; therefore, these teachers were not chosen to represent a comparison group considering the stability accuracy of the teacher ability measures could not be ensured.

Three treatment groups are considered for comparison: (a) a teacher receiving a kit and professional development (Cohort 1 and 2 – course group); (b) a teacher receiving a kit only (kit group); and (c) a control teacher (Cohort 3). Teachers were chosen with similar ability estimates from the post assessment calibration under the condition that the teacher was not highlighted as resulting in high fit statistics (see Figure 2). The teacher estimated with the same ability for the post assessment as the teacher receiving the kit (Measure=1.73) only was highlighted as a misfitting person; therefore, the teacher estimated with the next closest ability estimate (Measure=1.45) for the post assessment was chosen for comparison. A control teacher from cohort 3 was chosen with an ability estimate of 0.96.

The variation in the student scores for the teacher receiving only the kit was much greater than that of the other two teacher groups (see Figure 3). The teacher receiving only the kit had the greatest increase in the student average from baseline to post
assessment (see Table 2). Consequently, the kit group of students resulted in a significantly greater mean ($F = 11.40; p$-value $= 0.000$) on the post assessment compared to students of the teacher receiving the kit and professional development and students of the control teacher.

Discussion

To accurately depict the impacts of the NU course on middle school physical science student learning of temperature and heat concepts, it is essential to independently determine the impact of the NU kit in the absence of the NU professional development course. This study compared the scores of students whose teachers received the NU kit with no training course (kit group), the NU kit with the NU training course (course group), and neither (control group). A Rasch calibration of teacher post assessment scores was used in choosing teachers to represent each comparison group in an attempt to control for the confounding variable of teacher knowledge for the comparison of student assessment scores across the groups. Based on the results from this study, the conclusion was that students whose teacher received only NU kit without NU professional development scored significantly higher on average compared to students whose teachers received the NU kit and the NU training course as well students whose teachers received neither the kit nor the NU professional development course.

Caution must be taken when interpreting these results for several reasons. First, it is important to note the kit group teacher was estimated with the highest ability of the three groups. The argument could be made that the better student performance of the kit only group was attributed to more teacher subject knowledge. For this reason, it is problematic that only one teacher represents each comparison group even though each
teacher may have an ample number of students for comparison. Now with regards to the student groups, students in the NU kit only group were 9th graders from an urban Kentucky school while the other groups were comprised of 6th, 7th, and 8th graders from rural Kentucky and Virginia schools. The rebuttal for these findings could be that 9th graders have experienced more exposure to these concepts than middle school students and thus are more likely to improve after review than middle school students who may be learning temperature and heat for the first time. However, the 9th grade students did not score on the baseline assessment significantly higher on average than middle school students, indicating that baseline knowledge was similar between the two groups.

Although this study results in inconclusive findings on the impact of the NU kit in the absence of the NU training course on middle school student learning of temperature and heat, the research questions still remain valid and essential to NU and other professional development courses. With the same type of research design employed, the questions guiding this study will be examined with middle school teachers from rural Kentucky regions in the near future. Implications of findings from this on-going study will identify whether time and effort should be expended on providing professional development along with kits to NU participants. Furthermore, findings will encourage other professional development to explore the impact of their training independent of variables potentially sharing an impact on teacher and student learning.
References


Teaching and Learning (pp. 177-210). New York: MacMillan.


Impact of NU Kit

Appendix

Table 1

*Descriptive statistics for post assessment raw scores*

<table>
<thead>
<tr>
<th>N</th>
<th>Mean</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>All teachers</td>
<td>65</td>
<td>15.338</td>
</tr>
<tr>
<td>Cohort 1</td>
<td>34</td>
<td>16.971</td>
</tr>
<tr>
<td>Cohort 2</td>
<td>13</td>
<td>14.154</td>
</tr>
<tr>
<td>Cohort 3</td>
<td>17</td>
<td>12.706</td>
</tr>
<tr>
<td>Kit teacher</td>
<td>1</td>
<td>20.000</td>
</tr>
</tbody>
</table>

*Figure 1. Dotplot of post assessment raw score totals separated by cohort*
Figure 2. Dotplot of calibrated teacher performance estimates separated by cohort

![Dotplot of Calibrated Teacher Performance Estimates](image)

Figure 3. Boxplot of student raw scores separated by cohort

![Boxplot of Student Raw Scores](image)
Table 2

Descriptive statistics of baseline and post student assessment scores by cohort

<table>
<thead>
<tr>
<th>Cohort</th>
<th>Test</th>
<th>N</th>
<th>Mean</th>
<th>St. Dev.</th>
<th>Min</th>
<th>Median</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kit</td>
<td>1</td>
<td>121</td>
<td>15.678</td>
<td>5.020</td>
<td>5</td>
<td>15.0</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>119</td>
<td>19.101</td>
<td>7.371</td>
<td>6</td>
<td>18.0</td>
<td>36</td>
</tr>
<tr>
<td>Kit/PD</td>
<td>1</td>
<td>93</td>
<td>15.892</td>
<td>4.322</td>
<td>6</td>
<td>16.0</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>90</td>
<td>16.222</td>
<td>4.820</td>
<td>4</td>
<td>16.0</td>
<td>31</td>
</tr>
<tr>
<td>Control</td>
<td>1</td>
<td>87</td>
<td>13.149</td>
<td>3.429</td>
<td>5</td>
<td>13.0</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>84</td>
<td>15.476</td>
<td>3.892</td>
<td>7</td>
<td>15.5</td>
<td>24</td>
</tr>
</tbody>
</table>