THE RELATION OF PHYSICS TEACHERS’ CONTENT KNOWLEDGE FOR TEACHING ENERGY AND TEACHING PRACTICE AS MEASURED BY THE QUALITY AND DEMAND OF THE ASSIGNMENTS AND ASSESSMENTS THEY DESIGN AND SELECT FOR INSTRUCTION

By Robert Zisk

A dissertation submitted to the
School of Graduate Studies
Rutgers, The State University of New Jersey
In partial fulfillment of the requirements
For the degree of
Doctor of Philosophy
Graduate Program in Education
Written under the direction of Eugenia Etkina
And approved by

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

________________________________________________________________________

New Brunswick, New Jersey
October 2021
ABSTRACT OF THE DISSERTATION

The Relation of Physics Teachers’ Content Knowledge for Teaching Energy and Teaching Practice as Measured by the Quality and Demand of the Assignments and Assessments they Design and Select for Instruction

by Robert Zisk

Dissertation Director: Eugenia Etkina

The construct of Content Knowledge for Teaching (CKT) specifies that teachers possess knowledge that is specific to the content area that they teach, that enables them to carry out tasks of teaching to help their students learn. As such, a teacher’s CKT should be related to all aspects of their practice. This dissertation investigates the relationship between high school physics teachers’ content knowledge for teaching energy (CKT-E) and the quality and demand of the assignments and assessments they design and select to use during their energy unit.

The first paper in this dissertation explores the relationship between teachers’ CKT-E and the quality and demand of the assignments and assessments they use during instruction. This paper describes the development of a protocol to assess the quality and demand of assignments and assessments in physics and lays out the validity argument for using scores produced by that protocol to measure quality and demand. Additionally, this paper provides evidence that teachers with more robust CKT-E use assignments and assessments that are of higher quality and demand than those with less robust CKT-E.
The second paper in this dissertation focuses specifically on the learning targets that physics teachers address in their instructional goals and on their end of unit assessment, and the relationship of the learning targets addressed to their CKT-E. This paper provides evidence that while all teachers tend to address the same breadth of learning targets regardless of CKT-E, teachers with more robust CKT-E tend to address learning targets at a deeper level than those with less robust CKT-E.

The final paper in this dissertation describes four case studies that further investigate the relationship between CKT-E and classroom practice. This paper explores the coherence between teachers’ CKT-E, goals for instruction, classroom practice, assignment and assessments and student outcomes. This paper supports the findings of the first two papers in this dissertation by describing how there is generally coherence between CKT-E and all aspects of teaching practice and explains any discrepancies through teacher’s orientation towards teaching and teacher’s robust curricular knowledge in lieu of strong CKT-E.
Dedication

This work is dedicated to my best friend, partner, and teammate. Thank you for always supporting me, inspiring me every day to finish and pushing me to reach for, and lasso, the moon.
# Table of Contents

**Abstract of the Dissertation** ........................................................................................................ ii

**Dedication** ...................................................................................................................................... iv

**List of Tables** ................................................................................................................................. vii

**List of Figures** .............................................................................................................................. ix

**Chapter 1: Introduction** ............................................................................................................... 1

References ........................................................................................................................................ 5

**Chapter 2: The Relationship of High School Teachers’ CKT-E to the Quality and Demand of the Assignments and Assessments they Design and Select for their Energy Unit** ........................................................................................................ 7

Introduction ....................................................................................................................................... 7

**Theoretical Framework** .................................................................................................................. 11

Content Knowledge for Teaching ...................................................................................................... 11
Content Knowledge for Teaching Energy Framework (CKT-E) .......................................................... 13
CKT-E and the tasks of teaching energy ............................................................................................. 17
The cognitive demand of assignments and assessments ..................................................................... 22

Designing a protocol to measure the demand and quality of assignments and assessments in physics ........................................................................................................................................ 27

Protocol Design .................................................................................................................................. 28

Developing an argument for the validity of using artifacts to measure the enactment of CKT ......................................................................................................................................................... 38

**Methods** ....................................................................................................................................... 42

Participants ......................................................................................................................................... 42
Collection of Artifacts .......................................................................................................................... 43
CKT-E Assessment .............................................................................................................................. 44
Raters .................................................................................................................................................. 44
Scoring Procedure ............................................................................................................................... 45

Evaluation of the validity of using artifacts to make inferences regarding a teacher’s CKT ........ 49

The Scoring Inference ........................................................................................................................ 49
The Generalization Inference .............................................................................................................. 55
The Extrapolation Inference ............................................................................................................... 57

Discussion ......................................................................................................................................... 60

References ......................................................................................................................................... 67

**Chapter 3: The Relationship of Teachers’ Content Knowledge for Teaching Energy and the Learning Targets for Instruction in Physics** ......................................................................................... 72

Introduction ....................................................................................................................................... 72

Content Knowledge for Teaching ...................................................................................................... 76
Content Knowledge for Teaching Energy Framework (CKT-E) ......................................................... 78
CKT-E and the “work” of teaching energy .......................................................................................... 82
Student Learning Targets in Times of the NGSS .............................................................................. 85
Capturing Information Regarding Teachers’ Goals and Unit Assessments ...................................... 87
## Methods

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants</td>
<td>91</td>
</tr>
<tr>
<td>CKT-E Assessment</td>
<td>91</td>
</tr>
<tr>
<td>Interviews</td>
<td>92</td>
</tr>
<tr>
<td>Assessment Artifacts</td>
<td>92</td>
</tr>
<tr>
<td>Coding of Interviews and Assessments</td>
<td>93</td>
</tr>
<tr>
<td>Determining the Breadth and Depth of the Learning Targets Within Teachers’ Goals and Unit Assessments</td>
<td>97</td>
</tr>
<tr>
<td>Determining the Alignment Between Teachers’ Goals and Their Assessments</td>
<td>98</td>
</tr>
</tbody>
</table>

## Findings

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>CKT-E Scores</td>
<td>99</td>
</tr>
<tr>
<td>What is the Relationship Between Teachers’ Stated Goals and CKT-E?</td>
<td>99</td>
</tr>
<tr>
<td>What is the Relationship Between the Targets Addressed by Teachers’ Assessments and CKT-E?</td>
<td>105</td>
</tr>
<tr>
<td>What is the Alignment Between Teachers’ Instructional Goals and Their Unit Assessments?</td>
<td>107</td>
</tr>
</tbody>
</table>

## Discussion

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Targets addressed in Teachers’ Goals for Instruction</td>
<td>109</td>
</tr>
<tr>
<td>Alignment</td>
<td>112</td>
</tr>
<tr>
<td>What Do These Findings Tell Us about the Construct of CKT?</td>
<td>113</td>
</tr>
<tr>
<td>Implications for Practice</td>
<td>114</td>
</tr>
<tr>
<td>Future Work and Limitations</td>
<td>115</td>
</tr>
</tbody>
</table>

## References

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chapter 4: Investigating patterns in a relationship among Content Knowledge for Teaching Energy multiple measures of teaching quality and student learning</td>
<td>122</td>
</tr>
</tbody>
</table>

## Chapter 4: Investigating patterns in a relationship among Content Knowledge for Teaching Energy multiple measures of teaching quality and student learning

### Introduction

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>CKT and Its Relationship to Teaching Quality</td>
<td>126</td>
</tr>
<tr>
<td>Content Knowledge for Teaching Energy Framework (CKT-E)</td>
<td>127</td>
</tr>
<tr>
<td>The Current Study</td>
<td>130</td>
</tr>
</tbody>
</table>

## Methods

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study Participants</td>
<td>131</td>
</tr>
<tr>
<td>Data Sources</td>
<td>131</td>
</tr>
<tr>
<td>Case Selection</td>
<td>141</td>
</tr>
</tbody>
</table>

## Cases

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher A</td>
<td>145</td>
</tr>
<tr>
<td>Teacher B</td>
<td>156</td>
</tr>
<tr>
<td>Teacher C</td>
<td>171</td>
</tr>
<tr>
<td>Teacher D</td>
<td>181</td>
</tr>
</tbody>
</table>

## Discussion

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Implications</td>
<td>197</td>
</tr>
<tr>
<td>Limitations</td>
<td>199</td>
</tr>
</tbody>
</table>

## References

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chapter 5: Summary</td>
<td>204</td>
</tr>
</tbody>
</table>
## List of Tables

### Chapter 2

<table>
<thead>
<tr>
<th>Table Number</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Scored Dimension Reliability – Assessments</td>
<td>50</td>
</tr>
<tr>
<td>2</td>
<td>Scored Dimension Reliability – Instructional Artifacts</td>
<td>51</td>
</tr>
<tr>
<td>3</td>
<td>Final Assessment – Context, Descriptive Statistics</td>
<td>52</td>
</tr>
<tr>
<td>4</td>
<td>Instructional artifacts – Context, Descriptive Statistics</td>
<td>52</td>
</tr>
<tr>
<td>5</td>
<td>Content Addressed – Descriptive Statistics, Number of Targets Addressed by the Final Assessments and Instructional Artifacts</td>
<td>52</td>
</tr>
<tr>
<td>6</td>
<td>Scored Dimensions - Descriptive Statistics for the Final Assessments</td>
<td>53</td>
</tr>
<tr>
<td>7</td>
<td>Scored Dimensions - Descriptive Statistics for the Instructional Artifacts</td>
<td>53</td>
</tr>
<tr>
<td>8</td>
<td>Correlation Between scored Dimension scores – Final Assessment</td>
<td>54</td>
</tr>
<tr>
<td>9</td>
<td>Correlation Between Scored Dimension Scores – Instructional Artifacts</td>
<td>55</td>
</tr>
<tr>
<td>10</td>
<td>Variance Components – Final Assessment (% Variance Accounted for)</td>
<td>56</td>
</tr>
<tr>
<td>11</td>
<td>Variance Components – Instructional Artifacts (% Variance Accounted for)</td>
<td>57</td>
</tr>
<tr>
<td>12</td>
<td>Relationship Between the Content Addressed by Final Assessments and Instructional Artifacts and CKT-E Assessment Scores</td>
<td>59</td>
</tr>
<tr>
<td>13</td>
<td>Relationship Between Artifact Scored Dimension Scores and CKT-E Assessment Scores</td>
<td>59</td>
</tr>
</tbody>
</table>
Chapter 3

<table>
<thead>
<tr>
<th>Table Number</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Breadth and Depth of Teachers’ Unit Goals by Number of Target Categories and Individual Targets and Pearson Correlation Between Breadth and Depth of Content and CKT-E Score (Interview)</td>
<td>104</td>
</tr>
<tr>
<td>2</td>
<td>Number of Teachers Who Addressed Each Target Category in Their Unit Goals (Interview)</td>
<td>106</td>
</tr>
<tr>
<td>3</td>
<td>Point-biserial Correlation Between CKT-E Score and Teacher-stated Goals Related to Each Target Category (n=28)</td>
<td>108</td>
</tr>
<tr>
<td>4</td>
<td>Number of Categories and Targets Addressed and Addressed In-depth and Pearson Correlation between CKT-E and Breadth and Depth of Content (Assessment)</td>
<td>110</td>
</tr>
<tr>
<td>5</td>
<td>Number of Teachers Per Number of Discrepancies: Targets Addressed In-depth</td>
<td>112</td>
</tr>
</tbody>
</table>

Chapter 4

<table>
<thead>
<tr>
<th>Table Number</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Case Study Teacher Demographic Information</td>
<td>147</td>
</tr>
<tr>
<td>2</td>
<td>Case Study Teachers’ Ranks on CKT-E Assessments, Measures of Practice, and Student Assessments</td>
<td>148</td>
</tr>
</tbody>
</table>
## List of Figures

### Chapter 2

<table>
<thead>
<tr>
<th>Figure Number</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CKT-E Framework</td>
<td>14</td>
</tr>
<tr>
<td>2</td>
<td>Levels of Cognitive Demand with Examples</td>
<td>25</td>
</tr>
<tr>
<td>3</td>
<td>Analysis Dimension with Examples</td>
<td>33</td>
</tr>
<tr>
<td>4</td>
<td>Mathematics Dimension with Examples</td>
<td>34</td>
</tr>
<tr>
<td>5</td>
<td>Representations Dimension with Examples</td>
<td>35</td>
</tr>
<tr>
<td>6</td>
<td>Context Dimension with Examples</td>
<td>37</td>
</tr>
<tr>
<td>7</td>
<td>Kane’s (2006) Inferences and Examples</td>
<td>40</td>
</tr>
<tr>
<td>8</td>
<td>Assessment Item Level Scoring Process</td>
<td>46</td>
</tr>
<tr>
<td>9</td>
<td>Assessment Test Level Scoring Process</td>
<td>47</td>
</tr>
<tr>
<td>10</td>
<td>Instructional Artifact Scoring Process</td>
<td>49</td>
</tr>
</tbody>
</table>

### Chapter 3

<table>
<thead>
<tr>
<th>Figure Number</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CKT-E Framework</td>
<td>79</td>
</tr>
<tr>
<td>2</td>
<td>Student Energy Targets</td>
<td>81</td>
</tr>
<tr>
<td>3</td>
<td>Alignment Between Goals and Assessment Content</td>
<td>98</td>
</tr>
</tbody>
</table>
## Chapter 4

<table>
<thead>
<tr>
<th>Figure Number</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Sample CKT-E Item Concerning Systems</td>
<td>139</td>
</tr>
<tr>
<td>2</td>
<td>Sample CKT-E Item Concerning Elastic Energy</td>
<td>140</td>
</tr>
<tr>
<td>3</td>
<td>Sample Pre-Assessment Items</td>
<td>141</td>
</tr>
<tr>
<td>4</td>
<td>Sample Energy Post-Assessment Items</td>
<td>142</td>
</tr>
<tr>
<td>5</td>
<td>Dimensions of Artifact Protocol</td>
<td>143</td>
</tr>
<tr>
<td>6</td>
<td>Dimensions of Observation Protocol</td>
<td>145</td>
</tr>
<tr>
<td>7</td>
<td>Sample Multiple-Choice Assessment Items – Teacher A</td>
<td>156</td>
</tr>
<tr>
<td>8</td>
<td>Sample Open-Ended Assessment Item – Teacher A</td>
<td>157</td>
</tr>
<tr>
<td>9</td>
<td>Sample Assignment Item – Teacher A</td>
<td>158</td>
</tr>
<tr>
<td>10</td>
<td>Sample Open-Ended Assessment Item – Teacher B</td>
<td>169</td>
</tr>
<tr>
<td>11</td>
<td>Sample Open-Ended Assessment item, Conserved Quantity – Teacher B</td>
<td>170</td>
</tr>
<tr>
<td>12</td>
<td>Sample Open-Ended Assessment Item, Conceptual Question – Teacher B</td>
<td>171</td>
</tr>
<tr>
<td>13</td>
<td>Sample Assessment Items – Teacher C</td>
<td>182</td>
</tr>
<tr>
<td>14</td>
<td>Sample Multiple-Choice Assessment Items – Teacher D</td>
<td>189</td>
</tr>
<tr>
<td>15</td>
<td>Sample Multiple-Choice Assessment Items, Comparing Values – Teacher D</td>
<td>190</td>
</tr>
<tr>
<td>16</td>
<td>Sample Open-Ended Assessment Item – Teacher D</td>
<td>191</td>
</tr>
</tbody>
</table>
Chapter 1: Introduction

Teaching and creating opportunities for students to learn involves the coordination of many tasks, including understanding students’ ideas, designing learning experiences to help students construct their own understanding of science concepts, assessing students’ knowledge and skills, and using the results of these assessments to design future learning experiences. To carry out such tasks, teachers must have specialized knowledge that enables them to engage in the act of teaching. Understanding this knowledge has been the subject of continuous research. This work has led to practice-based conceptions of teacher knowledge that specify that teachers must have knowledge of the subject they are teaching, as well as the knowledge of how to help students learn that specific subject. In defining this knowledge, Shulman (1986) developed the construct of Pedagogical Content Knowledge which describes teacher knowledge as a combination subject-matter knowledge and knowledge of how to make the subject comprehensible to others. More recently, Ball and colleagues (2008) built on the work of Shulman and devised the construct of Content Knowledge for Teaching (CKT). The construct of CKT is built on the premise that teachers have a body of content-based knowledge that allow them to carry out critical tasks of teaching with the goal of helping students learn. This knowledge is distinct from the knowledge of the same subject of other experts who do not teach.

As CKT specifies that teachers have a body of knowledge that enables them to carry out the tasks of teaching, a teacher’s CKT should be related to the quality of their instruction. Studies to investigate this link have generally focused on studying instructional quality using classroom observations and data on student achievement.
(Baumert et al., 2010; Carlisle et al., 2009; Gitomer et al., 2014; Hill et al., 2005; Kersting et al., 2012; Phelps et al., 2014). Most of these studies have focused on broad content areas such as a year of elementary mathematics and investigated relationships between teachers’ knowledge for teaching and measures of instructional quality sampled throughout the school year. This dissertation builds on this work and has a goal of defining physics-specific CKT and measuring how this CKT relates to instructional quality in two ways. First, the papers in this dissertation investigate the link between CKT and instructional quality within one unit of a high school physics course, the unit on energy in mechanics. By focusing on one unit of instruction, it is possible to draw links between teacher knowledge related to specific content, how they carry out instruction related to that content, and how their students learn the content. Second, the papers in this dissertation investigate instructional quality through classroom artifacts, or the assignments and assessments that teachers design for instruction. While observations of instruction can provide insight into how teachers ask questions and explain content, and to the discourse patterns of the classroom, classroom artifacts can provide insight into the types of tasks teachers expect students to engage in, the cognitive demand of the classroom, and expectations of the classroom regarding what students are expected to know about a particular content area.

This work in this dissertation is one part of a larger research project, the Content Knowledge for Teaching Energy Project (CKT-E) (P.I. Drew Gitomer; National Science Foundation DRL 1222732 and 1222777), a multi-institutional research project which set out to define content knowledge for teaching energy, develop an assessment of teachers’ CKT-E, develop measures to assess instructional quality and student achievement related
to the teaching of energy, and investigate the link between CKT-E and instruction (Etkina et al., 2018; Phelps et al., 2020)

The three papers in this dissertation build on the construct of CKT and investigate the relationship between CKT-E and instructional quality as follows.

The first paper in this dissertation, co-authored with Courtney Bell and Drew Gitomer, describes the development of a protocol to assess the instructional quality and demand of classroom assignments and assessments. In this paper, we called upon work to measure the intellectual demand of assignments and assessments (Newmann et al., 2001; Wenzel et al., 2002) and the extent to which students are engaged in complex thinking when working on assignments and assessments (Clare & Aschbacher, 2001), as well as work to develop protocols to measure artifact quality in science and mathematics (H. Borko et al., 2007; Hilda Borko et al., 2005; Martínez et al., 2012). Using the above research, we developed a protocol to measure the quality of the assignments and assessments that teachers select and devise for their unit on energy. We then investigated the validity of the scores produced by the protocol using a validity argument approach (Kane, 1990; 2006). Finally, we investigated the relationship between teachers’ CKT-E and the quality and demand of their assignments and assessments.

The second paper in this dissertation, co-authored with Courtney Bell, Eugenia Etkina, and Drew Gitomer, focuses on the breadth and depth of the goals that teachers set for their unit on energy, how they assess students’ achievement of those goals on their end of unit assessment, and their CKT-E. In this study, we analyzed pre-instruction interviews to determine teachers’ goals, and then analyzed each item and task on their unit assessment to determine the content focus of each. We then compared the breadth
and depth of their goals, the breadth and depth of the content of their assessments and their CKT-E.

The final paper in this dissertation (Robert Zisk is the sole author) contains four case studies of teachers with varying levels of CKT-E and student achievement. In this study, I analyzed pre-instruction interviews, classroom observations, assignments and assessments, and student outcomes to describe how CKT-E influences all aspects of teaching practice and how CKT-E and classroom practice mediate student achievement.

All three papers in this dissertation provide insight into the relationship between CKT, instructional quality, and student outcomes, and advance the theoretical construct of Content Knowledge for Teaching. The papers also help us learn how teachers use that knowledge in practice.
References


Chapter 2: The Relationship of High School Teachers’ CKT-E to the Quality and Demand of the Assignments and Assessments they Design and Select for their Energy Unit

Abstract: Content Knowledge for Teaching (CKT) describes the knowledge that teachers possess for teaching a particular subject that enables them to carry out the work of teaching. As such, there should be a relationship between teachers’ CKT and their classroom practice. In this study we focus specifically on the relationship between teachers’ CKT and cognitive demand of the assignments and assessments they use during instruction. As part of a larger effort to create and validate a measure of content knowledge for teaching physics, specifically energy in the context of mechanics, we investigate the cognitive demand of the assignments and assessments that teachers design or select to use during their unit on energy. We then assess the relationship between the demand of teachers’ assignments and assessments and their content knowledge for teaching energy (CKT-E). We find that the teachers with higher CKT-E used more demanding assignments and assessments during instruction. This finding provides evidence of a relationship between a teacher’s knowledge for teaching a particular subject and specific teaching practice within that subject area.

Introduction

Researchers, policy makers and educators have paid significant attention to the importance of teachers providing students with high quality instruction in science in US science classrooms. In the US, and in other countries, science learning has begun to shift toward using inquiry to help students develop concepts, drawing connections between concepts and ideas, and building students’ ability to learn through engaging in the practices of science (Achieve, 2010; Martin et al., 2004; Tekkummru-Kisa et al., 2015).
The shift in the goals of science education was outlined by the Framework for K-12 Science Education (Quinn et al., 2012), which called for teachers to provide time for students to develop meaningful understanding of the content through deep exploration, by engaging in science practices (NGSS Lead States, 2013). This push has moved science learning away from developing students’ knowledge of facts and procedures, which has been found to be common practice in many science classrooms (Levin, 2008; Levin, Hammer & Coffey, 2009).

Though researchers and policy makers have stressed the importance of high-quality science teaching and learning, it is the teachers who ultimately need to implement the new standards. In order to be able to teach differently, teachers must have knowledge of the subject matter, curriculum, and instructional methods that enable them to design instruction that encourages the development of the core ideas of their science discipline through engaging students in the practices of science.

The idea that teachers possess content knowledge that is specific to teaching and enables them to carry out the everyday tasks of teaching is the basis of Content Knowledge for Teaching framework (Ball et al., 2008). This framework includes and expands upon the framework of Pedagogical Content Knowledge (Shulman, 1987) which described the knowledge necessary for teaching as a combination of subject-matter knowledge and the knowledge of how to make a subject comprehensible to others and an understanding of the conceptions, pre-conceptions and difficulties students may have regarding a subject. CKT framework built upon the work of Shulman by describing teacher knowledge as not only the knowledge of the content, but also as knowledge of the content that is specific to teaching that goes beyond what an educated adult might know.
about a particular subject and beyond what a non-teaching content expert knows about the subject. CKT is based on the premise that the knowledge a teacher has for teaching a particular subject should be related to classroom practice, and in the science classroom, should be related to a teachers’ ability to carry out high quality science instruction. As such, teachers with more developed CKT should carry out higher quality instruction than those with less developed CKT. However, this relationship has been explored in few studies (Park et al., 2011; Keller et al., 2017), and little is known about the specific differences in practice among teachers with different CKT.

The study described in this paper is based on the hypothesis that teachers with different CKT will enact practice in different ways, as those with more developed knowledge will have a more extensive set of resources to employ when planning for and carrying out classroom instruction. The goal of this study is threefold: first, to develop a protocol to measure instructional quality in physics, second, to assess the validity of the scores produced by the protocol to measure instructional quality, and third, to examine the relationship between instructional quality and teachers’ content knowledge for teaching. We explore the relationship between a teacher’s knowledge for teaching a particular domain, specifically, energy in the mechanics portion of the first high school physics course, and the enactment of such knowledge in practice. To measure teacher’s knowledge for teaching a particular domain we used content knowledge for teaching energy (CKT-E) assessment (Etkina et al., 2018) and to measure the enactment of this knowledge in practice we used classroom artifacts – assignments and assessments that teachers design as they provide insight into the content of instruction, the types of tasks the students are provided, and the expectations of the classroom.
To evaluate teacher designed assignments and assessments we developed a protocol that evaluates the quality and demand of classroom artifacts in physics and then examined the validity of the scores produced by the protocol. We then explored the relation of the scores produced by the protocol to the scores of the teachers on the CKT-E assessment (Etkina et al, 2018) in order to make inferences regarding if and how teachers with different CKT-E as measured by the CKT-E assessment differ with regard to assignments and assessments they design and select for instruction. Specifically, we pose the following questions:

1) *How does the quality and complexity of assignments and assessments vary across teachers with different CKT-E?* And

2) *How is content knowledge for teaching as measured by the CKT-E assessment related to the quality and complexity of the assignments and assessments high school physics teachers design and/or select for their units that focus on energy in the context of mechanics?*

We address these questions in the context of a larger project to create and validate measures of Content Knowledge for Teaching in high school physics, specifically the knowledge of teaching energy in the context of mechanics (The Content Knowledge for Teaching Energy Project, Etkina et al., 2018). To interrogate the validity of the claim that teachers with different CKT enact practice in different ways, we employ an argument approach to validity (Kane, 2006) in which we investigate the reliability of, and variation in, the scores produced by the artifact protocol, and the relation of the protocol scores to the estimates of teachers’ content knowledge for teaching energy.
Theoretical Framework

Content Knowledge for Teaching

Content Knowledge for Teaching (CKT) (Ball, Thames and Phelps, 2008) describes the knowledge that teachers possess that enables them to carry out the daily tasks of teaching in the classroom with the goal of helping their students learn. This knowledge includes the knowledge of the content of their subject area, but also includes knowledge of the content that is specific to teaching, or “specialized content knowledge,” which goes beyond what an educated adult should know about a subject, and pedagogical content knowledge (Shulman, 1987) which includes orientation towards teaching, knowledge of students’ ideas and difficulties, the curriculum, and context (Gess-Newsome, 2015; Magnusson et al., 1999). More recently, researchers have worked to develop practice-based conceptions of CKT that describe content knowledge for teaching as enabling teachers to carry out keys tasks or works of teaching when both planning and carrying out instruction (Mikeska et al., 2017, 2018; Park & Oliver, 2008; Park & Suh, 2015; Geoffrey Phelps et al., 2014). Common to these task-based conceptions of CKT is the ability to develop instructional tasks and assessments that meet the goals of instruction among many others.

To illustrate the difference between more common content knowledge and the specialized knowledge needed for teaching, consider the following example: Those knowledgeable in physics content know Newton’s second law and commonly describe it as \( \text{Force} = \text{Mass} \times \text{Acceleration} \). The knowledge that is common to most who have taken a physics course is that there is a relationship between the forces exerted on an object, the objects mass and its acceleration. However, those with more specialized knowledge for
teaching understand that $F = ma$ is not productive and is not entirely correct as it is only true when there is just one force exerted on an object. Therefore, they think of Newton’s second law as \textit{Acceleration} = \textit{Sum of the Forces} / \textit{Mass}. Structured this way, the equation not only allows to accurately determine the acceleration of an object but also describes the cause-and-effect relationship between the object’s acceleration and its mass and emphasizes that it is the sum of the forces exerted on the object that matters not just one force. They also understand why it is more productive to use this representation of the Newton’s second law in the (Etkina et al., 2019). This example helps describe the specialized knowledge that teachers need to call upon during the work of teaching.

Several studies have worked to develop assessments of teachers’ content knowledge for teaching particular subjects including mathematics (Baumert et al., 2010; Blömeke et al., 2011; Foundation, n.d.; Hill et al., 2004; Krauss et al., 2008), English-language Arts (Carlisle et al., 2009; Foundation, n.d.; Kucan et al., 2011; G. Phelps & Schilling, 2004) and science (Coffey et al., 2011; Forbes et al., 2015; K. J. Roth et al., 2011).

As the construct of CKT specifies the knowledge that teachers have for teaching a particular subject, a hypothesis that naturally arises is that teachers with different levels of CKT for a particular subject will also differ in their instructional practice and in turn, have students who attain different outcomes. As such, several studies have focused on assessing the relationship between teachers’ CKT and measures of classroom practice and student outcomes. These studies have reported positive relationships between classroom observation scores and teachers’ CKT (Hill & Charalambous, 2012; Hill et al., 2007; Hill et al., 2005; Hill et al. 2012) , as well as between a teachers’ CKT and student
outcomes as measured by standardized assessments (Hill & Charalambous, 2012; Hill et al., 2005; Hill et al., 2012). In science, Park, Jang, Chen, and Jung (2011) measured biology teachers’ knowledge for teaching photosynthesis and heredity and their ability to carry out reformed teaching practices, and found a significant, positive relationship between teacher knowledge and reformed teaching. Keller, Neumann and Fischer (2017) also examined the influence of teachers’ knowledge for teaching physics on student achievement and found a positive correlation between a teacher’s knowledge and the achievement of their students. Each of these studies has provided insight into the correlations between a teacher’s knowledge for teaching a subject and other measures associated with teaching quality and effectiveness. However, while these studies show that there is some difference in classroom practice between teachers with different CKT, they do not provide information as to the specifics of how their practice differs. For example, teachers with higher CKT may be more proficient at carrying out classroom discussions or questioning, or with regard to classroom artifacts, teachers with higher CKT may create more complex assignments and assessments.

Content Knowledge for Teaching Energy Framework (CKT-E)

The study described in this paper is part of a larger effort to define the framework of content knowledge for teaching energy (CKT-E) and to create both an assessment and practice-based measures of content knowledge for teaching physics in the domain of energy with an overarching goal of developing a nuanced understanding of the relationship between CKT-E and classroom practice (Etkina et al., 2018). The project focused on one content domain in order to develop a coherent theoretical link between CKT and practice.
Central to this effort was the development of the framework of CKT-E. The framework was based on the assertion that teachers need to have knowledge of the important student learning targets within the domain of learning and the knowledge to enact tasks of teaching to support students in meeting those targets. However, the knowledge of learning targets is not always sufficient to interpret the range of student ideas and responses that can arise during instruction. This additional knowledge described as *Horizon Knowledge*. As such, the framework is comprised of three components, the *Tasks of Teaching*, the *Targets for Students*, and *Horizon Knowledge* within the domain of energy (Etkina et al., 2018) (See Figure 1).

**Figure 1**

*CKT-E Framework*

The tasks of teaching describe the activities that teachers engage in that help to support and promote student learning (Ball & Bass, 2000). Within our framework, these tasks include: (I) anticipating student thinking around science ideas; (II) designing, selecting, and sequencing learning experiences and activities; (III) monitoring, interpreting, and acting on student thinking; (IV) scaffolding meaningful engagement in a
science learning community; (V) explaining and using examples, models, representations, and arguments to support students’ scientific understanding; and (VI) using experiments to construct, test, and apply concepts (Etkina et al., 2018. The knowledge to carry out the tasks of teaching described here belongs within pedagogical content knowledge in Ball and colleagues’ (2008) model.

In the study described in this paper, we are specifically interested in the teachers’ ability to carry out two key tasks of teaching: designing and selecting learning activities and experiences (task II), and developing assessments that match the goals for instruction (task III).

Along with the tasks of teaching, the second component of the CKT-E framework is the set of Student Energy Targets. The targets focus on specific content within the domain of energy in mechanics and articulate the important concepts, practices and knowledge representations that are critical to the domain and are in line with existing research on the learning and teaching of energy (Etkina et al., 2018). They are comprised of both common energy content and specialized energy content that is specific to the work of teaching in Ball and colleagues’ (2008) model. The targets are organized into seven target categories with sub-targets within each. The target categories consist of:

- connections of energy and everyday experiences,
- choice of system,
- identification and differentiation of different forms of energy and other physics concepts,
- transfer of energy,
- use of mathematics,
• use of representations, and
• science practices.

These targets go beyond simple terms, definitions, and equations, and instead focus on the disciplinary core ideas related to energy, the crosscutting concepts of systems and energy and matter, and the science practices. For example, instead of just knowing that the definition of the conservation of energy means that any change in the total energy of a system can be accounted for by the work done on the system, the student energy target related to conservation describes that the student should develop an understanding that the energy of a system is always conserved, and, in some cases, it could be constant (Energy Target D1). In energy, constancy means that the total energy of a system does not change (remains constant in time), while conservation means that any change in the total energy of the system can be accounted for – the total change in energy of the system will equal the energy transferred into or out of the system (Seeley et al., 2019).

In each of the first six target categories, the sub-targets are specific to energy. In contrast, the final target category, science practices, uses the science practices from the Next Generation Science Standards (NGSS Lead States, 2013). While the sub-targets within the practices category are not written as specific to the learning of energy, the framework focused solely on the enactment of science practices that helped students develop knowledge of the energy targets.

The third component of the CKT-E framework is Horizon Knowledge. Though this knowledge is not a direct focus of this study, to have a complete understanding of the student energy targets and how to help students develop knowledge of the targets,
teachers must call upon their horizon knowledge. Consider a situation in which a student says that they know that the gravitational potential energy of system that contains an object and Earth increases as the object moves farther and farther from the surface of Earth, so an object far in space must have an infinite amount of gravitational potential energy. In this case, the teacher needs to be aware that \( mgh \) is a mathematical approximation which is used approximate gravitational potential energy near Earth’s surface and that for objects farther away from Earth, \( -\frac{Gm_1m_2}{r} \) provides a more accurate approximation.

The framework of CKT-E is centered on teachers’ knowledge of the content targets within the domain, and their ability to carry out the tasks related to teaching those targets. For example, in order to be able to carry out the task of teaching of designing a learning activity, such as developing a set of problems, or the task of designing an assessment item, teachers need to be aware of the important learning targets within the domain, and then have the knowledge and ability to craft problems and items that provide the opportunity for students to achieve the learning targets or, in the case of an assessment item, provide information as to if the students achieved the learning target.

**CKT-E and the tasks of teaching energy**

In order to help students achieve the important learning targets within the domain of energy in mechanics, teachers must engage in several tasks of teaching. Consider a teacher trying to help students develop understanding of energy constancy and conservation and the difference between the two (Energy Target D1). Energy conservation is a disciplinary core idea and embedded within the *energy and matter* cross-cutting concept of the NGSS. To help students construct this idea, the teacher must
help the students develop the following elements: (1) the energy of any system is always conserved but not necessarily constant (Energy Target D1), (2) the energy of a system can remain constant in some processes (Energy Target D1) and (3) sometimes one system makes the analysis of a process easier than another (Energy Target B2). To address these targets within their unit on energy, a teacher must first engage in the task of teaching of setting a goal for instruction related to conservation and constancy. To do so, the teacher needs to have two different sets of knowledge. First, they need to possess the knowledge of constancy and conservation. Second, they need to know that when conducting an energy analysis of a process, it is possible to define the system in different ways. Analyzing a phenomenon using one system we might find that no work was being done on it and the energy of the system is constant and analyzing the same phenomenon using a different system we might find that the energy changes as work is being done on the system. In this case, the energy is conserved but is not constant. Third, the teacher needs to know when some systems are more useful for analysis than others.

Once the teacher sets the above goal(s), they then need to engage in the tasks of teaching that enable the students to meet the goal. Examples of such tasks are helping students carry out experiments and analyze patterns in the data, engaging the class in discussions, selecting and using examples related to the target and choosing questions to ask the students. Along with those tasks, another key task of teaching that a teacher needs to engage in is designing and selecting activities and problems to both help the students meet the target and to assess their knowledge of the target. To begin to craft learning activities and problems, the teacher first needs to have understanding of the types of activities that will help students meet the goal. In the case of the target highlighted here, a
teacher could ask what energy conservation means, or what energy constancy means, but that would not address the idea that the energy of a system is always conserved and sometimes constant.

In order to help students meet the above goals and energy targets, the students need to be provided with the opportunity to analyze different phenomena in which the energy of a system is both conserved and constant and in which energy is not constant. In this case, the teacher could choose to create a set of problems in which the students analyze a situation in which a small falling ball hits the ground and stops. This process could be analyzed using several different systems and the teach would have to choose the system carefully in order to meet the goals they set for the students. The teacher could set the system as Earth, the ball and the floor. As the ball falls, it speeds up, as the gravitational potential energy in the system is converted to kinetic energy. As it hits the ground, the ball stops, and the system no longer has kinetic energy (assuming that Earth’s kinetic energy is negligible) but both the ball and the floor warm up a little. As the system includes the floor in the system, the energy analysis is rather simple – the kinetic energy is converted into internal energy of the ball and the floor, and the energy of the system is constant. Alternatively, the system could be set as the ball and the floor. In this case, as the ball falls, Earth does work on the ball and the kinetic energy of the system increases, and as in the previous system, when the ball hits the ground, the kinetic energy is converted into internal energy, and the total energy of the system would be conserved, but not constant because Earth did work on the system. Finally, the system could be set as just the falling ball. Like the previous system, the system gains kinetic energy as the ball falls and Earth does work on the system. However, since the floor is not in the system, it
appears that the ball stops, and the kinetic energy of the ball goes to zero and the energy of the system is not constant. Where did the energy go? One answer is that the floor did negative work on the ball, but the floor did not move, thus it could not do work. In addition, the floor warmed up a little – where did this energy come from? In this case, for a student the energy of a system is not only not constant, but it is not conserved either. If this system is presented to students, it may appear to them that choosing different systems (one in which the floor is in the system and one in which it is not) leads to two different answers. To help reconcile this for the students, the teacher needs to employ a third type of knowledge, their horizon knowledge. The horizon knowledge in this case is the idea of pseudo-work. As the ball hits the ground, the ground compresses the ball slightly, doing work on its center-of-mass, and stopping the ball. To a student, it appears that no work is done on the ball because there is no obvious displacement, but with horizon knowledge, the teacher can reconcile the student’s idea and help them to see that the energy of the system is conserved.

To create a problem or problems to address the above goals and targets, the teacher first needs to decide if they are going to provide one of the above systems for the students, or if they are going to let the students choose their own system. Additionally, the teacher needs to have an awareness of the different tools that students can use to analyze the situation. Will the students be asked to set up a conservation of energy equation and analyze the problem mathematically? Will the students be asked to use different representations while analyzing the problem, such as bar charts (Van Heuvelen & Zou, 2001), or pie charts? In creating any learning activity, problem or assessment item, these are examples of the types of choices a teacher needs to make. A teacher needs
to employ all three types of knowledge, the knowledge of the student energy targets, the knowledge of the tasks of teaching, and horizon knowledge to develop such activities (Etkina et al., 2018).

All three sets of knowledge (knowledge of student learning targets, knowledge of the tasks of teaching and horizon knowledge) are essential to the work of teaching. Consider a physicist who has a strong personal grasp of the energy targets but does not have the knowledge to carry out the task of teaching of designing learning activities and problems. If this was the case, they would not necessarily be able to develop activities and problems related to the targets. Conversely, if a teacher believes that developing an understanding that energy conservation is an important energy target (note that just developing an understanding of conservation does not rise to the expectations of the energy target stated at the outset of this example), and their knowledge related to the energy target of conservation is simply that conservation means the energy is conserved in an isolated system (which is a common, incomplete understanding of energy conservation and constancy (Seeley, Vokos, & Etkina, 2019), then the items they create might be definition items where the students are asked to define conservation or problems where no work is done on the system.

The framework and example presented here illustrate how a teacher’s CKT can influence their enactment of the everyday tasks of teaching. Teachers’ knowledge of student learning targets, the tasks of teaching and ultimately their CKT, can influence all areas of teaching practice, including the content of their goals, and the activities and problems they use during instruction and on their assessments. In this study, we explore the relationship between teachers’ CKT-E as measured by the CKT-E test (Etkina et al.,
2018) and the quality of the assignments and assessments they design and the selection of the assignments and assessments for instruction. The quality of the assignments and assessments will be analyzed using a protocol developed based on the CKT-E framework that was intended to measure the quality and demand of assignments and assessments used during energy instruction.

**The cognitive demand of assignments and assessments**

When developing assignments and assessments for instruction, a teacher must first decide what learning targets to focus on and then decide what type of problem or activity to use to address those targets. Consider the following two problems focused on the learning target of work as the change in energy of a system:

1) *A waiter exerts 4 Newtons lifting a dinner tray 1.2 meters. How much did the energy of system tray-Earth change?*

2) *A waiter catches a 2kg bag of flour that has fallen off a 2-meter-high shelf. If the waiter starts catching the bag .8 meters above the ground and stops it at the ground level, what is the average amount of force must the waiter exert catching the bag to stop the flour bag? Consider the system to be the bag of flour and Earth. Show your answer using both energy bar charts and mathematics.*

To answer the first question, the student needs to know that to calculate the change in energy using the equation for work \( \text{Work} = \text{Force} \times \text{Distance} \times \cos \theta \) and then be able to plug the values for force, distance, and \( \cos \theta = 1 \) to solve. In the second question however, the student similarly must know that work leads to the change in energy, but then also needs to know that at .8 meters, the bag of flour-Earth system has both gravitational potential energy and kinetic energy, and that in order to stop the bag
before it hits the ground, the work done on the system by the waiter must be negative. They then need to be able to set up a bar chart representing the process and the generalized work-energy principle to solve the problem. Once they complete those steps, then they can plug in the appropriate values to solve. One thing to note about the second problem is that the value for distance is not directly provided. The student needs to have the awareness that if the bag starts to be caught at .8 meters, this is the needed distance.

While both of these problems focus on the same subject matter and learning target (work as the change in energy), they clearly differ in the knowledge and skills needed to solve them. The first problem only requires the knowledge of work as the transfer of energy and the generalized work-energy principle. However, in the second problem, the student needs to know much more, including how to choose the correct system (in this case, the bag of flour and Earth), which initial and final state to choose, how to properly construct a bar chart that represents the bag of flour falling and being caught, and then how to set up the generalized work-energy principle that will enable them to calculate the average amount of force needed to stop the bag before it hits the ground. This difference in difficulty and the knowledge and skills necessary to solve the problems can be described as a difference in the cognitive demand of each item.

When analyzing tasks and problems, cognitive demand has been used to describe the level of thinking demanded of students to successfully engage in the task (Stein et al., 2001; Tekkumru-Kisa et al., 2016). Tasks that are highly structured and require minimal steps to solve are considered to be low-demand tasks, while tasks with multiple steps, more than one solution path and open-ended tasks are considered to be more demanding.
In science, Tekkumru-Kisa and colleagues (2016) developed a framework for analyzing the cognitive demand of tasks. At the lowest level (level 1) are memorization tasks that only require the reproduction of definitions or equations seen before. At level 2, the problem requires the students to solve a problem using a script that they have employed before. Problem 1 above can be considered a level 2 task under this framework as the student would only need to know the equation for work and then plug in the values. Level 3 and 4 tasks within this framework are less structured but still have possible solution paths that the students have followed before. However, these tasks cannot be followed mindlessly (just plugging in values, for example). The second problem above would fall into one of these two levels. Finally, at the highest level (level 5) are “doing science tasks” in which students are engaged in science practices to help develop an understanding of natural phenomena. Tasks at the higher levels of demand are in line with the expectations of the Next Generation Science Standards, which call for students to develop a knowledge of the core ideas of science through engaging in the science practices. Figure 2 below provides examples of each level of demand.
Levels of Cognitive Demand with Examples

<table>
<thead>
<tr>
<th>Level of Cognitive Demand</th>
<th>Example Item</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>What is the kinetic energy of a 2-kg ball moving at 5 m/s?</td>
</tr>
<tr>
<td>2</td>
<td>What is the change in energy of 2-kg box pushed across a smooth floor for 2 meters with a force of 20 N?</td>
</tr>
<tr>
<td>3 - 4</td>
<td>A waiter catches a 2-kg bag of flour that has fallen off a 2-meter-high shelf. If the waiter starts catching the bag 0.8 meters above the ground and stops it at the ground level, what is the average amount of force must the waiter exert catching the bag to stop the flour bag? Consider the system to be the bag of flour and Earth. Show your answer using both energy bar charts and mathematics.</td>
</tr>
<tr>
<td>5</td>
<td>You are given a meter stick, a spring that obeys Hooke’s Law, a 500-g object, a spring scale and a cell phone with a slow-motion camera. Design an experiment to test conservation of energy. What data will you collect? What data will support conservation of energy? What are some possible sources of uncertainty?</td>
</tr>
</tbody>
</table>

Along with this framework of cognitive demand in science, there have been efforts to describe the levels of demand and problem-solving pathways needed to complete tasks and solve problems specifically in physics. Tuminaro and Redish (2007) captured the steps students took while responding to algebra-based introductory physics problems. At the highest and most sophisticated level, when solving problems, students mapped meaning to mathematics. To do this, students had to develop a conceptual understanding of the physical situation or phenomenon they were analyzing and then progress to a solution. Problems that required this approach could not be solved by simply plugging in numbers to an equation, and these types of problems would fall into the higher levels (levels 3 or 4) of the cognitive demand framework described previously. At the lowest level of sophistication, Tuminaro and Redish observed students engaging in
plug and chug or multi-step (recursive) plug and chug problems. These problems required little cognitive effort and fall into the lower levels of the cognitive demand framework. Teodorescu and colleagues (2013) also characterized the levels of demand of typical problems in physics. In their framework, problems at the lower levels only basic retrieval of facts or equations, and at more complex levels, require students to engage in analysis of a physical situation or engage in experimentation.

Though these frameworks have identified a wide-range of tasks with regard to cognitive demand, research into the types of tasks and problems that provided to students in science classrooms indicate that teachers focus on much lower-level tasks such as definitions, the use of step-by-step equations, scripted scientific investigations and problems and tasks with little or no connection to scientific ideas (Abd-El-Khalick et al., 2008; Abd-El-Khalick, 2012; Levin, 2008; Levin et al., 2009; Quinn et al., 2012; K. Roth & Givvin, 2008). This lack of cognitive demand in science classrooms is problematic because without exposure to more demanding questions, students will not be able to develop the reasoning skills to solve such problems in the future and develop the understanding of the subject-matter that goes along with developing those reasoning skills.

In this study, we attempt to find a relationship between the cognitive demand of the assignments and assessments that teachers use during their energy unit and their CKT-E. As CKT-E specifies the knowledge teachers have that enables them to carry out the work of teaching, there should be a relationship between CKT-E and the demand of the assignments and assessments that teachers design and select for instruction. Definition problems, plug and chug and problems with only one right answer are simple
to create, so teachers with less developed knowledge can easily develop those items for their class. However, more complex problems, those with multiple solution paths, multiple steps, and possibly multiple answers are more difficult to create and as such require a more developed knowledge of the subject matter and of student reasoning. Teachers with more developed CKT have more resources to employ when creating assignments and assessments and therefore should be able to create more demanding tasks.

**Designing a protocol to measure the demand and quality of assignments and assessments in physics**

As the goal of this study was to measure the demand and quality of the assignments and assessments physics teachers use during their unit on energy and then relate that demand to their content knowledge for teaching energy, we developed a protocol to measure the demand of the tasks provided to students during the unit. We focused specifically on assignments and assessments as designing and selecting assignments and assessments for instruction is a critical task of teaching (Etkina et al., 2018), and as such, they can offer insight into the nature and quality of classroom interactions and instruction (Borko et al., 2007; Gitomer & Bell, 2013; Matsumura & Pascal, 2003). Assignments and assessments have been used to measure the intellectual demand of math and English-language arts classrooms (Newmann et al., 2001; Wenzel et al., 2002), the level in which students are engaged in complex thinking and use content knowledge when working on classroom tasks (Clare & Aschbacher, 2001), and the extent to which classroom artifacts exhibit evidence of reform-oriented instruction in both math and science (Borko et al., 2007; Borko et al., 2005; Martínez et al., 2012). Bol and Strage
(1996) and Koh (2011) used artifacts, specifically assessment items, as evidence for the cognitive demand that teachers place on their students during assessments.

By analyzing classroom artifacts, one can gain insight as to what extent the teacher is engaging in this task of teaching. For example, if an assessment from one teacher contains a significant number of definition-type problems, you would be able to infer that the teacher focuses mainly on declarative knowledge. However, if an assessment from a second teacher contains definition-type problems along with problems that ask students to analyze complex phenomena, and other problems that do not have a clear solution path, you would be able to infer that the teacher focuses on the development of multiple types of knowledge in the content area. If these patterns were repeated across multiple assignments and assessments from each of the teachers, we would then be able to make inferences regarding to what extent and level of quality each teacher carries out the task of teaching related to addressing multiple types of knowledge and in turn, infer that the second teacher with the more complex assignments and assessments has the ability to create or select tasks that focus on multiple types of knowledge and tends to have the students focus on multiple types of knowledge in the classroom.

**Protocol Design**

**Identifying the Dimensions of the Protocol**

As we were interested in the demand and quality of teachers’ assignments and assessments related to their ability to carry out the tasks of teaching learning targets within energy, we constructed the protocol around the tasks of teaching and energy
learning targets developed for the CKT-E project (Etkina et al. 2018). As an example, consider the following task of teaching:

“The teacher elicits student understanding and help them express their thinking via multiple modes of representation.”

The use and interpretation of multiple representations is a key target within energy, and the cognitive demand related to representations could vary significantly across assignments or assessment items. Some items may not require students to use multiple representations, such as asking students for a definition, or asking students to complete a simple math operation, while others may require students to construct multiple representations or translate between them, such as asking students to develop a bar chart based on a phenomenon and then a mathematical model based off of the bar chart. For example, a question could ask what type of energy is associated with the motion of an object or to calculate the kinetic energy of a moving object, which would not require any representations to solve, or the question could be similar to the example of the falling bag of flour described previously that asked the student to draw a bar chart describing the process and then to develop a mathematical representation that is consistent with the bar chart.

To develop the protocol, we reviewed the tasks of teaching and student energy targets for common themes. Within the tasks of teaching, some emerging themes were designing tasks that focus on key concepts, productive representations, mathematical models and experiments; using and developing models and representations to analyze phenomena and seeking consistency between those representations; using and developing mathematical models of phenomena; and engaging with the science practices. Similarly, within the energy targets, we grouped the categories into four themes: The analysis of

We then used the common themes from the targets and tasks of teaching as the foundation for the development of the dimensions of the artifact protocol. Based on these common themes, we identified a set of dimensions to capture the enactment of the tasks of teaching in the assignments and assessments that teachers design and select for instruction. These dimensions are:

- **Analysis:** The depth of analysis required by the problems on the assignments or the assessment;
- **Mathematics:** The depth of mathematics (algebra) needed to solve the problems;
- **Representations:** The types of interpretation and use of representations needed to solve the problems.

We incorporated “depth” and “use” into the domains to allow for the distinction between the level of cognitive demand required of the assignments and assessments. A problem that requires mathematics could be quite easy and not necessarily meet the expectations of instruction and learning set by the tasks of teaching and content targets, or it could be quite demanding.

Each of the dimensions had the purpose to capture the quality of instruction related to the extent to which the teacher designed or selected artifacts that required students to carry out different cognitive tasks with different levels of demand. The
analysis dimension was meant to capture the complexity of the phenomena that the students were asked to work through on the assignments and assessments; the mathematics dimension was intended to capture the depth and complexity of the mathematics the students needed to engage in when solving the problems on the artifact; and the representations dimension was intended to capture the representations that were present in the problems on the assignments or assessments that the students needed to interpret to solve the given problems and the representations that the students needed to use to solve the problems.

**Defining the scale points**

As each dimension was intended to capture the level of demand the artifact required related to the dimension, we developed a set of scale points related to demand. For example, for the representations dimension, when designing an assignment or assessment problem, a teacher could create or select problems that do not contain any representations and do not require the use of representations other than algebraic equations, problems that require the use of multiple representations in order to solve or something in between. To develop the scale points, we defined cognitive demand as the level of reasoning required to complete the task (Stein et al., 2001; Tekkumru-Kisa et al., 2015). At lower levels of demand are tasks that focus on the correct response (definitions and “plug and chug” type problems) that have a well-defined path to completion, and at higher levels of demand are tasks that are less well defined and require a deeper analysis of the phenomenon. To further define our scale points for each dimension, we consulted the literature into problem-solving in physics (Teodorescu et al., 2013; Tuminaro & Redish, 2007), physics textbooks and experts in both physics teaching and content. This
resulted a four-point scale (with scores of 1 – 4) for each dimension, with scoring levels that range from low to high complexity and demand related to the dimension.

For the analysis dimension, we developed the scale points to capture the complexity of the phenomenon that the students were asked to work through within the problem or assignment. Tasks that scored at the low end of the analysis dimension would be problems that do not ask the students to work through the phenomenon in order to solve, and tasks scored at the high end of the scale would be problems that ask students to reason through complex phenomena (such as phenomena with multiple objects in the system or without defined initial and final states) or engage in higher level cognitive tasks such as analyzing errors or justifying their responses with evidence. Figure 3 provides the full description of the scale points and examples items for the analysis dimension.
### Figure 3

**Analysis Dimension with Examples**

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Score</th>
<th>Example items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analysis</td>
<td>1 – Items on the assessment or assignments do not generally require the analysis of a problem situation/process and can be solved through the recall of facts or definitions or through “plug and chug.”</td>
<td>________ energy is the energy associated with a moving object.</td>
</tr>
<tr>
<td></td>
<td>2 – Items on the assessment or assignments generally require the analysis of a simple problem situation or process.</td>
<td>What is the kinetic energy of a 4 kg ball right before it hits the ground if it is dropped from a height of 3 meters above Earth?</td>
</tr>
<tr>
<td></td>
<td>3 – Items on the assessment or assignments generally require the analysis of a problem situation with states containing multiple types of energy or without defined initial or final states.</td>
<td>A .5 kg bouncy ball is dropped from a height of 4 m. After it hits the ground the first time, it rises to a height of 3.2 meters. What is the kinetic energy of the ball after the first bounce when the ball is 2 meters off the ground? If the system is the ball, Earth and the surface, how much energy is converted into internal energy during the bounce?</td>
</tr>
<tr>
<td></td>
<td>4 – Items on the assessment or assignments generally require the student to develop an experiment, to justify, explain or defend their responses or to analyze or evaluate a solution or error.</td>
<td>You are given a cart, a ramp, a photo gate, a scale and a ruler. Design an experiment to test the conservation of energy. Make sure to list all of the assumptions that you made when designing your experiment and describe the calculations that you would make.</td>
</tr>
</tbody>
</table>

For the *mathematics* dimension, we developed the scale points to capture the complexity of the mathematics that the students were asked to engage in when solving the problem or tasks on an assignment. In this dimension, we focus specifically on the level of algebra and mathematical operations needed to solve the problem. Problems that did not require mathematics were scored at the low end of the scale and problems that required the development a mathematical model in order to solve the problem were scored at the high end. It is important to note that a problem was not rated higher just because it required a higher level of mathematics (Calculus as opposed to Algebra, for example). Instead, the
scores were based entirely on the reasoning required to solve the problem. Figure 4 provides the full description of the scale points and examples items for the mathematics dimension.

**Figure 4**

*Mathematics Dimension with Examples*

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Score</th>
<th>Example Item</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mathematics</td>
<td>1 – Items on the assessment or assignments do not generally require the use of mathematics.</td>
<td>________ energy is the energy associated with a moving object.</td>
</tr>
<tr>
<td></td>
<td>2 – Items on the assessment or assignments generally require the explicit application of a definition-type equation with given variables in order to solve</td>
<td>What is the kinetic energy of a 4-kg ball moving at 5 m/s?</td>
</tr>
<tr>
<td></td>
<td>3 – Items on the assessment or assignments generally require application of a definition-type equation and requires preliminary steps in order to calculate the values for the variables needed to solve the equation, or is an item that asks the student to describe a physical process mathematically without solving.</td>
<td>How much power is needed to stop a 10-kg ball moving 10 m/s across a smooth floor in 10 seconds?</td>
</tr>
<tr>
<td></td>
<td>4 – Items on the assessment or assignments generally require multiple steps and require the development of a mathematical model of the process in the problem in order to solve and does not contain explicit scaffolding.</td>
<td>A 500-kg roller coaster cart is moving 10 m/s at the top of a 20m hill. How high can you build the next hill so that the cart goes it at no less than 2 m/s?</td>
</tr>
</tbody>
</table>

Finally, for the *representations* dimension, we developed the scale points to capture the extent to which the students are asked to use representations when working through a problem or task. The use of multiple representations is a common tool for problem solving and an aide in the understanding of physics concepts. At the low end of
the scale, the problem or task did not require the students to use representations, while at the high end, the students were asked to translate between representations and use representations to solve the problem. Figure 5 provides the full description of the scale points and examples items for the representations dimension.

**Figure 5**

*Representations Dimension with Examples*

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Score</th>
<th>Example Item</th>
</tr>
</thead>
<tbody>
<tr>
<td>Representations</td>
<td>1 – Items on the assessment or assignments do not generally ask for the use of representations other than mathematics (and students do not use any other representations).</td>
<td>________ energy is the energy associated with a moving object.</td>
</tr>
<tr>
<td></td>
<td>2 – Items on the assessment or assignments generally require the interpretation of a representation or the interpretation or use of a representation to find values to solve a definition-type equation or problem.</td>
<td>Based on the bar chart shown below, is energy conserved? Is it constant?</td>
</tr>
<tr>
<td></td>
<td>3 – Items on the assessment or assignments generally require the direct translation of one representation into another and consistency between the representations.</td>
<td>Based on the drawing below, draw an energy bar chart for positions A, B and C. Be sure to indicate your system.</td>
</tr>
<tr>
<td></td>
<td>4 – Items on the assessment or assignments generally require the use of or translation between multiple energy related representations and reasoning through the representation(s) in order to solve a physics problem. The items require more than the direct application of a definition type equation.</td>
<td>A 500-kg roller coaster cart is moving 10 m/s at the top of a 20-m hill. How high can you build the next hill so that the cart goes it at no less than 2 m/s? First, draw a bar chart for the situation, and then use the bar chart to develop the mathematical model before solving.</td>
</tr>
</tbody>
</table>
**Descriptive Coding**

In addition to capturing the demand of the assignments and assessments, we also designed the protocol to capture certain descriptive aspects of the artifacts including the content focus, science practices and problem context.

The content coverage dimension was designed to capture the content that is addressed by the problems or tasks. Content tracking allows for an understanding of the breadth of content addressed by the problems or tasks on the artifacts and also allows for an understanding of the types of tasks teachers design around certain content. For this dimension, the content was coded using the CKT-E content targets. To illustrate how items on teachers’ unit assessments were coded for content using the energy content targets, consider the following example from a teacher’s assessment:

*Give an example of a system that loses energy because negative work is done on it.*

To answer this item, students need to have knowledge of two content targets within the domain of energy. Students first need to understand the concept of a system and how the choice of a system can determine the types of energy present within the system. In addition, students need to understand how work, specifically negative work, can change the energy of a system. In analyzing this item using the CKT-E targets, we see that the item addresses the following targets:

1. **B1 - Understands that the energy accounting in a phenomenon depends on the choice of system**

2. **D4 - Recognizes that work is the way in which energy is transferred mechanically and may result in a change in temperature in some cases**

For the problem context dimension, we were interested in learning the context of the problems and tasks that the teacher asked the student to engage in. For this
dimension, the problems or tasks were coded if they were context free (such as a simple plug and chug mathematics problem or definition), if they were based in a “physics” context (such as a ramp or pulley system), or if they required the students to work through a real-world context while solving. Examples of each context are shown in figure 6 below.

**Figure 6**

*Context Codes with Examples*

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Score</th>
<th>Example Item</th>
</tr>
</thead>
<tbody>
<tr>
<td>Context</td>
<td>Context-Free</td>
<td>The type of energy most commonly associated with motion is:</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Or</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Calculate the kinetic energy of a 5kg object moving 4m/s</td>
</tr>
<tr>
<td>Physics Context</td>
<td></td>
<td>Two balls are held at the top of two different, frictionless, 5m high ramps. The first ramp is 4m long, and the second is 3m long. Which ball will be going fast at the bottom? Explain your reasoning using multiple representations.</td>
</tr>
<tr>
<td>Real-World Context</td>
<td></td>
<td>You are designing a roller coaster for a new theme park, and you are trying to figure out how high to make the second hill of the ride. The first hill is 30m high, and the maximum weight of the car is 1000kg. Assuming the coefficient of friction between the wheels and track is .16 and the distance between the top of the two hills is 170m, how high can the second hill be? Consider air resistance to be negligible.</td>
</tr>
</tbody>
</table>

Finally, to capture any science practices, we coded the artifacts for the presence of any of the science practices defined by the Next Generation Science Standards (NGSS Lead States, 2013). For example, consider the following question:

*Using a track, a small car, a photogate, a stopwatch and a ruler, describe a procedure to test the constancy of energy in an isolated system. Provide an example of*
data that would support constancy of energy. Make sure to list any assumptions you would make.

In this item, students have to design an experiment (NGSS science practice *Designing and Carrying Out Investigations*), and provide examples of data that would fit the constancy of energy (NGSS science practice *Analyzing and Interpreting Data*).

**Developing an argument for the validity of using artifacts to measure the enactment of CKT**

Along with developing a protocol to measure the quality and demand of physics assignments and assessments, a second goal of the study was to investigate the relationship between the quality and demand of the assignments and assessments that teachers develop for their energy units and their knowledge for teaching energy in the context of mechanics (CKT-E) to support the claim that teachers with different CKT-E enact practice in different ways when designing and selecting instructional artifacts. To investigate the plausibility of this claim, we employ the validity framework developed by Kane (2006). Under this framework, the validity of a claim can be assessed through the evaluation of a set of inferences. Kane’s framework consists of the formative stage or interpretive argument and the summative stage or validity argument itself (Kane, 1990; 2006). The interpretive argument specifies the inferences and assumptions connecting the theoretical scores produced by the assessment and the conclusions, claims or decisions based on those scores. The validity argument then works to evaluate the inferences and assumptions based on evidence and previous research.

According to Kane (2006), the first step in assessing the validity of the inferences and interpretations made based on scores produced by some measure is to lay out the interpretive argument, or the set of theoretical assumptions that would connect the scores
and the interpretation of those scores. In developing the interpretive argument, Kane described four assumptions and inferences that must be validated as part of assessing the validity of scores produced by the measure: the scoring inference, the generalization inference, the extrapolation inference, and the decision inference. Each of these inferences provides different evidence towards establishing validity. In figure 7, I describe the types of evidence that can be used to support each of the inferences as described in Kane (2006) and Bell and colleagues (2012).
### Kane's (2006) Inferences and Examples

<table>
<thead>
<tr>
<th>Inference</th>
<th>Description</th>
<th>Assumptions to be tested</th>
</tr>
</thead>
</table>
| **Scoring** | The scoring inference concerns the ability for the measure to take a observed performance and translate that performance into a score | - Does the protocol appropriately describe the levels of demands with regard to assignment and assessment tasks in physics?  
- Can the scoring rule be applied accurately and reliably?  
- Do the scores produced by the protocol reflect a range of quality and demand? |
| **Generalization** | The generalization inference concerns the ability for the scores from an observed performance to generalize to the universe of performances | - Is there variation in the demand of the instructional assignments used by a teacher?  
- Can variation in the quality and demand of classroom assignments and assessments be attributed to the teacher or is the variation due to the scoring by the raters? |
| **Extrapolation** | The extrapolation inference concerns the ability for the scores produced by a measure to extrapolate to a level of skill or a level of a trait. | - Is there a relationship between the quality and demand of teachers’ assignments and assessments and their content knowledge for teaching energy (CKT-E) as measured by the CKT-E test or observations or other measures? |
| **Decision** | The decision inference concerns the ability to draw conclusions from the scores produced by the measure. | - Are the decisions based on the measure appropriate? (For example, should the scores produced by the measure be used to evaluate teachers?)  
- Do the scores produced by the measure support the decisions? |
The first inference described by Kane (2006) is the scoring inference. The scoring inference is concerned with the ability of the measure to take a performance or an item and translate the performance or item into a score. For example, with artifacts, can a protocol designed to assess the quality of the artifacts accurately and reliably be used to score artifacts based on their quality? With regard to a protocol meant to assess the quality and demand of classroom artifacts, the protocol should be based on research into cognitive demand and problem solving in physics, and the scores should reflect a range of demand and quality across a sample of artifacts. Additionally, to ensure that the scores can be applied reliably, inter-rater reliability checks must be performed.

The second inference to be evaluated is the generalization inference. This inference is concerned with the relation of the sample used for scoring to the universe of samples. To assess the quality of the artifacts that a teacher uses throughout a year, for example, the quality of a single artifact from the beginning of the year may not be representative of the quality of artifacts throughout the year. Instead, a larger sample with artifacts from throughout the year could possibly be more representative of the artifact quality in the classroom. By using a larger sample, one can control for sampling error that may affect the scores. Additionally, some consideration needs to be made for variance across assignments from the same teacher. When trying to capture the quality and demand of instruction, we should not expect every assignment to be highly demanding. Variability in demand should coincide with the students’ ability and readiness to engage with the more demanding tasks.

The third inference to be evaluated is the extrapolation inference. This inference is concerned with the relation of the scores on one measure to the scores on related
measures, as well as scores on measures that believed to not be related. If two measures are related, the scores produced by the measures should be correlated (subjects who score high on one, should score high on the other), while two measures that are not related would not necessarily produce related scores. In the case of the artifact protocol in this study, the scores produced by the artifact protocols should be related to other measures meant to assess CKT-E, such as observation scores, and scores on an assessment of CKT-E.

Finally, the decision inference concerns the conclusions that can be drawn from the scores produced by the measure. For example, can the artifact protocol be used for professional development to help guide the development of assignments and assessments? The decision inference is not investigated in this study.

In the next sections, we will describe the collection and scoring of the assignments and assessments from teachers who were subjects in the CKT-E study and the evaluation of the validity argument with regard to the claim that teachers enact Content Knowledge for Teaching when designing and selecting assignments and assessments for instruction.

**Methods**

**Participants**

The participants in the study were 32 teachers from the Content Knowledge for Teaching Energy Project. The teachers were selected by responding to a call for participants that was sent to physics teaching mailing lists and groups near each research site. Eleven of the 32 teachers were women. Eighty-eight percent of the teachers had a master’s degree, and the same percentage of teachers took at least three physics courses
as an undergraduate. Each teacher taught at least one physics class, ranging from conceptual physics to Advanced Placement® (AP®) physics and International Baccalaureate® (IB®) physics, and 94 percent of the teachers taught physics for three or more years.

**Collection of Artifacts**

Teachers were asked to submit 4 instructional artifacts from their energy unit. One of those artifacts was to be the final assessment (referred to as “final assessment) from the unit, and the other 3 were artifacts used during the unit (referred to as “instructional artifacts). These artifacts could include worksheets, lab assignments and quizzes, for example. The teachers were given no other instructions as to how to select the artifacts. For each of the artifacts collected, teachers were asked to submit 9 samples of student work, 3 that the teacher considered to be of high quality, 3 medium quality, and 3 low quality. As the study was intended to measure teaching practice and teacher knowledge, we did not explicitly score the student work. Instead, it was used as an additional source of evidence when scoring teacher-created problems and questions. For example, if a problem did not ask for the use of multiple representations, but each student used representations when solving, we took this as evidence that the teacher has an expectation that the students use multiple representations. The teachers were given no other guidance on how to pick the assignment or choose the student work samples other than asking them to space the artifacts out over the whole unit.

Of the 32 teachers in the study, 28 provided a full set of artifacts including the final assessment and three instructional artifacts. Only those who submitted a full set of artifacts were included in this study.
**CKT-E Assessment**

As part of the CKT-E project, each teacher completed the CKT-E assessment (Etkina et al., 2018; Phelps et al., 2020). This assessment was comprised of items that were based on the tasks of teaching energy in a high school physics class. For example, some items required the teacher to analyze and respond to student ideas while others asked the teacher to sequence instruction in a way that would be most effective in helping students develop certain concepts.

The assessment was administered to and field-tested with 329 physics teachers from across the United States. Following the scoring of the assessment, each teacher was assigned a standardized item response theory (IRT) estimate of their CKT-E. The estimates of CKT-E for the teachers in the field test ranged from -1.95 to 2.16, with a median estimate of .01 (Phelps et al., 2020). CKT-E IRT estimates for the 28 teachers in this study ranged from -1.44 to 2.16. The mean CKT-E estimate for these teachers was .55 (SD=.87), indicating that the CKT scores for teachers in the study were above the average of the field test sample.

**Raters**

Four raters to scored the artifacts. As the protocol is specific to physics, all of the raters had knowledge of physics and physics teaching. Three of the raters had degrees in physics, 1 had a degree in chemistry, and 4 were enrolled in a graduate program focusing on physics education. Two of the raters were practicing teachers, one for 5 years, and one in his first year. The other 2 raters had experience teaching physics at the college level. The first author served as the master rater.
Raters were trained before scoring. During the training sessions, the raters were first provided with the protocol for review, dimension by dimension. Following the review of each dimension, they were asked to score up to 15 individual items from assignments and assessments, and then were able to discuss their scores to address any confusion or unclear aspects of the protocols. This process was repeated for each dimension. Following the initial scoring of the individual items, the raters were asked to score complete 5 training assignments and assessments independently. After scoring, the raters were again allowed to discuss their scores. Once the raters completed training, they began live scoring.

**Scoring Procedure**

For each teacher, the final assessment and the 3 non-test artifacts were scored separately. The final assessments were scored first, followed by the assignments. Each artifact was double scored. The scoring of the assignments and assessments was separated because, (a) the assignments were often short and less varied in their content coverage, while the tests were longer and tended to address multiple aspects of content; and (b) the instructional assignments and assessments were used during the unit, while the final assessment was used after instruction. Throughout the text, we will distinguish between the final assessment scoring, and the scoring of the instructional artifacts other than the final assessment.

**Final Assessment Scoring Procedure**

The final assessments were scored two ways, at the item level and at the test level. To score the tests at the item level, two independent raters scored each item on each assessment on each of the scored dimensions and the descriptive dimensions. This
process produced scores on each dimension for each item, as well as descriptive information (context, content). Following the scoring of each item, the item scores on each dimension were averaged together to get an overall score on each dimension for the assessment. Therefore, each teacher had a score for each of the three scored dimensions, as well as an overall demand score, which was the average of the three dimensions. This process is described in figure 8 below.

Figure 8

Assessment Item Level Scoring Process

1. Each item was double scored on the analysis, mathematics and representations dimensions

2. The dimension scores from the two raters were averaged together for each item creating item dimensions scores (TEST IDS)

3. All items IDSes were averaged together to produce an overall score for each dimension (TEST ODS): analysis, mathematics and representations

4. The ODSes were averaged together to get an overall demand score for the assessment

After completing the item level scoring, two additional raters independently scored the test as a whole on each of the 3 scored dimensions (analysis, mathematics, and representations). (Test Level Scoring) They were instructed to review the whole test, and then make a holistic judgment when scoring each dimension. The total score for the final
assessment was calculated using the average of the 3-dimension scores. This process is described in figure 9 below.

**Figure 9**

*Assessment Test Level Scoring Process*

1. Each test was double scored. Each rater read through the whole test and scored it on the analysis, mathematics and representations dimension.

2. These dimension scores from each rater were averaged together to produce an analysis, mathematics and representations score for each assessment.

3. The three dimension scores were averaged together to produce an overall demand score for each assessment.

Any discrepancies between the two raters were resolved by a third, who was the master rater.

After scoring, we reviewed the descriptive statistics of the scores produced by both the item level scoring and the test level scoring, including reliability, and the distribution of scores and chose to use the scores from the assessment scoring produced at the test level. In addition, we were concerned that scoring at the item level would give equal weight to simple items, as well as more complex items, which would cause the overall scores of some assessments to be lower, despite being more demanding. A full
description of the scoring and aggregation methods will be described in a further publication.

**Instructional Artifacts Scoring**

The instructional assignment and assessments were scored after the final assessment scoring was completed. As each teacher submitted a set of instructional artifacts, two independent raters scored each artifact in the set submitted by the teacher separately (Artifact Level Scoring). They were instructed to read through the assignment or assessment, and then score each artifact in the set on each scored dimension and descriptive dimension. The raters would score the first artifact on each dimension, then the second, and then the third. The master rater resolved any discrepancies between the two raters. The scores on each dimension from each artifact in the set were then averaged to get a score on the set of artifacts for each of the dimensions. The overall score for the instructional artifacts was calculated using the average of the 3 dimension scores. This resulted in the teacher having four scores, one for each dimension, which was the average of the dimension scores across the 3 instructional artifacts and an overall demand score which was the average of each dimension score. This process is described in figure 10 below.
*Figure 10*

**Instructional Artifact Scoring Process**

1. Each rater read through each assignment and scored each on on the analysis, mathematics and representations dimensions

2. The dimension scores from the two raters were averaged together for each assignment

3. The dimension scores from each assignment were averaged to produce an overall analysis, mathematics and representation score the set of assignments

4. The overall dimension scores were averaged together to get an overall demand score for the set of assignments

**Evaluation of the validity of using artifacts to make inferences regarding a teacher’s CKT**

Following the scoring of the instructional artifacts and final assessments, the scores were used to investigate the validity of the each of the inferences of the validity argument. In the following sections, each of the inferences are evaluated.

**The Scoring Inference**

**Reliability**

The scoring inference concerns whether the scoring rules were applied accurately and consistently (Kane, 2006; Bell et al., 2012). To assess whether the scoring rules were applied accurately and consistently, we first assessed the inter-rater reliability of the codes and scores produced by the protocol. To assess the reliability of the descriptive dimensions, agreement was calculated between the two raters. With regard to assessment
and assignment context, the agreement between the two raters for the assessment was 79 percent, while the agreement for the individual assignments was 84 percent. With regard to the content addressed by the assignments and assessments, the agreement on the assessment was 74 percent, and the agreement for the individual assignments was 78 percent. The agreement on the science practices dimension was 80% for the assessment and 83% on the assignments.

To assess the reliability of the scores produced by the scored dimensions, we computed both the agreement between the two raters and weighted Cohen’s Kappa (Cohen, 1968). Cohen’s Kappa is a measure of agreement between two raters after accounting for agreement by chance, while weighted kappa takes into account the distance between the two scores (two raters that produce scores that are one off from each other will result in a higher Kappa value than scores that were two off, for example). Tables 1 and 2 present the reliability statistics for the scored dimensions of both the assignment scoring and the assessment scoring.

**Table 1**

<table>
<thead>
<tr>
<th>Dimensions</th>
<th>% Exact Agreement</th>
<th>% Scores within 1</th>
<th>Weighted K</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analysis</td>
<td>84</td>
<td>96</td>
<td>.76</td>
</tr>
<tr>
<td>Math</td>
<td>84</td>
<td>96</td>
<td>.77</td>
</tr>
<tr>
<td>Representations</td>
<td>80</td>
<td>93</td>
<td>.82</td>
</tr>
<tr>
<td>Total Score</td>
<td>96</td>
<td>1</td>
<td>.94</td>
</tr>
</tbody>
</table>
Table 2

**Scored Dimension Reliability – Instructional Artifacts**

<table>
<thead>
<tr>
<th></th>
<th>% Exact Agreement</th>
<th>% Scores within 1</th>
<th>Weighted K</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analysis</td>
<td>67</td>
<td>97</td>
<td>.54</td>
</tr>
<tr>
<td>Math</td>
<td>83</td>
<td>94</td>
<td>.80</td>
</tr>
<tr>
<td>Representations</td>
<td>77</td>
<td>98</td>
<td>.76</td>
</tr>
<tr>
<td>Total Score</td>
<td>80</td>
<td>98</td>
<td>.70</td>
</tr>
</tbody>
</table>

With regard to the scored dimensions, the exact agreement between the raters for the final assessment scoring was over 80 percent for each of the dimensions, and over 95% of the scores fell within 1 score point for each dimension. Similarly, for the instructional artifact scoring, the simple agreement between the raters, while slightly lower than the final assessment scoring, indicates a high level of agreement between the raters. The weighted Kappa levels also indicate a substantial level of agreement when the distribution of the scores is taken into account (Shavelson & Webb, 1991; Viera & Garret, 2005).

**Distribution of Codes and Scores.**

Along with the reliability of the scores produced by the protocol, when assessing the scoring inference, we should examine the distribution of scores produced by the protocol. If only a portion of the scale is being used when scoring, it is possible that the protocol is not designed to capture the differences between the items and tasks on the artifacts. Table 3 provides the mean and range for the proportion of items on each teacher’s final assessment that were context free, had a physics context, or had a real-world context and Table 4 provides same analysis for the number of instructional artifacts with each context (out of 3 artifacts submitted). Number of each context was used for instructional artifacts due to each teacher submitting the same number (3). Table 5 shows
the mean and range for the number of content targets addressed by the final assessments and instructional artifacts. With regard to science practices, we found only 3 assessments that contained any science practices other than using multiple representations, and a similar result among the instructional artifacts.

**Table 3**

*Final Assessment – Context, Descriptive Statistics*

<table>
<thead>
<tr>
<th>Context</th>
<th>Minimum Proportion</th>
<th>Maximum Proportion</th>
<th>Mean Proportion</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Free</td>
<td>0</td>
<td>.67</td>
<td>.16</td>
<td>.20</td>
</tr>
<tr>
<td>Physics Context</td>
<td>.30</td>
<td>1.00</td>
<td>.71</td>
<td>.21</td>
</tr>
<tr>
<td>Real World Context</td>
<td>0</td>
<td>.50</td>
<td>.13</td>
<td>.15</td>
</tr>
</tbody>
</table>

**Table 4**

*Instructional artifacts – Context, Descriptive Statistics*

<table>
<thead>
<tr>
<th>Context</th>
<th>Minimum Proportion</th>
<th>Maximum Proportion</th>
<th>Mean Proportion</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Free</td>
<td>0</td>
<td>1</td>
<td>.07</td>
<td>.27</td>
</tr>
<tr>
<td>Physics Context</td>
<td>1</td>
<td>3</td>
<td>2.22</td>
<td>.75</td>
</tr>
<tr>
<td>Real World Context</td>
<td>0</td>
<td>2</td>
<td>.70</td>
<td>.67</td>
</tr>
</tbody>
</table>

**Table 5**

*Content Addressed – Descriptive Statistics, Number of Targets Addressed by the Final Assessments and Instructional Artifacts*

<table>
<thead>
<tr>
<th></th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Final Assessment</td>
<td>4</td>
<td>17</td>
<td>8.89</td>
<td>3.13</td>
</tr>
<tr>
<td>Instructional Artifacts</td>
<td>0</td>
<td>10</td>
<td>4.15</td>
<td>2.45</td>
</tr>
</tbody>
</table>
When examining the distribution of the context of the items on the final assessments and instructional artifacts, we found that the proportions (for the final assessment) and the number of artifacts (for the instructional artifacts) that based in each context have a distribution a range that indicates a variation in the contexts of teachers’ assessments and instructional artifacts. Similarly, for the content addressed by the final assessments and instructional artifacts, we see a range from few or no targets addressed to 17 targets addressed by a final assessment and 10 addressed by a set of instructional artifacts (out of 30 possible targets).

For the scored dimensions, we calculated the mean and range for each of the dimensions for the assessment scoring and the instructional artifact, as well as the total score (the average of the dimension scores). The descriptives are shown in table 6 and table 7 respectively.

**Table 6**

*Scored Dimensions - Descriptive Statistics for the Final Assessments*

<table>
<thead>
<tr>
<th></th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analysis</td>
<td>1.5</td>
<td>4</td>
<td>2.50</td>
<td>.77</td>
</tr>
<tr>
<td>Math</td>
<td>1</td>
<td>4</td>
<td>2.91</td>
<td>.76</td>
</tr>
<tr>
<td>Representations</td>
<td>1</td>
<td>4</td>
<td>2.29</td>
<td>.86</td>
</tr>
<tr>
<td>Total Score</td>
<td>1.67</td>
<td>4</td>
<td>2.57</td>
<td>.64</td>
</tr>
</tbody>
</table>

**Table 7**

*Scored Dimensions - Descriptive Statistics for the Instructional Artifacts.*

<table>
<thead>
<tr>
<th></th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analysis</td>
<td>1.33</td>
<td>3.67</td>
<td>2.57</td>
<td>.55</td>
</tr>
<tr>
<td>Math</td>
<td>1.33</td>
<td>4.00</td>
<td>2.59</td>
<td>.72</td>
</tr>
<tr>
<td>Representations</td>
<td>1.00</td>
<td>3.67</td>
<td>2.48</td>
<td>.78</td>
</tr>
<tr>
<td>Total Score</td>
<td>1.39</td>
<td>3.56</td>
<td>2.55</td>
<td>.55</td>
</tr>
</tbody>
</table>
When examining the distribution of the scores, each dimension for both the final assessment scoring and the instructional artifact scoring have minimums near the bottom of the scale, and maximums near the top. The range is slightly reduced for the instructional artifact scoring because it is the average of dimension scores for each of the artifacts in the set. The mean for each of the dimensions falls near the middle of the 1 – 4 scale.

**Independence of Dimensions**

A final concern when evaluating the scoring inference and the scores produced by the protocol is whether the dimensions of the protocol are independent, or measuring different aspects of tasks and problems within the artifact. While all the dimensions of the protocol should be somewhat related, if the scores from on each dimension are highly correlated, that is evidence that the dimensions are picking up on similar aspects of the problems or tasks. The correlations between the dimensions for the final assessment scoring and the instructional artifact scoring are shown in Table 8 (final assessment) and Table 9 (instructional artifacts). The final column of the tables shows the correlation between the row, and the overall averaged scores of the other two dimensions.

**Table 8**

*Correlation Between scored Dimension scores – Final Assessment*

<table>
<thead>
<tr>
<th></th>
<th>Analysis</th>
<th>Math</th>
<th>Rep</th>
<th>Overall W/O Dimension</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analysis</td>
<td></td>
<td>.18</td>
<td>.57**</td>
<td>.43*</td>
</tr>
<tr>
<td>Math</td>
<td></td>
<td></td>
<td>.60**</td>
<td>.46*</td>
</tr>
<tr>
<td>Representations</td>
<td></td>
<td></td>
<td></td>
<td>.76**</td>
</tr>
</tbody>
</table>

**p<.01, *p<.05**
Table 9

Correlation Between Scored Dimension Scores – Instructional Artifacts

<table>
<thead>
<tr>
<th>Analysis</th>
<th>Math</th>
<th>Rep</th>
<th>Overall W/O Dimension</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analysis</td>
<td>.47*</td>
<td>.55**</td>
<td>.60**</td>
</tr>
<tr>
<td>Math</td>
<td></td>
<td>.46*</td>
<td>.53**</td>
</tr>
<tr>
<td>Representations</td>
<td></td>
<td></td>
<td>.58**</td>
</tr>
</tbody>
</table>

**p<.01, *p<.05

For both the final assessment scoring and the instructional artifact scoring, each of
the dimensions are significantly correlated, which we would expect, as they are each
related to the demand of the problems and tasks on the assignments and assessments.
However, the correlations are still quite moderate, indicating that the dimensions are
picking up on different aspects related to assignment demand.

The Generalization Inference

The generalization inference concerns whether the scores produced by the
protocol based on the sample of data (the final assessment and the instructional artifacts)
are representative of the universe of artifacts created by the teacher (Kane, 2006). While
the assessment of every artifact a teacher creates for instruction would possibly yield a
more accurate estimate of the teacher’s ability to develop and select assessments and in
turn their CKT, we are only able to score a portion of their artifacts. Ideally, the sampled
data would be representative of the teachers’ ability.

To assess of representativeness of the sample, we conducted a G-Study
(Shavelson & Webb, 1991). A G-study analyzes the scores produced by the protocol to
estimate the sources of variance within those scores. For example, when examining the
dimension scores on the final assessment, variation in the scores could be due to the
teacher, or it could be due to the raters scoring the final assessment. Ideally, a significant portion of the variance should be attributed to the teacher.

Similarly, for the instructional artifact scoring, where raters scored each of the artifacts in the set separately and the total set scores for each dimension were produced by aggregating the dimension scores from each artifact, the sources of variance in the dimension scores for the set could be attributed to the teacher, the variance in the artifact scores or the rater. As with the assessment, a high level of variance attributed to the teacher would indicate that we are assessing differences in a teacher’s ability to design and select artifacts for instruction and in turn their CKT. However, if there is a high level of variance attributed to the individual assignments, then that would indicate the scores on the individual assignments are not a reliable estimate of a teachers’ ability to develop and select assignments for instruction because of the variance in scores from one assignment to the next for each teacher.

To assess the sources of variance in the final assessment scoring, we created a model with 4 possible sources of variance: the teacher, the rater, the interaction of the rater and the teacher, and any residual error. Table 10 indicates the percent of the variance in each of the dimension scores accounted for by each of the sources for the final assessment.

Table 10

Variance Components – Final Assessment (% Variance Accounted for)

<table>
<thead>
<tr>
<th>Analysis</th>
<th>Teacher</th>
<th>Rater</th>
<th>Teacher x Rater</th>
<th>Residual</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mathematics</td>
<td>88</td>
<td>&lt;1</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>Representations</td>
<td>88</td>
<td>2</td>
<td>10</td>
<td>0</td>
</tr>
</tbody>
</table>

Variance Components – Final Assessment (% Variance Accounted for)
Regarding the variance in the final assessment scores, a large amount of variance was attributed to the teacher. As there were few possible sources of variance, and the raters scored with high reliability, so the high levels of variance associated with the teacher are to be expected.

To assess the sources of variance in the instructional artifact scoring, we created a model with 5 possible sources of variance: the teacher, the rater, the assignment within the teacher, the interaction of the rater and the teacher, and any residual error. Table 11 indicates the percent of the variance in each of the dimension scores accounted for by each of the sources.

Table 11

| Variance Components – Instructional Artifacts (% Variance Accounted for) |
|-----------------|-----------------|-----------------|-----------------|-----------------|
| Teacher | Artifact | Rater | Rater x Teacher | Residual |
| Analysis | 14 | 58 | <1 | 7 | 20 |
| Mathematics | 12 | 53 | <1 | 7 | 28 |
| Representations | 39 | 40 | <1 | 5 | 16 |

When examining the sources of variance for the instructional artifacts, we found that while there is some variance in the set scores attributable to the teacher, most of the variance is attributed to individual artifacts. This indicates a high level of variability in the assignments that the teachers use during instruction, which reflects the normal day to day variation in instruction.

The Extrapolation Inference

The extrapolation inference concerns the relationship between the scores produced by the protocol and the scores produced by other measures intended to assess similar constructs (Kane, 2006). In the case of the CKT-E project, along with the artifact
protocol, we created a test-based measure of CKT-E. If teachers enact CKT while developing and selecting artifacts for instruction, then there should be a relationship between the scores produced by the artifact protocol and the scores on the CKT-E assessment. Teachers who score higher the CKT-E assessment should produce artifacts that are more complex, and therefore should score higher on the artifact protocol.

To assess the relationship between the scores produced by the artifact protocol and the CKT-E assessment, we analyzed the correlations between each the descriptive dimensions and scored dimensions and CKT-E assessments scores. For the descriptive dimensions, to assess the relationship between the context of the assessment, we correlated the proportion of items on each assessment that were either context free, had a physics context or a real-world context to teachers’ CKT-E scores. For the assignments, we correlated the proportion of the assignments that had each context. There was a significant negative correlation between teachers’ CKT-E scores and the proportion of items on their final assessment that were context free \( r_p(28) = -.38, p < .01 \) indicating that teachers with lower CKT-E scores tended to include more context free items on their final assessments. There were no significant correlations for physics context items or real-world context items on the final assessment and no correlations between the context of the instructional artifacts and CKT-E.

Regarding the relationship between the content of the assignments and assessments and teachers’ CKT-E scores, we correlated the number of targets addressed by teachers’ assessments and assignments, respectively, and their CKT-E assessment scores. The correlations are shown in table 12. A more in-depth analysis of the content of teachers’ final assessments can be found in Zisk, Gitomer, Bell, and Etkina (2019).
Table 12

*Relationship Between the Content Addressed by Final Assessments and Instructional Artifacts and CKT-E Assessment Scores*

<table>
<thead>
<tr>
<th>Instructional Artifacts</th>
<th>Teachers CKT-E assessment scores and Number of Targets Addressed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Final Assessment</td>
<td>.63**</td>
</tr>
</tbody>
</table>
**p<.01

For the final descriptive dimension, science practices, we found no correlations between the number of science practices addressed on teachers’ assessments and instructional artifacts and their CKT-E score. This lack of relationship was driven by the absence and minimal presence of science practices on teachers’ assessments and instructional artifacts.

To assess the relationship between each of the scored dimensions and CKT-E, we correlated the average score for each of the dimensions (as well as the overall score, the average of the three dimensions) and teachers’ CKT-E scores for both the assessment and instructional artifacts. Table 13 shows the correlation between the scored dimension scores for both the assignments and assessments and the CKT-E assessment scores.

Table 13

*Relationship Between Artifact Scored Dimension Scores and CKT-E Assessment Scores*

<table>
<thead>
<tr>
<th>Final Assessment and CKT-E assessment score</th>
<th>Instructional Artifacts and CKT-E assessment score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analysis</td>
<td>.27</td>
</tr>
<tr>
<td>Math</td>
<td>.35</td>
</tr>
<tr>
<td>Reps</td>
<td>.31</td>
</tr>
<tr>
<td>Total</td>
<td>.39*</td>
</tr>
</tbody>
</table>
**p<.05

We did not find any significant relationships between each of the dimension scores (analysis, math and representations) and the scores on the CKT-E assessment.
However, there was a significant relationship between the total final assessment score (the average of the 3 dimension scores) and the CKT-E assessment scores. With regard to the instructional artifacts, the only significant relationship was between the analysis dimension and the CKT-E score.

**Discussion**

The CKT-E framework hypothesizes that when teaching, teachers use their knowledge of the important targets within the domain of energy and their knowledge of how to help students achieve these targets (the tasks of teaching). Therefore, the framework predicts that teachers with different content knowledge for teaching should carry out instruction in different ways. One critical task of teaching is the ability to design and select assignments and assessments for instruction. In this study, we set out to investigate the relationship between the quality and demand of the instructional artifacts that physics teachers design for their unit on energy in the context of mechanics and their knowledge for teaching the same content (CKT-E) as measured by the CKT-E assessment (Etkina et al., 2018). In doing so, we developed a protocol to measure the quality and demand of assignments and assessments in physics and investigated the validity of the claim that teachers enact content knowledge for teaching when designing or selecting artifacts for instruction.

We first investigated whether the quality and demand of the assignments and assessments that physics teachers use vary across a unit of instruction. As teachers use both knowledge of the targets within the domain of energy and the knowledge of how to design and select assignments and assessments to address those targets, there should be variation across teachers with respect to the assignments they use during instruction. The
wide range in the demand scores of each of the three dimensions indicates that there is significant variation in assignment and assessment demand across teachers. Students in different classrooms are subject to a wide range of expectations. In some classes, instruction, as indicated by the assignments and assessments, is more traditional, with a focus on rote memorization of terms and simple “plug and chug” math-oriented problems. This type of instruction was found to be quite common in science classrooms throughout the country (Hofstein & Lunetta, 2004; Roth et al., 2006) and this is exactly the type of instruction that the Framework for K-12 Science Education (Quinn et al., 2012) and the NGSS (NGSS Lead States, 2013) attempted to improve upon. However, there were several classrooms that set significantly more demanding expectations, where students were engaged in deep analysis of phenomena and were using science practices. The instruction in these classes was more in line with recent science education standards and would fall more towards the high level of demand on other task analysis frameworks (Stein, et al., 2001; Tekkumru-Kisa et al., 2015).

This difference in the demand of the assignments and assessments could be stem from several sources. First, it is possible that different teachers have different knowledge for teaching that results in them designing or selecting more or less demanding assignments. For example, consider teachers who use less demanding assignments and assessments that focus on facts and “plug and chug” math oriented problems. It is possible that these teachers use these assignments due to a limited knowledge of the targets within the domain of energy. A less robust knowledge of the targets could result in instructional tasks that are focused on less demanding content such as the knowledge of definitions. It is also possible that teachers with less demanding assignments have an
in-depth knowledge of the targets within energy but do not have the knowledge to carry
out the task of designing assignments and assessments to address such targets.
Conversely, consider the teacher who designs or selects very demanding assignments that
address the important targets within energy. To develop such instructional activities, the
teacher must possess an in-depth knowledge of the targets and the knowledge to carry out
the task of designing assignments to meet those targets. This deficiency in knowledge
could also explain the absence of science practices in the assignments and assessments
provided to the students. In order to create items that include the science practices, the
teacher must have knowledge of the important learning targets, the science practices, and
how the science practices can be used to develop knowledge of the targets. It is much
easier create and score a more traditional problem than to create an item that requires, for
example, the students to develop an experiment or analyze data in order to find a
mathematical relationship. Pellegrino (2013) addressed this difficulty and outlined the
challenges in developing assessments that included science practices and were in-line
with the NGSS.

Beyond teacher knowledge, it is also possible that the teachers are simply given
the assignments and assessments to use during instruction or are following an assigned
curriculum and as such, have little to no influence on the assignments they use in class.

Despite the wide range of demand across teachers, it is important to note though
that even among the assignments and assessments of the teachers that had the highest
levels of demand, there were tasks and assignments that were not as demanding. This is
quite evident when examining the sources of variance in the demand scores of the
instructional artifacts. While there was some variation at the teacher level, the majority of
the variation was attributed to the assignments themselves. This indicates that within
assignments of each teacher, there is a wide range of demand. Some assignments are
much easier than others. This variation is not necessarily a negative. It is not reasonable
to expect every assignment to be at the high level of demand. Additionally, to be
successful with more demanding tasks, students first need to find success with those that
are less demanding. For example, it is possible that when students are starting to learn the
material, assignments might be less demanding, and even later in the unit, the teacher
might include less demanding tasks for a number of reasons, including giving the
students who might be struggling an opportunity to succeed. However, this variation
needs to be explored further, as in this study, we only collected three assignments.
Investigating the variation across a whole unit of assignments might lend insight to how
practice varies in the day-to-day flow of the classroom, which is something that has not
been fully explored in the context of artifacts or classroom observations.

Along with investigating the variation in the quality and demand of teachers’
assignments and assessments, we investigated the relationship between the demand of
teachers’ assignments and assessments and their knowledge for teaching energy (CKT-E)
as measured by the CKT-E assessment to evaluate the validity of the claim that teachers
with different CKT-E enact practice in different ways when designing and selecting
artifacts for instruction. To explore the relationship between knowledge and practice and
to investigate the claim that CKT-E is related to instructional practice, we employed the
validity argument approach (Kane, 2006). The reliability of the scores produced by the
protocol and the variance in the scores both indicated that the scoring rules of the
protocol were applied correctly, and that the protocol was able to capture differences in
the demand and quality of the assignments and assessments. The sources of the variance in the demand scores could be attributed to the teacher for both the assessment and assignments, though there was a large variation in the scores across instructional artifacts, as was discussed previously. The appropriate application of the scoring rules and the attribution of variance to the teacher support the use of the scores produced by the protocol to make claims regarding differences in teaching practice.

The final inference we evaluated concerned the relationship between teachers’ knowledge and the quality and demand of their assignments and assessments. When trying to assess the claim the teachers with different CKT-E enact practice in different ways, the relationship between knowledge and practice is essential. We found that teachers with higher CKT-E have more demanding assessments overall than those with lower CKT-E, require deeper levels of analysis on their assignments, address more content on both their assignments and assessments, and tend to have fewer problems and tasks that are context free. These findings indicate that teachers with higher CKT-E are creating assignments and assessments that are more demanding and address more targets within energy. The lack of a relationship between the math, representations and overall demand of the assignments could be explained by the varied nature of assignments throughout a unit, though the significant relationship between teachers’ CKT and the analysis dimension provides evidence that teachers with higher CKT may produce assignments that require the students to engage in more complex reasoning, even if the mathematical and representational demands of the assignments are not as high.

The results of the validity investigation indicate that teachers with different CKT-E do enact practice in different ways. Teachers with lower CKT-E tend to create lower
demand assignments and address fewer energy targets, while those with higher CKT-E create high demand assignments that address more targets. The assignments and assessments used by teachers with more developed CKT-E are more in-line with the type of instruction that is expected within the US science classroom (NGSS Lead States, 2013; Quinn et al., 2012) and emphasize more “active-thinking” on the part of the students that has been associated with learning gains (Minner et al., 2010).

The relationship of teachers’ CKT-E and the demand of their assignments support the assertion that teachers with more robust knowledge for teaching energy have first, a more in-depth knowledge of the targets within the domain of energy, as their assignments and assessments address more energy targets and second, a more developed knowledge of how to develop more demanding instructional tasks that meet those targets. The alignment of higher CKT-E teachers’ assignments and assessments to the instruction described by the Framework for K-12 Science Education (Quinn et al., 2012) and the NGSS (NGSS Lead States, 2013) may also indicate that they have more knowledge of the current standards or/and that they have the knowledge to meet those standards. These results lend support to the construct of CKT as the knowledge that teachers use to carry out the day-to-day tasks of teaching (Ball et al., 2008; Gess-Newsome, 2015; Magnusson, et al., 1999; Mikeska, et al., 2017; Phelps et al., 2014), as those teachers with more advanced knowledge design more demanding instructional tasks.

Along with supporting the construct of CKT, the results of this study also have implications for the advancement of policy regarding high-quality science instruction and professional development. In this study, we found that the teachers with higher scores on the CKT-E assessment were more likely to provide their students with tasks that were in-
line with current standards and research into high quality instruction. This finding indicates that a population of teachers does not have the knowledge to carry out the instruction necessary to help students meet the learning targets expected by the current science standards. To ensure that teachers are able to carry out the expected quality instruction, professional development needs to help teachers learn how to create more demanding tasks that are aligned with those standards for their students.

Though this study has provided evidence for the relationship between a teacher’s knowledge for teaching and specific aspects of classroom practice, it has several limitations. First, it is possible that teachers are using assignments and assessments provided to them by their school. If that is the case, it is possible that the teacher had little influence on the design of the tasks. Furthermore, some teachers do not necessarily believe that unit tests are the most effective way of assessing students or the best use of class time. This could possibly lead them to create less demanding or less time-consuming assessments. Additionally, teachers often make many considerations when determining the assignments and assessments they use during instruction. For example, if the teacher knows class time will be short for some reason, they may choose to include a shorter task in lieu of a more demanding assessment. This lack of demand would not necessarily be reflective of their CKT. Finally, regarding the assignments, we only collected three out of the possible numerous assignments and assessments that could have been used during a unit and gave teachers the ability to choose what artifacts to provide (outside of the final assessment). This did not allow us to have a full sense of the overall demand of the assignments used throughout the unit nor did it allow us to capture fully how the demand of the assignments varied throughout the unit.
References


Chapter 3: The Relationship of Teachers’ Content Knowledge for Teaching Energy and the Learning Targets for Instruction in Physics

Abstract: Content Knowledge for Teaching (CKT) describes the knowledge that teachers have for teaching a particular subject. As such, there should be a relationship between teachers’ CKT and their classroom practice. In this study we focus specifically on the relationship between teachers’ CKT and the learning targets they set for instruction as measured by their instructional goals and their unit assessments. As part of a larger effort to create and validate a measure of content knowledge for teaching in physics, specifically energy in the context of mechanics, we investigate the depth and breadth of teachers’ goals for their energy unit and the learning targets addressed by their unit assessments as well as the alignment between teachers’ goals and assessments. We then assess the relationship between the teachers’ learning targets and their Content Knowledge for Teaching Energy (CKT-E) as measured by an assessment of teacher’s knowledge of teaching energy. While all teachers in the study addressed the same breadth of learning targets, teachers with high CKT-E scores addressed the targets at a deeper level. The findings provide evidence of a relationship between a teacher’s knowledge for teaching a particular subject and specific teaching practice within that subject area.

Introduction

Attempting to understand, characterize, and capture the knowledge that enables teachers to carry out the work of teaching has long been a subject of inquiry and research (Gitomer & Zisk, 2015). Content Knowledge for Teaching (CKT) describes teacher knowledge as not only knowledge of the domain one is teaching but also knowledge of how to help students learn within the domain, knowledge of students, and the curriculum (Ball et al., 2008). It is with this knowledge that teachers carry out the daily work of
teaching, and as such, CKT is fundamentally based on the conception that teacher knowledge should be related to teaching practice. Teachers with different content knowledge for teaching should enact practice in different ways. While there has been work to define CKT within different domains (Gess-Newsome, 2015; Magnusson et al., 1999; Mikeska et al., 2018; Mikeska et al., 2017; Park & Oliver, 2008; Park & Suh, 2015; Phelps et al., 2014), and to relate teachers CKT to scores on other measures of practice (Hill & Charalambous, 2012; Hill et al., 2007; Hill et al., 2005; Hill et al., 2012; Hill et al., 2005; Hill et al., 2012), little is known about the particulars of how practice varies among teachers with different content knowledge for teaching.

In science, one area of teaching practice that has received particular attention has been the knowledge and skills that teachers are addressing and helping students to develop during instruction. In the US, and in other countries around the world, science instruction has begun to stress developing deep knowledge of science through drawing connections between concepts and ideas and building students’ ability to learn through engaging in the practices of science (Martin et al., 2004; Tekkumru-Kisa et al., 2015). The shift in instruction was outlined by the Framework for K-12 Science Education (National Research Council, 2012) that called for teachers to provide time for students to develop meaningful understanding of the science through deep exploration, and the Next Generation Science Standards (National Research Council, 2013) that called for the development of science concepts through the engagement in the practices of science. Under the framework and the NGSS, learning science now means more than just developing disciplinary core ideas (e.g., laws of physics), but also developing an
understanding of concepts that cut across the science disciplines and the practices of science through which the ideas and concepts are developed.

As new standards have been implemented across the country (currently 42 states and the District of Columbia have either adopted the NGSS or have implemented standards based on the K-12 science framework (NASTA, 2014), classroom teachers have been responsible for designing and implementing lessons that meet those standards. However, in order to teach science at the level expected by the new standards, teachers must possess the knowledge to teach to those standards effectively, and yet little is known about the relationship between teacher knowledge and how they enact instruction. Do teachers have the knowledge of the learning targets for students expected under the new standards, and are they able to develop tasks and assessments that address those targets? Developing an understanding of how teachers’ knowledge for teaching influences their targets and goals for instruction can both expand the theory of content knowledge for teaching and assist those who educate teachers to target the knowledge necessary to develop teachers’ abilities to address learning targets that are aligned with the current science standards.

In this study, we examine the learning targets that high school physics teachers address in their instruction on energy in the context of mechanics. Learning targets that teachers set within a unit can vary in many ways. First, teachers can address more or fewer topics, and second, teachers can address the same topics but at different levels of depth, such as focusing on just terms and equations or in contrast, helping students construct a deeper understanding of the core ideas in the discipline and important concepts through exploration and the engagement with the science practices. In the study
reported in this paper we examine the relationship between teachers’ learning targets for their energy unit and their content knowledge for teaching energy.

This study is based on a research initiative to define a framework of content knowledge for teaching energy (CKT-E), develop an assessment of that knowledge, and examine the relationship between teachers’ CKT-E and how they enact tasks of teaching to support students’ learning of energy (Etkina et al. 2018). In this paper, we first examine the breadth and depth of the learning targets that high school physics teachers address in their instruction on energy in mechanics through the goals they set for instruction and their unit assessments. We ask what core ideas teachers address during their energy unit (breadth), and whether they address them at a surface level (the memorization of terms and definitions, for example) or at a more in-depth level (depth). Second, we examine how the breadth and depth of the learning targets addressed within a teacher’s energy unit is related to their CKT as measured by the test we designed (Etkina et al. 2018; Phelps et al., 2020). Finally, the teachers’ stated goals for instruction and the targets addressed on their final unit assessment are compared to determine alignment between them. The study asks the following research questions:

1) What is the relationship between the breadth and depth of teachers’ goals for instruction and their content knowledge for teaching energy?

2) What is the relationship between the breadth and depth of the learning targets that teachers assess on their final unit assessment in energy and their content knowledge for teaching energy as measured by the CKT-E assessment?
3) What is the alignment between the goals that teachers set for instruction and their unit assessments in energy, and how is this alignment related to a teacher’s content knowledge for teaching energy as measured by the CKT-E assessment?

Content Knowledge for Teaching

The construct of content knowledge for teaching (Ball et al., 2008) describes the knowledge that teachers have that allows them to perform daily tasks in the classroom with the goal of helping students learn. The construct is based on the assertion that the knowledge needed to help students learn goes beyond knowledge of the subject matter of a content expert and includes specialized knowledge related to teaching and learning that is central to the work that teachers face in the classroom. This knowledge includes knowledge of the subject matter and as well as knowledge of the curriculum, of students, and how students think about the subject matter. To illustrate the difference between more common content knowledge and the specialized knowledge needed for teaching, consider the following example: Those knowledgeable in physics content know Newton’s second law and commonly describe it as \( F = \text{Mass} \times \text{Acceleration} \). The knowledge that is common to most who have taken a physics course is that there is a relationship between the forces exerted on an object, the objects mass and its acceleration. However, those with more specialized knowledge for teaching understand that \( F = ma \) is not productive and is not entirely correct as it is only true when there is just one force exerted on an object. Therefore, they think of Newton’s second law as \( \text{Acceleration} = \frac{\text{Sum of the Forces}}{\text{Mass}} \). Structured this way, the equation not only allows to accurately determine the acceleration of an object but also describes the cause-and-effect relationship between the object’s acceleration and its mass and emphasizes that it is the sum of the forces
exerted on the object that matters not just one force. They also understand why it is more productive to use this representation of the Newton’s second law in the classroom (Etkina et al., 2018) This example helps describe the specialized knowledge that teachers need to call upon during the work of teaching.

More recently, researchers have expanded upon the conceptualization of CKT to develop practice-based conceptions of CKT that describe content knowledge for teaching as enabling teachers to carry out keys tasks or works of teaching when planning and carrying out instruction and assessment (Gess-Newsome, 2015; Magnusson et al., 1999; Mikeska et al., 2018, 2017; Park & Oliver, 2008; Park & Suh, 2015; Phelps et al., 2014). These frameworks each include key tasks related to planning of instruction, to carrying out instruction in the classroom and to assessing student learning, and describe teachers’ content knowledge of teaching as being essential to their ability to carry out those tasks.

Central to the theory of CKT is the idea that teachers with more robust knowledge for teaching a particular subject carry out more effective teaching practice (Hill et al., 2007; Schilling & Hill, 2007) and, as such, there should be a relationship between teachers’ CKT and their classroom practice and differences in how teachers with different CKT carry out the work of teaching a particular subject or topic. Previous efforts to connect knowledge and practice have compared teachers’ scores on assessments of CKT and their scores on general observation instruments (Hill & Charalambous, 2012; Hill et al., 2007; Hill et al., 2005, Hill et al., 2012) or students’ scores on standardized assessments (Hill & Charalambous, 2012; Hill et al., 2005; Hill et al., 2012). Keller, Neumann, and Fischer (2017) also examined the influence of teachers’ knowledge for teaching physics on student achievement and found a positive correlation between
teachers’ knowledge and the achievement of their students. These studies have provided insight into relationships between CKT and other measures associated with teaching quality and effectiveness. However, we still do not understand the particulars of teaching practice that underlie these patterns of relationships.

One area in particular that has not been explored is the relationship between teachers’ CKT and the learning targets that they address during instruction. If teachers with varying levels of CKT have more or less developed specialized content knowledge, then their level of knowledge should be reflected in the content of their instruction. Those with more developed CKT should have the resources to address more and possibly more robust content in their classrooms, while those with less developed CKT should not be expected to address content as robust.

**Content Knowledge for Teaching Energy Framework (CKT-E)**

The work in this paper is part of a larger effort to define the framework of content knowledge for teaching energy (CKT-E) and to create both an assessment and practice-based measures of content knowledge for teaching physics in the domain of energy with an overarching goal of developing a nuanced understanding of the relationship between CKT-E and classroom practice. Our project focused on one content domain in order to develop a coherent theoretical link between CKT and practice.

Central to this effort was the development of the framework of CKT-E. The framework was based on the assertion that teachers need to have knowledge of the important student learning targets within the domain of learning and the knowledge to enact tasks of teaching to support students in meeting those targets. However, the knowledge of learning targets is not always sufficient to interpret the range of student
ideas and responses that can arise during instruction. This additional knowledge described as *Horizon Knowledge*. As such, the framework is comprised of three components, the *Tasks of Teaching*, the *Targets for Students*, and *Horizon Knowledge* within the domain of energy (Etkina et al., 2018) (See Figure 1).

**Figure 1**

*CKT-E Framework*

![CKT-E Framework Diagram]

The tasks of teaching describe the activities that teachers engage in that help to support and promote student learning (Ball, 2000). Within our framework, these tasks include: (I) anticipating student thinking around science ideas; (II) designing, selecting, and sequencing learning experiences and activities; (III) monitoring, interpreting, and acting on student thinking; (IV) scaffolding meaningful engagement in a science learning community; (V) explaining and using examples, models, representations, and arguments to support students’ scientific understanding; and (VI) using experiments to construct,
test, and apply concepts (Etkina et al., 2018). In this study, we are specifically interested in the teachers’ ability to carry out two key tasks of teaching, defining goals for instruction (task II), and developing assessments that assess the goals for instruction (task IV).

Along with the tasks of teaching, the second component of the CKT-E framework are the Student Energy Targets. The targets focus on specific content within the domain of energy in mechanics and articulate the important concepts, skills and knowledge representations that are critical to the domain and are in line with existing research on the learning and teaching of energy (Etkina et al., 2018). The targets are organized into seven target categories with sub-targets within each (a complete list of the targets is shown in Figure 2). These targets go beyond simple terms, definitions, and equations related to energy, and instead focus on the disciplinary core ideas related to energy, the crosscutting concepts of systems and energy and matter, and the science practices.

The framework of CKT-E is centered on teachers’ knowledge of the content targets within the domain, and their ability to carry out the tasks related to teaching those targets. For example, in order to be able to carry out the task of teaching of setting goals for instruction, teachers need to be aware of the important learning targets within the domain. Similarly, to design assessment items teachers need to be aware of the content targets for the domain and the goals they set for instruction in order to be able to assess students’ knowledge related to those targets.
Figure 2

Student Energy Targets

A) Connection of Energy and Everyday Experiences/ The student:
1) Uses energy ideas to interpret or explain everyday phenomena
2) Recognizes the important role of internal energy in interpreting or explaining everyday phenomena

B) Choice of System/ The student:
1) Understands that the energy accounting in a phenomenon depends on the choice of system
2) Explains the relative advantage of a given system choice (i.e., relative ease of analysis)
3) Recognizes that the choice of the system determines whether springs or Earth do work
4) Identifies and differentiates between forms of energy and other physics concepts

C) Identification of and Differentiation Between Forms of Energy and Other Physics Concepts/ The student:
1) Understands that energy cannot be observed directly and knows how different forms of energy correspond to different measurable physical quantities
2) Recognizes and maintains a consistency of scale during energy analysis
3) Differentiates between energy and related ideas
4) Distinguishes between forms of energy and energy transfers

D) Transfer of Energy/ The student:
1) Understands that the energy of a system is always conserved but might not be constant
2) Recognizes that work is the way in which energy is transferred mechanically and may result in a change in temperature in some cases
3) Recognizes when to use compensatory models for tracking energy into and out of a system and when quantitative models are of limited use

E) Use of Mathematics/ The student:
1) Understands that when considering potential energy, it is the change that is important. The zero level of potential energy is arbitrary but the change is not.
2) Understands the linear and non-linear mathematical relationships between forms of energy and the factors on which they depend
3) Understands and can account for vector and scalar quantities in energy analysis
4) Understands that conservation serves as a mathematical constraint on the outcomes of possible processes
5) Understands that the mathematical analysis of energy-related processes depends on the choice of initial and final state

F) Use of Representations/ The student:
1) Selects/creates and uses appropriate verbal, mathematical, and graphical/pictorial representations to describe, analyze, and/or communicate a physical situation or process
2) Interprets different representations used to describe, analyze, and/or communicate a physical situation or process
3) Understands the relationships between different representations of the same phenomenon
4) Understands standard technical representations and language used to communicate energy-related ideas

G) Use of NGSS Science Practices (NRC, 2013)/ The student:
1) Asks questions and defines problems
2) Develops and uses models
3) Plans and carries out investigations
4) Analyzes and interprets data
5) Uses mathematics and computational thinking
6) Constructs explanations and designs solutions
7) Engages in arguments from evidence
8) Obtains, evaluates, and communicates information
CKT-E and the “work” of teaching energy

To illustrate CKT-E as including the knowledge of student learning targets and knowledge of the tasks of teaching, consider a teacher trying to help students develop understanding of energy constancy and conservation and the difference between the two. In energy, constancy means that the total energy of a system does not change (remains constant in time), while conservation means that any change in the total energy of the system can be accounted for – the total change in energy of the system will equal the energy transferred into or out of the system (Seeley et al., 2019). Energy conservation is a disciplinary core idea and embedded within the energy and matter cross-cutting concept of the NGSS. To help students learn this idea, the teacher must help the students develop the following elements: (1) the energy of any system is always conserved but not necessarily constant, (2) the energy of a system can remain constant in some processes (Energy Target D1) and (3) sometimes one system makes the analysis of a process easier than another (Energy Target B2). In order to enact the task of teaching of setting a goal for instruction related to conservation and constancy, the teachers need to have two different sets of knowledge. First, they need to possess the knowledge of constancy and conservation. Second, they need to know that when you are conducting an energy analysis of a process, you can define your system in different ways. Analyzing a phenomenon using one system we might find that no work was being done on it and the energy of the system is constant and analyzing the same phenomenon using a different system we might find that the energy changes as work is being done on the system. In this case, the energy is conserved but is not constant. Third, they need to know when some systems are more useful for analysis than others.
Once the teacher sets the above goal, they begin to choose phenomena for the students to analyze to help them develop the ideas of system choice, constancy, and conservation. As an example, consider the teacher using an example of a small falling ball hitting the ground and stopping. This process could be analyzed using several different systems and the teacher would have to choose the system carefully in order to meet the goals they set for the students. To help meet the goal of helping students to understand energy constancy, the teacher could set the system as Earth, the ball and the floor. As the ball falls, it speeds up, as the gravitational potential energy in the system is converted to kinetic energy. As it hits the ground, the ball stops, and the system no longer has kinetic energy (assuming that Earth’s kinetic energy is negligible) but both the ball and the floor warm up a little. As the system includes the floor in the system, the energy analysis is rather simple - the kinetic energy is converted into internal energy of the ball and the floor, and the energy of the system is constant. However, as the teacher wants the students to realize that some systems make the analysis of a system easier or more difficult, they would then have to present a second system with which to analyze the same process - falling ball. In this case, the teacher decides not to include the floor in the system. In this case it appears that the ball stops and the kinetic energy of the ball goes to zero and the energy of the system is not constant. Where did the energy go? One answer is that the floor did negative work on the ball, but the floor did not move, thus it could not do work. In addition, the floor warmed up a little – where did this energy come from? In this case for a student the energy of a system is not only not constant but it is not conserved either. It looks like it is not easy to analyze the process including the floor in the system but not so easy if the floor is not in the system. The students will be really
confused now – choosing different system leads to different answers! The teacher however then needs to employ a third type of knowledge, their horizon knowledge, to help the student reconcile this idea. The horizon knowledge in this case is the idea of pseudo-work. As the ball hits the ground, the ground compresses the ball slightly, doing work on its center-of-mass, and stopping the ball. To a student, it appears that no work is done on the ball because there is no obvious displacement, but with horizon knowledge, the teacher can reconcile the student’s idea and help them to see that the energy of the system is conserved. In this example, the teacher employs their knowledge of the student energy targets, their knowledge of the tasks of teaching to set goals and chose examples for instruction, and their horizon knowledge.

All three sets of knowledge (knowledge of student learning targets, knowledge of the tasks of teaching and horizon knowledge) are essential to the work of teaching. Consider a physicist who has a strong personal grasp of the energy targets, but does not have the knowledge to carry out the task of teaching of setting goals for students. If this was the case, they would not necessarily be able to develop a goal related to the target. Conversely, if a teacher believes that developing an understanding that energy conservation is an important energy target (note that just developing an understanding of conservation does not rise to the expectations of the energy target stated at the outset of this example), and their knowledge related to the energy target of conservation is simply that conservation means the energy is conserved in an isolated system (which is a common, incomplete understanding of energy conservation and constancy (Seeley et al., 2019), then their goal for instruction would look much different than that of a teacher with a strong grasp of the energy targets and students will not develop the idea that
energy is constant when no work is done on the system, and that energy is conserved, but not constant when work is done on it.

The framework and example presented here illustrate how a teacher’s CKT can influence their enactment of the everyday tasks of teaching. Teachers’ knowledge of student learning targets, the tasks of teaching and ultimately their CKT, can influence all areas of teaching practice, including the content of their goals, and the content addressed by their assessments. In this study, we explore the relationship between teachers’ CKT-E as measured by the CKT-E test (Etkina et al., 2018) and the student learning targets addressed within their energy unit through the analysis of two key tasks of teaching: setting goals for instruction and assessing student learning. The targets that they address in these two tasks will be analyzed using the lens of the student energy learning targets outlined in the CKT-E framework.

In the following section, we explain why it is important to focus on the targets of instruction.

**Student Learning Targets in Times of the NGSS**

Over the past 15 years the understanding within the science education community of what should be a learning target in science changed significantly. Reports and standards such as *Taking Science to School* (National Research Council [NRC], 2007), the *Framework for K–12 Science Education* (NRC, 2012), and the *Next Generation Science Standards* (NGSS; NRC, 2013), claimed that learning science should not be limited to the learning terms, facts, definitions and mathematical equations. Instead, students should be provided with opportunities to develop in-depth understandings of science topics through the engagement in science practices. Each of these reports and
frameworks indicated that surface-level instruction, such as the teaching of terms, definitions, and mathematical equations, was not sufficient to help students develop the core ideas of the science disciplines and concepts that cut across the science disciplines. Further, the reports also stressed that while learning those the core ideas in the disciplines, students should be engaged in the practices of science, building explanations and models through the use of evidence.

The shift towards learning the core ideas of science through the engagement in the science practices echoed the research into deep learning and the building of deep knowledge. This work described the differences between surface learning, or the memorization of terms and formulas, and deeper knowledge, which is developed through the opportunities to connections between different ideas and pieces of knowledge (Whittrock, 1994). For example, knowing the definition of work is not enough to develop an understanding of the role of work as a mechanism of transfer of energy into or out of a system. This difference was also captured Scardamalia and Bereiter (2006) when they described the difference between knowing about and knowing of. They described knowing about as declarative knowledge, or terms, formulas, the typical work of school, and knowledge of as the integration of procedural knowledge and understanding of the concepts. It is with knowledge of that students are able to solve more robust problems that go beyond simple recall and require the integration of skills and concepts. For example, in physics, knowledge about could the knowledge of the equation for work and the definition of a system in physics, while knowledge of could be analyzing a falling ball and defining the correct system to be able to calculate the amount of work being done on the ball by Earth.
The *Framework for K–12 Science Education* (NRC, 2012) and the *Next Generation Science Standards* (NRC, 2013) have provided a set of standards that call for a different approach to teaching and learning of science. Both documents argue for students learning disciplinary core ideas and crosscutting concepts (such as systems and conservation) through active engagement with the science practices (the integration of knowledge and skills). Learning science now means developing questions regarding natural phenomena, planning and carrying out investigations, analyzing data and looking for patterns, and using these patterns to develop explanations and models (NRC, 2012). This is a very different approach to learning compared to listening to lectures and doing cook-book labs.

In this study, we distinguish between surface learning targets (terms, definitions, equations) and in-depth learning targets. We define in-depth learning targets as those that meet the *student energy targets* developed for the CKT-E project. These targets, described previously, and shown in Figure 1 go beyond knowledge of terms and definitions and address the NGSS disciplinary core ideas related to energy, the crosscutting concepts, and the science practices.

**Capturing Information Regarding Teachers’ Goals and Unit Assessments**

In this study, we examine teachers’ learning targets that they set for their students learning units on energy in the context of mechanics through the goals they set for their students, and DCIs, CCCs and science practices addressed on their final assessment for the energy unit. Interviews, and particularly surveys (Kurz et al., 2010; Porter, 2002; Porter et al., 2011), have been used to capture teacher self-reports of their instructional goals as well as what they plan to teach during a particular class or unit. Teachers may
also be asked to complete an interview or survey after teaching and indicate the topics or standards they addressed, which provide information regarding the focus and learning targets of a particular lesson or unit of instruction (Kurz et al. 2010; Porter et al. 2011).

Along with surveys and interviews, researchers have attempted to gain insight into teachers’ values during a unit of instruction through examination of artifacts, or the assignments and assessments that teachers use for a given unit. Artifacts contain evidence of what students are actually asked to do in the classroom and should reflect what teachers value most in terms of knowledge and skills—or at least what they communicate to their students as valued knowledge and skills.

Previous work has shown that artifacts can be important sources of evidence of the quality of classroom instruction (Borko et al., 2007; Gitomer & Bell, 2013). Other researchers have used artifacts to measure the cognitive and intellectual demand of the classroom (Newmann et al., 2001; Wenzel et al., 2002) and as evidence of reform-oriented instruction in both mathematics and science (Borko & Stecher, 2006; Borko et al., 2005; Borko et al., 2007). Martínez, Borko, Stecher, Luskin, and Kloser (2012) focused on assessments in science as evidence of quality instruction in science. Though these studies did not focus specifically on capturing learning target through assignments and assessments, they do suggest that classroom assessments can provide valuable information of the nature of classroom instruction. This assertion is supported by the work of Baumert and colleagues (2010), who found that classroom assessments are reflective of the focus of classroom instruction.
Measuring the Alignment Between Teachers’ Goals and Unit Assessments

Along with analyzing teachers’ goals and the learning targets addressed on their unit assessments, a question that arises is if there is alignment between the goals that teachers set for instruction and their unit assessments. Curriculum and assessment guidance documents have suggested that there should be an alignment between teachers’ instructional goals, instruction, and the assessments teachers use (The Center on Standards and Assessment Implementation, 2018; Wiggins & McTighe, 2005). Further, alignment has been shown to be positively related to student achievement (Gamoran, et al., 1997; Kurz et al., 2010).

Alignment between goals and instruction has been studied in several ways. Porter (2002) developed the Surveys of Enacted Curriculum to measure the alignment between the intended curriculum (the standards), the enacted curriculum, and the assessed curriculum and other studies have addressed the alignment between teachers’ goals and expectations for instruction and their classroom assessments. Such studies found that there is a strong relationship between the teachers’ expectations and the expectations of their unit assessments (Baumert et al., 2010; Martínez et al., 2012). In addition, there is evidence from other studies that while teachers may set higher-level instructional goals, their assessment tasks tend to be less cognitively demanding (Bol & Strage, 1996; Koh, 2011).

The Current Study

This study examines the breadth and depth of the learning targets teachers plan to address within a single unit of instruction, the assessment of those learning targets on their unit assessment, and the relationship of those learning targets to teachers’ CKT-E as
measured by the CKT-E assessment (Etkina et al., 2018). As CKT-E defines the knowledge teachers have for teaching a subject, teachers with more robust CKT-E should address a wider array of learning targets and set more in more depth learning targets than those with less robust CKT-E. We examine learning targets using two types of evidence. For both types of evidence, we examine whether particular learning targets are addressed and whether the learning targets are addressed in some depth.

To gather information about teachers’ goals for an instructional unit on energy in mechanics in high school physics we used pre-instructional interviews. These goals specified what participating teachers believed their students should know and be able to do by the end of the unit, and what teachers planned to teach. The second source of evidence came from the final unit test. We selected this particular artifact for its potential to provide insight into the breadth and depth of the learning targets that teachers addressed during their unit on energy, as unit assessments should contain items that are related to the targets within that unit.

From the analysis of both the teachers’ goals and unit assessments, we also examined the alignment between teachers’ goals and their final unit assessments.

Finally, the study examined the relationship of these measures with teachers’ CKT-E. Estimates of teachers’ CKT-E were calculated based on their responses on an assessment of CKT-E developed for the Content Knowledge for Teaching Energy Project (Etkina et al., 2018).
**Methods**

**Participants**

The participants in the study were 32 teachers from the Content Knowledge for Teaching Energy Project. The teachers were selected by responding to a call for participants that was sent to physics teaching mailing lists and groups near each research site. Eleven of the 32 teachers were women. Eighty-eight percent of the teachers had a master’s degree, and the same percentage of teachers took at least three physics courses as an undergraduate. Each teacher taught at least one physics class, ranging from conceptual physics to Advanced Placement® (AP®) physics and International Baccalaureate® (IB®) physics, and 94 percent of the teachers taught physics for three or more years.

**CKT-E Assessment**

As part of the CKT-E project, each teacher completed an assessment of CKT-E. This assessment was comprised of items that were based on the work of teaching energy in a high school physics class. For example, some items required the teacher to analyze and respond to student ideas while others asked the teacher to sequence instruction in a way that would be most effective in helping students develop certain concepts (Etkina, et al., 2018).

The assessment was administered to and field-tested with 329 physics teachers from across the United States. Following the scoring of the assessment, each teacher was assigned a standardized item response theory (IRT) estimate of their CKT-E. The estimates of CKT-E ranged from -1.95 to 2.16, with a median estimate of .01 (Phelps et al., 2020).
Interviews

Prior to teaching a unit on energy, each teacher in the study participated in a pre-instruction interview. The interview was meant to capture their goals for the energy unit, the structure and sequence of their unit, major instructional activities, and other information such as student difficulties and assessment plans. For the purpose of this study, we focused on four questions in order to gain insight into the goals the teachers set for the energy unit:

1a) What are your broad learning goals for your students this year in physics?

1b) What are your specific learning goals for your students in the energy unit?

2) What are the major activities that you will use to address each goal in the energy unit?

3) How do you chunk and sequence the energy unit?

The final two questions do not explicitly ask teachers to describe their goals for the unit, but when discussing major activities and how their unit was organized, teachers often mentioned additional goals for instruction.

Assessment Artifacts

Teachers were asked to submit their final assessment for the energy unit. Of the 32 teachers in the study, 28 teachers provided a final assessment. Of those 28 teachers, two teachers taught their energy unit at the end of the year and submitted their cumulative final exam. For those two teachers, only the items that were related to energy were
included in the study. Teachers who did not submit a unit assessment were excluded from this study.

**Coding of Interviews and Assessments**

**Instrument development**

Coding of interviews and artifacts was done with respect to the energy targets developed for the CKT-E project that define what a student should know and be able to do at the end of a unit on energy in high school physics (Etkina et al., 2018).

**Coding of interviews**

Responses to the interview questions were coded based on the teacher-stated goals and were referenced to the CKT-E student energy targets. The responses to each interview question were coded individually, and each goal that the teacher mentioned was first coded for the broad target category it addressed and then for the individual target(s) within the category. If a goal met the standard of an energy target, it was considered to address the learning target in-depth and was coded as *Target Category - In-depth* along with the corresponding individual target that the item addressed. If a goal was related to an energy target but did not meet the standard of the target, it was considered a superficial goal and coded as *Target Category - No Target*.

To illustrate the difference between a goal being coded as meeting the standard of an energy target and a similar goal being coded *No Target* but related to an energy target, consider the following statements from two different teachers. In one interview, the teacher states, “*It’s important for the students to know about work and how to calculate it,*” while a second teacher states, “*I think it is important for students to understand the idea of work and how work is one of the ways that energy is transferred into or out of a*
system.” In both statements, the teachers describe that they think work is an important topic. However, in the first statement, the teacher only describes that he or she believes students should know about work and the equation used to calculate it. The teacher does not make an explicit statement regarding what a student should learn about work beyond the equation. In contrast, the second teacher focuses on the students learning about work as a mechanism of energy transfer. This idea is much more in-depth than a student just being able to calculate work, and it directly addresses one of the CKT-E student energy targets: D4 - Recognizes that work is the way in which energy is transferred mechanically and may result in a change in temperature in some case. Using the coding scheme, the first statement would be coded as Transfer of Energy – No Target, and the second statement would be coded as meeting the target category Transfer of Energy – In-depth and the individual target D4.

The first author developed the coding scheme and coded all of the interviews. A second rater with more than 30 years of experience teaching physics to high school students and to physics educators was used to establish reliability. Out of 28 interviews, 112 total responses were coded (four per interview, to questions 1a, 1b, 2, and 3). Twenty percent of those questions, selected at random, were double coded, with 89 percent agreement at the target category level and 77 percent agreement at the individual target level. Disagreements were resolved through a conference between both coders.

**Coding of assessment artifacts**

To document the learning targets addressed by teachers’ unit assessments, each item on the assessment was coded individually based on the CKT-E targets addressed.
This coding was completed as part of an effort to measure the quality and demand of the final assessments teachers use during their energy unit (Zisk et al., 2017).

Coding was broken down into two levels. First, the item was coded based on the broad target category or categories it addressed (e.g., Target category A, Connections of Energy and Everyday Experiences). Then, the rater coded the item for the individual target(s) within the target category or categories in terms of depth. Items that met the CKT-E target completely were considered to have addressed the target in-depth and were coded as Target Category - In-depth along with the corresponding individual target that the item addressed. If the item was related to an energy target but did not meet the standard of the target, it was considered to have addressed the target category superficially and was coded as Target Category - No Target along with the target category it most closely matched.

To illustrate how items on teachers’ unit assessments were coded using the energy targets, consider the following example from a teacher’s assessment:

A waiter exerts a 10 Newton force lifting a tray 0.5 meters. How much work did the waiter do?

This item is rather simple with regard to the knowledge and skills required to solve it. The students need to simply plug the values for force and displacement into the equation for work simplified for the case of constant force and angle of zero degrees between the force and displacement vectors (Work = Force x Displacement; assuming the force that the waiter exerts is constant and points upward) and solve. While this item addresses the topic of work and the energy targets contain a target that addresses work (Target Category - Transfer of Energy; Individual Target - D4 - Recognizes that work is the way...
in which energy is transferred mechanically and may result in a change in temperature in some cases), just mentioning work in the problem and having students solve the equation only superficially addresses the target. This item would be coded as *Transfer of Energy - No Target*.

In contrast with the previous item, a teacher could ask the following:

*How much energy is transferred into a tray when a waiter exerts a 10 Newton force while lifting the tray 0.5 meters?*

While this question can also be answered using the same simplified equation for work, in order to answer it the students need to first recognize that work is a way energy is transferred into or out of a system mechanically (Target Category - *Transfer of Energy*; Individual Target - *D4 - Recognizes that work is the way in which energy is transferred mechanically and may result in a change in temperature in some cases*) and then plug the values for force and displacement into the equation. As such, this item addresses the same target as the previous item but does so at a deeper level than just asking the students to complete a “plug and chug” problem on the topic of work. Therefore, this item would be coded as *Transfer of Energy - In-depth, Target D4*.

Other items may address one or two targets within a domain. For example:

*Give an example of a system that loses energy because negative work is done on it.*

To answer this item, students need to have knowledge of two energy targets. These are the concept of a system and how the choice of a system can determine the types of energy present within the system and the concept of work and how work (in this case, negative work) can change the energy of a system. In analyzing this item using the CKT-E targets, we see that the item addresses the following targets:
3. **B1 - Understands that the energy accounting in a phenomenon depends on the choice of system**

4. **D4 - Recognizes that work is the way in which energy is transferred mechanically and may result in a change in temperature in some cases**

This item would be coded as *Choice of System - In-depth* and *Transfer of Energy - In-depth* as well as the individual targets shown above (B1 and D4, respectively).

Finally, other more complex items may require students to demonstrate their knowledge of more than two targets when responding.

Four raters coded the assessments. All of the raters had knowledge of physics and physics teaching. Three of the raters had degrees in physics, one had a degree in chemistry, and all four were enrolled in a graduate program focusing on physics education. Two of the raters were practicing teachers, one for five years and one in his first year. The other two raters had experience teaching physics at the college level. The first author was the master rater and had participated in the design of the energy targets.

Each assessment was coded by two different raters. At the target category level, agreement between the two raters was 88 percent, and agreement at the individual target level was 74 percent. A third rater resolved any discrepancies between codes.

**Determining the Breadth and Depth of the Learning Targets Within Teachers’ Goals and Unit Assessments**

We measured both the breadth (targets addressed *both* superficially and in-depth) and depth (targets addressed superficially *or* in-depth) of the energy targets that teachers addressed within their goals and on their unit assessments.

To measure the targets addressed at different depths, two analyses were conducted. To measure the targets addressed superficially, we counted targets categories that were addressed only at the superficial level (coded as *Target Category - No Target*).
To measure the targets addressed in-depth, we only counted the target categories that were addressed at the in-depth level (coded as *Target Category - In-depth*). At the in-depth level, we also counted the number of individual energy targets addressed in both teachers’ goals and on their unit assessments.

**Determining the Alignment Between Teachers’ Goals and Their Assessments**

To determine the alignment between teachers’ goals for instruction and their unit assessments, we counted the number of discrepancies between teachers’ goals and their assessments. In this study, we focused on two types of inconsistencies. The first one is coverage inconsistency. The coverage inconsistency was recorded when there was a mismatch between goals and assessments. For example, a learning target (at a superficial or in-depth level) was included in teachers’ goals but not included on the assessments and vice versa. The second inconsistency that was recorded was a depth inconsistency – a mismatch in the depth of the goals and assessments. For example, a teacher states a goal at an in-depth level but only addresses the target superficially on the assessment and vice versa. In this study we measured alignment in four ways as shown in Figure 3.

**Figure 3**

*Alignment Between Goals and Assessment Content*

| Coverage, Goals - Assessment: | The teacher mentioned a target (either superficially or in-depth) within their goals but did not include items on their assessment related to that target. |
| Coverage, Assessment - Goals: | The teacher addressed a target (either superficially or in-depth) on the assessment but did not state a goal related to that target. |
| Depth, Goals - Assessment: | The teacher mentioned an in-depth goal related to a target but only addressed that target superficially on the assessment. |
| Depth, Assessment - Goals: | The teacher addressed a target in-depth on the assessment but only mentioned a superficial goal related to that target. |
Findings

CKT-E Scores

Estimates of teachers’ CKT-E were determined using Item Response Theory (IRT). IRT models are a type of latent variable model where observed responses are considered manifestations of a construct that cannot be observed, in this case, CKT-E. The model assigns an estimate of CKT-E to each teacher, with average performance assigned an estimate of 0. CKT-E IRT estimates for the 28 teachers in this study ranged from -1.44 to 2.16. The median CKT-E estimate for these teachers was .67, indicating that the CKT-E scores for teachers in the study were above the average of the field test sample (estimates of CKT-E from the teachers in the national field test ranged (n = 362) from -1.95 to 2.16, with a median estimate of .01).

What is the Relationship Between Teachers’ Stated Goals and CKT-E?

In terms of the breadth of targets, most teachers stated goals that addressed 5–6 target categories. With regard to depth, for most teachers, no more than one target category was treated superficially. When restricting the analysis to only the targets addressed at an in-depth level, most teachers addressed 4–5 target categories and 7–8 individual targets (see Table 1).

To measure the relationship between the teachers’ goals and their CKT-E, we calculated the Pearson correlation between teachers’ CKT-E scores and the number of target categories addressed overall (breadth), the number of target categories addressed only at the superficial level, the number of target categories addressed at the in-depth level, and the number of individual targets addressed in their unit goals at the in-depth level (see Table 1).
Table 1

*Breadth and Depth of Teachers’ Unit Goals by Number of Target Categories and Individual Targets and Pearson Correlation Between Breadth and Depth of Content and CKT-E Score (Interview)*

<table>
<thead>
<tr>
<th>Range of Categories/Targets</th>
<th>Mean Number of Categories/Targets</th>
<th>SD of Categories/Targets</th>
<th>Pearson Correlation Between CKT-E and Breadth and Depth of Content (r_p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target Categories Covered (Breadth)</td>
<td>3–7</td>
<td>5.53</td>
<td>1.17</td>
</tr>
<tr>
<td>Target Categories Addressed Superficially</td>
<td>0–2</td>
<td>0.73</td>
<td>1.00</td>
</tr>
<tr>
<td>Target Categories Addressed In-depth</td>
<td>1–7</td>
<td>4.75</td>
<td>1.69</td>
</tr>
<tr>
<td>Individual Targets Taught In-depth</td>
<td>1–17</td>
<td>7.90</td>
<td>3.89</td>
</tr>
</tbody>
</table>

**p ≤ .01.

There was no relationship between CKT-E scores and the breadth teachers’ goals.

All teachers in the study tended to set goals related to the same breadth of targets.

However, with regard to depth, there was a significant negative correlation between CKT-E scores and targets addressed superficially and positive correlations between CKT-E scores and the targets teachers addressed at an in-depth level, both for categories and
individual targets. This indicates that, among teachers in the study, those with lower CKT-E scores were more likely to state superficial goals while teachers with higher CKT-E scores tended to set more in-depth goals and address more individual targets for their energy units.

We also examined whether the intended teaching of particular target areas was associated with CKT-E scores (see Table 2). There was a wide range in the number of teachers who addressed each of the target categories. All of the teachers in the study mentioned goals that addressed the target categories, Identification and Differentiation of Types of Energy and Transfer of Energy, while 16–23 teachers addressed each of the remaining target categories in their unit goals. The targets addressed least frequently across teachers were Systems and System Choice, Use of Representations, and Science Practices.
Table 2

Number of Teachers Who Addressed Each Target Category in Their Unit Goals (Interview)

<table>
<thead>
<tr>
<th>Target Category</th>
<th>Number of Teachers with Goal That Addressed Content Related to Target Category</th>
<th>Number of Teachers with Only Superficial Goal Related to Target Category</th>
<th>Number of Teachers with In-depth Goal Related to Target Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Connection Between Energy and Everyday Life</td>
<td>23</td>
<td>2</td>
<td>21</td>
</tr>
<tr>
<td>Systems and System Choice</td>
<td>16</td>
<td>4</td>
<td>12</td>
</tr>
<tr>
<td>Identification of Energy</td>
<td>28</td>
<td>0</td>
<td>28</td>
</tr>
<tr>
<td>Transfer of Energy</td>
<td>28</td>
<td>2</td>
<td>26</td>
</tr>
<tr>
<td>Use of Mathematics</td>
<td>23</td>
<td>10</td>
<td>13</td>
</tr>
<tr>
<td>Use of Representations</td>
<td>19</td>
<td>3</td>
<td>16</td>
</tr>
<tr>
<td>Science Practices</td>
<td>18</td>
<td>1</td>
<td>17</td>
</tr>
</tbody>
</table>

Goals associated with target category, Identification and Differentiation of Types of Energy, were almost always discussed in-depth. For almost all other categories, if the teacher addressed the category it was done at an in-depth level. The one exception was, Use of Mathematics, which was treated superficially by a large proportion (10 of 23) of the teachers who mentioned the category.

We then correlated the nature of target category coverage and CKT-E scores using point-biserial correlations (see Table 3). There was no relationship between
teachers’ CKT-E scores and whether or not they addressed any specific target category; all teachers were equally as likely to mention a goal related to any of the target categories. However, teachers with lower CKT-E scores were more likely to address categories superficially for the categories, *Connection of Energy to Everyday Life, Transfer of Energy, Use of Mathematics,* and *Use of Representations.* Teachers with higher CKT-E scores were more likely to address the following categories in-depth: *Systems and System Choice, Transfer of Energy,* and *Use of Mathematics.*

---

1To some extent the magnitude of these correlations was affected by the presence of several (2) teachers with CKT-E scores that were substantially lower than those for other teachers in the sample.
Table 3

*Point-biserial Correlation Between CKT-E Score and Teacher-stated Goals Related to Each Target Category (n=28)*

<table>
<thead>
<tr>
<th>Target Category</th>
<th>CKT-E and Content Addressed</th>
<th>CKT-E and Content Only Addressed Superficially</th>
<th>CKT-E and Content Addressed In-depth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Connection of Energy and Everyday Life</td>
<td>-.01</td>
<td>- .45*</td>
<td>.26</td>
</tr>
<tr>
<td>Systems and System Choice</td>
<td>.33</td>
<td>-.24</td>
<td>.50**</td>
</tr>
<tr>
<td>Identification of Energy</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Transfer of Energy</td>
<td>N/A</td>
<td>- .47*</td>
<td>.47*</td>
</tr>
<tr>
<td>Use of Mathematics</td>
<td>.17</td>
<td>-.49*</td>
<td>.60**</td>
</tr>
<tr>
<td>Use of Representations</td>
<td>- .11</td>
<td>-.50**</td>
<td>.21</td>
</tr>
<tr>
<td>Science Practices</td>
<td>.28</td>
<td>-.16</td>
<td>.34</td>
</tr>
</tbody>
</table>

*Note.* Correlations of N/A could not be calculated as every teacher addressed or addressed in-depth a goal related to that category.  
* *p ≤ .05. ** p ≤ .01.

Overall, regarding teachers’ goals, there was no relationship between teachers’ CKT-E and the breadth of their goals. However, when focusing on the depth, teachers with higher CKT-E scores were more likely to have in-depth goals, while teachers with lower CKT-E scores were more likely to set superficial goals.
What is the Relationship Between the Targets Addressed by Teachers’ Assessments and CKT-E?

The unit assessment was also analyzed in terms of the breadth and depth of the energy targets addressed (see Table 4). In terms of the breadth of the targets addressed, the assessments tended to contain items that addressed 5–6 target categories. When focusing on targets assessed at the in-depth level, we found that teachers’ assessments also tended to address 5–6 target categories and 8–9 individual targets. At the superficial level, there were only two instances among all of the teachers in the study in which a target category was addressed at only the superficial level. If teachers addressed a target category on their assessments, they tended to do so at some level of depth.
Table 4

*Number of Categories and Targets Addressed and Addressed In-depth and Pearson Correlation between CKT-E and Breadth and Depth of Content (Assessment)*

<table>
<thead>
<tr>
<th></th>
<th>Range of Categories/Targets</th>
<th>Mean Number of Categories/Targets</th>
<th>SD of Categories/Targets</th>
<th>Pearson Correlation between CKT-E and Breadth and Depth of Content (r_p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target Categories Addressed</td>
<td>3–7</td>
<td>5.32</td>
<td>.94</td>
<td>.39*</td>
</tr>
<tr>
<td>Target Categories Addressed In-depth</td>
<td>3–7</td>
<td>5.25</td>
<td>1.00</td>
<td>.41*</td>
</tr>
<tr>
<td>Individual Targets Addressed In-depth</td>
<td>4–17</td>
<td>8.89</td>
<td>3.13</td>
<td>.59**</td>
</tr>
</tbody>
</table>

*p ≤ .05. **p ≤ .01.

The depth and breadth of the targets addressed on the assessment were then correlated with CKT-E scores (see Table 4). There was a significant positive relationship between teachers’ CKT-E scores and the breadth of the targets addressed on their assessments as well as targets addressed in-depth at the category level and at the individual target level. Teachers with higher CKT-E scores addressed more targets and addressed more targets in-depth than teachers with lower CKT-E scores.
What is the Alignment Between Teachers’ Instructional Goals and Their Unit Assessments?

We evaluated the alignment between teachers’ goals and their assessments in two ways. First, we measured the alignment of the targets addressed within teachers’ unit goals and the targets addressed on the assessments by counting the discrepancies between teachers’ goals and their assessments. To measure the number of discrepancies in the addressed targets, we counted all of the instances in which teachers stated a goal related to a target category (either at the superficial or in-depth level) and did not address the target category on their unit assessments, as well as the number of instances in which teachers included targets on their assessments that were not included in their unit goals.

In counting the number of instances in which teachers included targets in their goals but did not address the target on their assessments, teachers in the study tended to mention, on average, one target category that they did not include on their assessments (mean number of discrepancies= .95, SD=.74), and the largest discrepancy was two target categories. A similar magnitude of discrepancies was noted for situations in which a target category was addressed in the assessment but had not been stated as a goal (mean number of discrepancies= 86, SD=.72), and the largest discrepancy was two target categories (see Table 5).
Table 5

Number of Teachers Per Number of Discrepancies: Targets Addressed In-depth

<table>
<thead>
<tr>
<th></th>
<th>Three Discrepancies</th>
<th>Two Discrepancies</th>
<th>One Discrepancy</th>
<th>No Discrepancies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Targets Included in Goals/Not Included on Assessment</td>
<td>0</td>
<td>7</td>
<td>13</td>
<td>8</td>
</tr>
<tr>
<td>Targets Included on Assessment/Not Included in Goals</td>
<td>0</td>
<td>4</td>
<td>13</td>
<td>11</td>
</tr>
<tr>
<td>Targets Addressed In-depth in Goals/ Superficially on Assessment</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>26</td>
</tr>
<tr>
<td>Targets Addressed In-depth on Assessment/ Superficially in Goals</td>
<td>1</td>
<td>4</td>
<td>7</td>
<td>17</td>
</tr>
</tbody>
</table>

We then correlated the number of discrepancies in target coverage with teachers’ CKT-E scores and found no significant correlations between teachers’ discrepancies in target coverage and their CKT-E scores. However, teachers with lower CKT-E scores were more likely to assess targets in-depth at the same time that they described the goals at a superficial level \(r_{pb}(28)=-.62, p<.01\).

Overall, there were few discrepancies when measuring the alignment between of teachers’ goals and their unit assessments. Similarly, there were only two instances in which a teacher stated an in-depth goal but assessed the target at a superficial level.
However, teachers with lower CKT-E scores tended to have more discrepancies in which they stated a superficial goal but then assessed the target at a more in-depth level.

**Discussion**

This study explored the breadth and depth of the learning targets addressed within high school physics units on energy in mechanics. Specifically, the study examined teachers’ goals for instruction, the learning targets addressed on their unit assessments, and the alignment between teachers’ goals and unit assessments. The results of the study expand our understanding of how a teacher’s CKT is related to the learning targets teachers address within the classroom.

**The Targets addressed in Teachers’ Goals for Instruction**

In examining the goals that teachers stated they had set for instruction, all teachers, regardless of their CKT-E score, tended to address the same breadth of energy targets at some level throughout their energy unit. All of the teachers had a common understanding of the topics that should be addressed within the energy unit. This common understanding may have been driven by curricular materials or standards documents that indicate the core topics within energy that should be taught. For example, in the *Next Generation Science Standards* (NRC, 2013), the disciplinary core ideas within energy are focused around definitions of energy, energy transfer and conservation, and mathematical models of energy and energy transfer. The majority of teachers in our study addressed each of these topics in their unit goals. However, it is important to note that, in this study if a teacher stated that, for example, conservation of energy was an important goal for instruction and then did not expand on that statement, we have no way of knowing if his
or her understanding of energy conservation is in line with the accepted understanding of energy conservation.

While the teachers in the study all discussed goals related to the same content, the teachers did not address the content at the same depth. Teachers with higher CKT-E scores stated more in-depth goals compared to teachers with lower CKT-E scores, and specifically, teachers with higher CKT-E tended to mention more in-depth goals related to target categories, *Systems and System Choice*, *Transfer of Energy*, and *Use of Mathematics*.

The discrepancy between the breadth of teachers’ goals and the depth of their goals and its relation to CKT-E could be rooted in differences in teachers’ knowledge of the student energy targets. Numerous researchers have described that a deep knowledge of the domain being taught is essential to carrying out the work of teaching (Ball et al., 2008; Elbaz, 1983; Gitomer & Zisk, 2015; Leinhardt & Greeno, 1986; Magnusson et al., 1999; Shulman, 1986, 1987). These researchers have highlighted the difference between general knowledge of the domain, which has been described as common content knowledge (Ball et al., 2008) or foundational knowledge (Rowland et al., 2005) and deeper knowledge of the content that is used to help teach the subject matter and to help students learn. Both types of knowledge are essential to a teacher’s content knowledge for teaching, but a general knowledge of the domain is not sufficient to carry out the work of teaching. As all of the teachers in the study addressed the same breadth of targets in their goals, it is possible that all of the teachers in the study have a similar knowledge of the general learning targets within energy. The energy target categories address all of the major topics and ideas within energy, and the teachers in the study may have gained
some knowledge of the targets through their educational preparation and while being on
the job. However, the difference in the depth of teachers’ goals points to a difference in
the depth of the knowledge of the student energy targets between teachers with lower
CKT-E and teachers with more developed CKT-E. A lack of a robust knowledge of the
energy targets among teachers with lower CKT-E could lead to those teachers not
understanding the topics fully in order to craft in-depth goals for their students.

The Targets Addressed on Teachers’ Assessments

We also examined the learning targets of teachers’ energy units through the
targets addressed on their unit assessments. Teachers with higher CKT-E addressed more
targets on their assessments and addressed more in-depth targets. As with teachers’ goals,
this difference in the targets addressed on the assessments could be attributed to teachers
with higher CKT-E having a more robust understanding of the student energy targets,
which would lead them to create more in-depth assessments. Additionally, since teachers
with lower CKT-E tend to set fewer in-depth goals, it is less likely that they would create
an assessment to address more in-depth targets.

Another possible explanation for the relationship between in-depth assessments
and CKT-E is that teachers with lower CKT-E may not have the knowledge to develop
items that address the energy targets at an in-depth level. Items that address more in-
depth targets are more difficult to create than items that address targets at a less robust
level. It is rather simple to create a fill-in definition question about a type of energy, for
example, or to create a “plug and chug” math problem, as closed questions that require
simple recall are easy to generate (Chin & Brown, 2000). In contrast, it is much more
difficult to create an item in which students have to develop a mathematical relationship
between different types of energy based on how they define the system of analysis. This difficulty in creating in-depth assessments was addressed by Pellegrino (2013), who highlighted the challenges in creating assessments that are aligned with the more in-depth NGSS. Guidance documents, such as Developing Assessments for the Next Generation Science Standards (National Research Council, 2014), have attempted to address the challenges in creating such assessments, but at the time of this study, these resources were still fairly new. It is possible that teachers with more developed CKT-E have a stronger knowledge of the subject matter and also how to carry out the task of teaching of designing assessments, which would allow them to create more in-depth items.

**Alignment**

The final research question of this study concerns the relationship between the goals that teachers set for instruction and the content of their assessments. If teachers’ goals describe what they expect students to know and be able to do by the end of their unit and the unit assessment is a measure of what the students learned, the content of both should be related.

Generally, all of the teachers in the study assessed the targets that they stated as goals for their unit. This falls in line with the findings of Baumert et al. (2010) and Martínez and colleagues (2012)—that assessments are reflective of the expectations of the classroom.

However, when we measured the number of instances in which teachers included items on their assessments that addressed targets that were not stated in their unit goals, we found that teachers with lower CKT-E scores were more likely to include items that addressed targets at an in-depth level despite stating a superficial goal related to that
target. The inclusion of in-depth targets on teachers’ assessments that was not included within their unit goals could be due to teachers responding to the standards, textbooks, or guidance from their school or district to assess certain content. This could result in items being included on an assessment even though a teacher does not state them as an important instructional goal. Similarly, if a teacher does not have a robust understanding of the energy targets, he or she may include items on the assessment from the standards (or some other resource) without understanding the purpose of the items. It is possible that teachers include items that seem to address targets at an in-depth level, but only expect superficial responses. Additionally, teachers may use assessment items or whole assessments that they did not create (e.g., a department-designed assessment), which could lead to this discrepancy.

**What Do These Findings Tell Us About the Construct of CKT?**

This work expands on the research into the nature of teachers’ CKT-E by providing evidence of how practice differs among teachers with different CKT-E as measured by the test that we developed. Previous work has found correlations between teachers’ CKT and measures of practice such as observation measures and measures of student learning such as standardized assessments (Hill & Charalambous, 2012; Hill et al., 2007; Hill et al., 2005; Hill et al., 2012). In this study, we focused specifically on the teaching of the important learning targets within the domain of energy—whether teachers with different CKT-E as measured by a CKT assessment differ in their goals and their treatment of learning targets in their unit assessments. This study provides evidence that teachers with different CKT-E carry out the work of teaching energy in different ways. The teachers in this study with higher CKT-E set more in-depth goals for their students
and addressed more in-depth learning targets within their energy unit as evidenced by their final assessments. This difference in their goals and their unit assessments supports the idea that the CKT is strongly based on the knowledge of the important learning targets within the domain—not just a general knowledge of the subject matter but a more in-depth knowledge of the learning targets within the domain and how to teach to those learning targets.

While all of the teachers in the study had a general knowledge of the energy learning targets as evidenced by the breadth of their goals, teachers with higher CKT-E were able to set more in-depth goals for their students and then create more in-depth assessment items around those goals, indicating a deeper knowledge of the learning targets within the domain of energy. Similarly, teachers with higher CKT-E were also able to develop assessments that addressed more targets and more in-depth targets, which points to a more developed knowledge of how to create assessment items around the energy targets. This difference in the depth of knowledge that teachers have for teaching energy and the relationship between CKT-E and teaching practice also provides evidence that assessments of CKT can be sensitive to differences in the knowledge for teaching a subject that influences how teachers carry out the work of teaching around that subject.

**Implications for Practice**

The findings for this study indicate that only teachers with more robust knowledge for teaching the content are able to both set deep goals for instruction and develop assessments that address learning targets that are in line with the new science standards. This need for a robust knowledge of the learning targets within a specific topic that we wish students to learn is indicative of a need for more education aimed at
developing both pre-service and in-service teachers’ knowledge of important learning targets for students and their knowledge of how to help students meet those targets. This is especially important in US physics classrooms where only 35 percent of physics teachers possess a degree in physics (Meltzer et al., 2012). Without a background in physics, these teachers may be already behind, particularly when it comes to the knowledge related to more in-depth learning targets in physics that is not needed for other majors. In fact, there have been several calls for improved training to develop teacher knowledge, such as Abell and colleagues (2009) who argued for increased training for doctoral students who plan to be teacher educators in order to better educate future teachers, and the work by Schneider (2015) and Friedrichsen and Berry (2015) who both set out to describe learning progressions for science educators to help develop knowledge for teaching. In physics specifically, a recent effort to highlight effective teacher education programs and improve pre-service physics teacher education (Sanifer & Brewe, 2015) stressed the importance of developing the specific knowledge and practices needed to be able to effectively teach physics (Etkina, 2010; Atkins & Salter 2015; McDermott et al. 2015; Palmquist & Jackson, 2015). Without this training, the goal of providing students with opportunities to more in-depth learning targets may not be fully realized.

**Future Work and Limitations**

Though in this study we found a relationship between teachers’ CKT-E and their energy goals and unit assessments, there are several limitations to the study. First, with regard to teachers’ unit goals, when a teacher stated a superficial goal in their interview, we had no way of knowing their understanding around that goal. A teacher could have
said that he or she thought conservation of energy was an important goal for their students, but if the teacher did not go into further detail, we had no way of knowing whether their understanding of the goal was more robust or fit with the accepted understanding of the concept.

With regard to the assessments that teachers used during their energy unit and submitted to the study, we made the assumption that the teachers either completely designed the assessments, selected the assessment items, or had a large part in designing the assessments. In the study, we controlled for this by asking the teachers if they created the assessments on their own. However, in generalizing the results, it is possible that teachers are required to give assessments that they did not create. Similarly, the teachers might be operating under certain curricular restraints such as district requirements. These constraints could certainly influence the learning targets addressed by both teachers’ goals and their assessments.

Finally, with regard to the sample of teachers in this study, while 28 teachers is not a particularly small sample, the sample of teachers was developed through recruitment and connections of project investigators to physics teachers in schools within the researchers’ geographic areas. This sample was not designed to be representative of the universe of physics teachers in the United States. However, there was significant variability in CKT-E scores and measures of practice across this group of teachers.

Future work could address some of these limitations and also provide a more complete understanding of the relationship between CKT-E and the targets of instruction. First, work needs to be done to understand how teachers with different CKT create their goals for instruction and their unit assessments. Do they refer to curriculum documents,
their knowledge of the subject, or perhaps district requirements? Additionally, in this study, we assessed teachers’ unit goals (i.e., the content they intend to teach) and the targets addressed by their unit assessments. Absent from this study is the day-to-day instruction within the energy unit. In future work, we will examine videos of classroom teaching in order to see if the same patterns we found in this study hold true while teachers are teaching the content and if there is alignment between teachers’ goals, their practice, and their assessments.
References


Chapter 4: Investigating patterns in a relationship among Content Knowledge for Teaching Energy multiple measures of teaching quality and student learning

Abstract: Content Knowledge for Teaching (CKT) describes the knowledge that teachers have and use to carry out tasks of teaching with the goal of helping their students learn. As such, the underlying premise is that CKT should influence all aspects of teaching practice, and in turn, student learning. In previous studies, we found that high school physics teachers’ Content Knowledge for Teaching energy (CKT-E) was positively related to the quality and demand of assignments and assessments that they use during their unit on energy in mechanics. In this study, I further investigate the relationship between CKT-E and teaching practice by conducting case studies of four teachers to examine the consistency between their CKT-E, all aspects of teaching practice, and student learning. I then use the framework of CKT to explain the consistencies and discrepancies among scores on measures of CKT-E and classroom practice. I find that there is generally consistency across CKT-E and all measures of practice, and discrepancies can be explained by teacher’s strong CKT which allow them to teach effectively in ways that are not typically described as effective teaching, or through a teacher’s strong knowledge of a specific curriculum and curricular approach.

Introduction

For the past several decades, researchers have been investigating the knowledge teachers need for high-quality teaching and how they enact that knowledge when planning for and carrying out instruction. In attempting to characterize the knowledge that allows teachers to prepare for and carry out instruction, Ball, Thames and Phelps (2008) developed the construct of content knowledge for teaching, or CKT. The assertion
of Content Knowledge for Teaching is that there is a set of knowledge that teachers need in order to carry out critical tasks of teaching. CKT model builds on Shulman's (1986) concept of Pedagogical Content Knowledge (PCK), which describes the knowledge necessary for teaching as a combination of subject-matter knowledge and the knowledge of how to make a subject comprehensible to others.

The work to define the construct of Content Knowledge for Teaching has led to several efforts to assess this knowledge using assessments that not only evaluate the knowledge of the content, but also the ability of the teacher to carry out everyday tasks of teaching (Etkina et al., 2018; Gitomer et al., 2014; Herbst & Kosko, 2014; Hill et al., 2004; Kersting et al., 2012; Krauss et al., 2008; Phelps & Schilling, 2004; Phelps et al., 2020; Sadler et al., 2013).

Several efforts have investigated the relationship between teachers’ CKT for teaching particular subjects as measured by written assessments, their classroom practice as measured through classroom observations, and student learning outcomes (Baumert et al., 2010; Carlisle et al., 2009; Gitomer et al., 2014; H. C. Hill et al., 2005; Heather C. Hill et al., 2008; Keller et al., 2017; Kersting et al., 2012; Park et al., 2011; Geoffrey Phelps et al., 2012). Some such studies looked across a whole year of instruction by sampling teachers’ content knowledge of a broad range of topics and instruction related to those topics (such as, Baumert et al., 2010; Gitomer et al., 2014; Hill et al, 2005), while others (for example, Kersting et al, 2012, Park et al. 2011) focused on a narrow range of content and student outcomes specifically related to that content in order to find specific relationships between knowledge for teaching, classroom practice, and student
outcomes. Taken as a whole, these studies have generally found positive relationships between CKT and measures of instructional quality and student outcomes.

Conceptualizations of Content Knowledge for Teaching specify that teachers’ knowledge for teaching should be reflected in all aspects of instruction. While traditionally, studies relating CKT to instruction have focused on classroom observations to assess practice, it is logical to think that teachers with different CKT should differ in how they carry out all aspects of teaching practice. Other windows into teaching practice can be found by assessing teachers’ goals and how they plan for instruction, and how they design and select artifacts (assignments and assessments) for their instructional units. Classroom artifacts in particular offer insight into the expectations that the teacher sets for their students, and how the teacher assesses student knowledge and abilities. Two studies have evaluated the relationship between high school physics teachers’ CKT for teaching energy (CKT-E) and the assignments and assessments they developed for their energy unit and found positive relationships between teachers CKT-E and the quality and intellectual demand of the assignments and assessments (Gitomer et al., 2020; Zisk, et al., 2019a).

Despite efforts to find relationships between Content Knowledge for Teaching and measures of classroom practice, there is a lack of research to evaluate coherence among multiple measures of CKT and classroom practice. For example, do teachers with high CKT perform better in the classroom, create higher quality artifacts, and have students who perform better than those with lower CKT? If there are cases in which all of the measures do not line up, what accounts for the discrepancies?
The work contained in this study is based on the Content Knowledge for Teaching Energy project (CKT-E project (Etkina et al., 2018; Phelps et al., 2020)) to define and validate the construct of Content Knowledge for teaching energy. The project resulted in the creation and validation of the test assessing physics teachers content knowledge for teaching energy in the context of mechanics. Over 500 teachers took the test. As a part of the project, 32 teachers out of those 500 participated in the intensive study. In that part of the study they video recorded all of their lessons on energy, collected samples of their assignments and assessments, responded to the interview questions prior to teaching the unit, and took the CKT-E test. Additionally, their students completed pre-unit and post-unit assessments designed by the project team. Previous studies based on this project have found a relationship between teachers’ CKT-E and classroom practice (Bell et al., 2017), the quality and demand of instructional artifacts (Gitomer et al., 2020; Zisk, et al., 2019a). In this study, we investigate closely four teachers from the group of deep study teachers. Using the case study approach, we examine their goals for instruction, their classroom practice, their instructional artifacts and their student outcomes in an attempt to explain coherence and discrepancies between different aspects of their instruction and their knowledge for teaching energy as measured by the CKT-E assessment.

Our goal is to answer the following research questions:

1. To what extent do teachers demonstrate consistency among their content knowledge for teaching as measured by an assessment and measures of teaching quality?
2. How does the framework of Content Knowledge forTeaching explain consistencies and discrepancies among different measures and student achievement?
Literature Review

**CKT and Its Relationship to Teaching Quality**

Theories of CKT specify that a teacher’s CKT should be related to all aspects of their teaching practice. Most studies investigating this relationship have focused on a broad range of topics within a subject area (such as teacher’s CKT for teaching first year algebra), and have examined the relationship between teachers’ CKT and other measures of teaching quality such as classroom observation scores and student achievement, as measured by assessments, such as scores on standardized tests. Such studies found significant positive relationships between classroom observation scores and teachers’ CKT (Hill et al., 2005; Hill et al., 2007; Hill et al., 2012; Hill & Charalambous, 2012). Similar relationships have been found between CKT and student outcomes as measured by standardized assessments (Hill et al., 2005; Hill et al., 2012; Hill and Charalambous, 2012).

While studies of CKT that focus on a broad range of topics in a subject area allow for generalizations to be made regarding the relationship of CKT and teaching quality, other studies have closely examined teachers’ knowledge for teaching specific content within a subject and their teaching and student outcomes related to that content, allowing for inferences to be made regarding how teacher knowledge mediates practice and student outcomes. In one such study, Park, Jang, Chen, and Jung (2011) measured biology teachers’ knowledge for teaching photosynthesis and heredity and their ability to carry out reformed teaching practices, and found a significant, positive relationship between teacher knowledge as measured by a rubric meant to assess teachers’ pedagogical content knowledge while teaching (Park & Oliver, 2008) and reformed
teaching, as measured by the Reformed Teaching Observation Protocol (RTOP) (Sawada et al., 2002). Keller, Neumann, and Fischer (2017) also examined the influence of teachers’ knowledge for teaching electricity within a high school physics course on student achievement and found a positive correlation between a teacher’s knowledge and the achievement of their students, based on an assessment of electricity and electrical energy.

Though there are studies that have examined the relationship between CKT and measures of practice, there have been no studies that have investigated the relationship between a teacher’s CKT and multiple measures that address all areas of practice, such as planning, the creation and selection of the assignments and assessments teachers use during instruction, classroom instruction, and student outcomes, and the coherence between all those measures.

**Content Knowledge for Teaching Energy Framework (CKT-E)**

In this paper, we describe a study that is part of a larger effort to define the construct of content knowledge for teaching energy (CKT-E), and assess the relationship between a teachers knowledge for teaching energy and their classroom practice as measured by multiple measures focused on practice related to only the topic of energy within a high school physics course. The measures develop for the project included an assessment of CKT-E (See Etkina et al. (2018) and Phelps et al. (2020) for a complete description of the assessment), a measure of classroom instruction and practice (See CITE), a measure develop to assess the quality and demand of classroom assignments and assessment (See Gitomer et al. (2020) and Zisk et al. (2019a) for a complete
description of the assignment and assessment protocol) and measures meant to assess student understanding of energy topics.

Central to development of these measures was the development of the framework of CKT-E. The framework was based on the assertion that CKT is knowledge of the important content and skills (student learning targets) within the domain of learning and the knowledge to carry out important practices (tasks of teaching) to support students in meeting those targets. Within the framework of CKT-E, the Student Energy Targets define the important concepts and skills in the domain of energy and the Tasks of Teaching define the key teaching practices that support students in meeting those targets.

Etkina and colleagues (2018) further refined the framework by disaggregating CKT-E into two components, CKT-D (disciplinary) and CKT-P (pedagogical). Etkina and colleagues defined CKT-D as disciplinary knowledge of physics that does not require detailed knowledge of student learning or the school context. For example, CKT-D would allow the teacher to analyze a phenomenon and determine whether the energy of a given system is conserved or constant. CKT-P calls upon disciplinary knowledge but also contains knowledge of student ideas and how to help students learn. For example, CKT-P would allow a teacher to choose the correct instructional intervention to assist students in developing the knowledge necessary to distinguish between energy constancy and conservation and how system choice determines if energy is constant. Etkina et al. (2018) analyzed the items on the CKT-E assessment and teachers’ performance on those items and found that while conceptions of CKT specify that teachers must have knowledge of the content to effectively guide students to learn, there were cases in which teachers were
able to build on student reasoning and select the appropriate pedagogical strategy (CKT-P) even when they themselves lacked the appropriate disciplinary knowledge (CKT-D)

Within our framework, The Student Energy Targets articulate the important concepts, practices and knowledge representations that are critical to the domain of energy in mechanics and are in line with existing research on the learning and teaching of energy (Etkina et al., 2018). Teachers with high CKT-D should have a knowledge of these targets. The targets are organized into seven target categories with sub-targets within each. The target categories consist of:

· connections of energy and everyday experiences,
· choice of system,
· identification and differentiation of different forms of energy and other physics concepts
· transfer of energy,
· use of mathematics,
· use of representations, and
· science practices.

These targets focus on the disciplinary core ideas related to energy, the crosscutting concepts of systems and energy and matter, and the science practices. For example, the targets specify that students should develop an understanding that the energy of a system is always conserved, and in some cases, the energy of the system could be constant (Energy Target D1), as opposed to just knowing that the definition of the conservation of energy means that any change in the total energy of a system can be accounted for by the work done on the system. In energy, conservation means that any
change in the total energy of the system can be accounted for – the total change in energy of the system will equal the energy transferred into or out of the system, while constancy means that the total energy of a system does not change (remains constant in time) (Seeley et al., 2019).

The Tasks of Teaching describe the activities that teachers engage in that help to support and promote student learning (Ball, 2000) and include: (I) anticipating student thinking around science ideas; (II) designing, selecting, and sequencing learning experiences and activities; (III) monitoring, interpreting, and acting on student thinking; (IV) scaffolding meaningful engagement in a science learning community; (V) explaining and using examples, models, representations, and arguments to support students’ scientific understanding; and (VI) using experiments to construct, test, and apply concepts (Etkina et al., 2018). Carrying out the instructional tasks with the goal of helping students learn the important energy targets calls upon a teacher’s CKT-P.

In carrying out any instructional task, teachers must call on their knowledge of the student energy targets and the tasks of teaching. For example, when designing an assessment item, teachers must have knowledge of the energy target they plan to assess, and then the knowledge that allows them to create an assessment item or task that will require students to show their knowledge related to that target.

**The Current Study**

The current study is based on the premise that a teacher’s CKT-E influences all aspects of their practice. As such, we build off of the results of previous studies of CKT-E that have found that CKT-E is a measurable construct (Etkina et al., 2018; Phelps et al., 2020), and is positively correlated to classroom instruction as measured by an
observation protocol, the quality and demand of classroom assignments and assessments, and student learning. In this study, we examine four teachers, each with varying CKT-E to illustrate the consistency between their performance on each of the measures and explain any discrepancies between among their measure scores.

Methods

Study Participants
The participants in the study were four teachers selected from a larger group of 32 teachers from the Content Knowledge for Teaching Energy Project (Etkina et al., 2018; Phelps et al., 2020). As participants in the study, each teacher took an assessment of their CKT-E, participated in a pre- and post-unit interview, recorded each lesson during their unit on energy, and submitted three assignments used during instruction and their unit assessment. Their students also took a general science assessment prior to starting the unit and an energy focused post-test.

Data Sources

Content Knowledge for Teaching Energy Assessment
Each teacher in the study completed an assessment of their Content Knowledge for Teaching Energy. This assessment consisted of items that assessed teachers’ knowledge of energy content and items that were based on the work of teaching energy in a high school physics class. For example, some items required the teacher to analyze and respond to student ideas while others asked the teacher to sequence instruction in a way that would be most effective in helping students develop certain concepts (Etkina, et al., 2018). The item shown in figure 1 is an example of an item that was designed to assess knowledge of systems, work as the change in energy of a system, and energy constancy
and conservation. In contrast, the item shown in figure 2, not only requires energy content knowledge, but also requires the teacher to evaluate student responses and then choose the most productive activity to help students further develop their knowledge of elastic potential energy. The assessment was administered to and field-tested with 329 physics teachers from across the United States. Following the scoring of the assessment, each teacher was assigned a standardized item response theory (IRT) estimate of their CKT-E. The estimates of CKT-E ranged from -1.95 to 2.16, with a median estimate of .01 (Phelps et al., 2020). Additionally, for the purpose of the case studies, we analyzed a subset of 8 items that focused on energy content knowledge to provide insight into the subjects’ knowledge of energy.

Figure 1
Sample CKT-E Item Concerning Systems

**Item 007a**

In a situation with a number of interacting objects, one may select any subset of them as the system of interest. The objects that have not been selected as belonging to the chosen system are therefore external to the system.

Ms. Inez wants to help her students realize that energy is a conserved quantity but the energy of a particular system may not be constant, depending on the specific scenario and the choice of system for analysis. She decides to have them focus on a scenario of a cyclist riding up a hill at a constant speed.

For each of the following systems, indicate whether the energy associated with the system increases, decreases, or remains approximately constant.

<table>
<thead>
<tr>
<th></th>
<th>Increases</th>
<th>Decreases</th>
<th>Approximately Constant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bicycle, rider, air, pavement, and Earth</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bicycle, rider, and Earth</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bicycle, air, and pavement</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bicycle and Earth</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Explain your reasoning for each of the systems.

This item assesses teachers’ energy content knowledge by providing them with different systems and asking them to determine if the energy of the system increases, decreases, or stays the same throughout the process of a bike being ridden up a hill at a constant speed.

---

Figure 2
Sample CKT-E Item Concerning Elastic Energy

**Item 016B**

Two students in Ms. Engel’s Physics Class are discussing the energetics of dribbling a basketball on a wooden floor. They agree that all of the kinetic energy gets converted into elastic energy for an instant when the basketball is compressed the most. They also recognize that many objects can be modeled as springs, even basketballs and wooden floors. They are uncertain about whether there would be equal amounts of elastic energy in the ball and the floor. They call Ms. Engel over to share their ideas with her and get some help.

Marcos says, “We are thinking that when the ball compresses against the floor, the forces that the ball and the floor exert on each other would be equal and opposite, so maybe the amount of elastic energy in the floor is the same as the elastic energy in the ball.”

Louisa responds, “I get that the forces are the same, but I am thinking that the ball compresses more than the floor, so shouldn’t there be more energy stored in the ball?”

Marcos replies, “But the floor is more rigid and would have a higher spring constant. I think the larger k of the floor compensates for the smaller Δx in the \( \frac{1}{2}k(\Delta x)^2 \) equation, and the elastic energies are equal.

**Item 016B**

Which of the following activities would be most likely to provide Ms. Engel’s students with additional insights about the relative amounts of elastic energy during the bounce of the basketball?

A. They could measure the spring constants and the displacements of both the floor and the ball and use those to compare elastic energies.

B. They could compare the elastic energies of two non-identical springs when they are compressed with the same force.

C. They could do an experiment to see if a basketball bounces higher on a soft carpet or a hard concrete floor.

D. They could do an experiment to show that the same basketball will not bounce as high off the gym floor if it has first been put in the freezer.

Explain your selection and how the activity you selected might provide students with additional insights about the relative amounts of elastic energy during the bounce of the basketball.

This item assesses teachers’ content knowledge for teaching energy by asking them to determine the most effective activity for helping students to understanding the role of elastic potential energy when a basketball is bounced up and down on the floor.

**Student Science Pre-Assessment and Energy Post-Assessment**
Prior to teaching their unit on energy, the teachers were asked to administer a
general science pre-assessment to their students. This assessment served as a measure to
evaluate the general level of knowledge of science reasoning and the practices of science
of the students in their classroom. There were also several items from the Force Concept
Inventory embedded in the assessment (Hestenes et al., 1992). An example from the
assessment can be found in figure 3.
Sample Pre-Assessment Items

**Question 13**

Can a model of an object (such as a car) be used to predict how an object will behave in certain situations (such as crashing into a wall)?

(A) No, a model is only useful for communicating to others what an object is like, not for making predictions about how an object will behave.

(B) No, predictions made with a model are never useful because a model is never exactly the same as the object it represents.

(C) Yes, a model will behave exactly as the object it is representing behaves because a model is exactly the same as the object it represents.

(D) Yes, a model can predict the behavior of the object it represents but the predicted behavior of the object may not be exactly the same as the actual behavior of the object because a model is never exactly the same as the object it represents.

**Question 14**

A student thinks that there are three variables (X, Y, and Z) that may affect the result of her experiment. What should the student do to find the effect of variable X on the result of the experiment?

(A) Change variable X and keep variables Y and Z the same.

(B) Change variables Y and Z at the same time and keep variable X the same.

(C) Change variables X and Y at the same time and keep variable X the same.

(D) Change variables X, Y, and Z at the same time.

After teaching the unit on energy, the teachers were asked to administer an assessment of energy knowledge to the students in their class. The questions on the assessment were drawn primarily from Diagnoser (FACET Innovations, 2003) and assessed students’ knowledge of the Energy Targets. Example items can be found in figure 4.

**Figure 4**
Sample Energy Post-Assessment Items

Javier pushes a rolling cart on a slippery table. In Trial 1, the kinetic energy of the cart at the end of the table is 5 Joules. In Trial 2, the kinetic energy of the cart at the end of the table is 20 Joules, i.e., four times as large.

Relative to the speed of the cart at the end of the table in Trial 1, the final speed of the cart in Trial 2 is

(A) Larger but less than double
(B) Double
(C) More than double but less than four times as large
(D) Four times as large

Question 8

Object 1 and Object 2 slide down a very slippery slide (effects of friction between objects and slide are very small). Both objects start at rest. Object 2 is twice the mass of Object 1.

Compared to the speed of Object 1 at the bottom of the slide, the speed of Object 2 is

(A) Equal
(B) Larger but less than double
(C) Double
(D) Four times as large

Classroom Assignments and Final Unit Assessment

Teachers in the study were asked to submit four classroom artifacts, three assignments used during their unit on energy, and the final unit assessment. In this study, assignments and assessments were used to gain insight into the types of assignments and assessments that teachers design and select for their energy unit with respect to the learning targets that they addressed (Zisk et al., 2019b) and their quality and demand (Zisk et al., 2019a). These artifacts were coded with respect to the set of student learning targets developed for the CKT-E project described previously (Zisk et al., 2019b). The artifacts were also coded for their quality and demand using a protocol designed to assess
the demand of the tasks within the artifact with respect to 3 dimensions: analysis, mathematics and representations. For each dimension, we developed four ordered scale points to represent the level of cognitive demand of each task within the assignment or assessment. Cognitive demand is defined as the level of reasoning required to complete the task (Stein et al., 2001; Tekkumru-Kisa et al., 2015). A complete description of the quality and demand study can be found in Zisk et al. (2019a), and a description of the three domains can be found in figure 5.

**Figure 5**

*Dimensions of Artifact Protocol*

- **Analysis**: This dimension considers the complexity of the phenomena that the students are asked to work through on the assignments or assessments.
- **Mathematics**: This dimension considers the depth and complexity of the mathematics that students need to engage in when solving the problems on the assignments or assessments.
- **Representations**: This dimension considers the extent to which students need to interpret or use representations to solve problems on the assignments or assessments.

**Pre- and Post-Unit Interview**

Prior to teaching their unit on energy and following the conclusion of the unit, each teacher in the study participated in an interview. The pre-unit interview asked the teacher to describe their learning goals for the unit, as well as the instructional activities in the unit designed to meet the goals and how students would be assessed during the unit. The post-unit interview asked the teachers to describe students’ success in meeting
the goals of the unit and possible sources of difficulty. A description of how the interviews were coded for learning goals can be found in Zisk and colleagues (2019b). In this study, we used the interviews to gain insight into the goals teachers set for instruction, and how they designed, selected and used classroom assignments and assessments to address those goals. We also used the interviews as a source of information as to teachers’ beliefs regarding instruction that may have influenced their practice. Teacher beliefs can help to provide a complete picture regarding why teachers carry out certain practices in the classroom (Charalambos, 2015).

**Observation Videos**

Teachers in the study were asked to video record each lesson that they taught during their unit on energy. Each teacher recorded at least 12 hours of video. Videos were coded using a protocol that was meant to assess quality of instruction developed for the Content Knowledge for Teaching Energy project. The observation protocol was developed to provide insight into the quality of classroom instruction and interactions during a teacher’s unit on energy. The protocol focuses on the nature of teaching around subject matter, in this case, the teaching of energy in physics. It was designed to capture instructional quality with regard to the structure of the lessons, the depth of material that the students had to learn, and the interactions between the teacher and the students, as well as between the students. Each video was scored with respect to a set of 6 dimensions. A description of each dimension can be found in figure 6. Four ordered scale points were developed for each dimension. For example, the discourse dimension was scored based on the presence of student talk focused on energy within the lesson. Lessons with little to no energy-focused student talk during the lesson were scored low on the
discourse dimension, while those lessons that featured students’ sharing ideas and making verbal connections between concepts or feedback loops, scored high on the dimension. For the purpose of this study, videos were used to gain insight into the type of instruction that the students experience in the classroom, the discourse in the classroom, the depth of content of instruction and the ways that assignments and assessments were used during instruction. We specifically focused on the content and discourse dimensions for the purpose of this study.

Figure 6
Dimensions of Observation Protocol

**Lesson Purpose** - This dimension addresses the extent to which the segment has a clear, explicit purpose centered around energy concepts or practices (CKT-E targets). The purpose of the lesson must be clearly communicated by the teacher or students, and the activities in the lesson must be tied to the purpose.

**Lesson Coherence** - This dimension addresses the coherence and connectedness among the concepts/practices and activities presented in the lesson within the context of a lesson purpose connected to CKT-E targets.

**Errors** - This dimension addresses the accuracy and correctness of the teacher's instruction and the teacher's acceptance of incorrect or incomplete student ideas.

**Content** - This dimension addresses the richness of the physics content of the segment in terms of the relevance, depth of content, quality and use of representations, and the quality of explanations developed and/or used by the teacher and students.

**Discourse** - This dimension concerns the quality of discourse within the segment. Discourse is defined as talk around physics topics within the classroom.

**Management** - This dimension concerns the management of behavior and time during the segment. Scores on this dimension are influenced by the behavior of students and the impact that the behavior has on teaching and learning during the segment.

**Case Selection**

Cases were selected to illuminate how teachers with different content knowledge for teaching design and use assignments and assessments during their unit on energy. As such, teachers were selected based on their CKT-E score and the quality and demand of their assignments and assessments. Four teachers were selected, Teacher A, Teacher B, Teacher C, and Teacher D. Teacher A was a fourth-year teacher at the time of the study. They majored in chemistry and completed their teacher preparation in a graduate teacher education program that focuses solely on preparing teachers to teach physics using reformed teaching practices. Out of the 32 teachers in the larger study, Their CKT-E score was 23rd, and their artifacts scored the ninth highest with regard to quality and
demand. Teacher B was a third-year teacher at the time of the study. They completed a major in physics and attended the same teacher preparation program as Teacher A. Their CKT-E was eighth out of 32 teachers, and their artifacts scored the highest with regard to quality and demand. Teacher C was a tenth-year teacher at the time of the study. They were certified to teach physics through an alternate-route certification program and also earned a master’s in science education. Their CKT-E score was fourth out of 32 teachers, while their artifacts were 28th out of 32 teachers. Teacher D was a sixth-year teacher who gained their physics certification through an alternative certification program that certifies physics teachers without the required major or preparation in physics to teach using a set curriculum. Their CKT-E was 28th out of 32 teachers and their artifacts were the 20th most demanding. The school context, grade taught and course taught varied across all 4 teachers. Table 1 provides descriptive information for each teacher and their classroom context.

Table 1
### Case Study Teacher Demographic Information

<table>
<thead>
<tr>
<th>Name</th>
<th>Training Details</th>
<th>Certification Details</th>
<th>Years Teaching</th>
<th>School Context</th>
<th>Grade</th>
<th>Course</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher A</td>
<td>Chemistry Major; Master’s in Education; Reformed-based physics education program</td>
<td>Traditional Route; Physical Science Certification (Can teach Physics or Chemistry)</td>
<td>4 years</td>
<td>Small, suburban school outside of a large eastern US city</td>
<td>9th Grade</td>
<td>College Prep. Physics</td>
</tr>
<tr>
<td>Teacher B</td>
<td>Physics Major; Master’s in Education; Reformed-based physics education program</td>
<td>Traditional Route; Physics Science Certification</td>
<td>3 years</td>
<td>Small, affluent suburban school in the eastern US</td>
<td>10th Grade</td>
<td>Honors Physics</td>
</tr>
<tr>
<td>Teacher C</td>
<td>Engineering Major; Master’s in Education; Science education program</td>
<td>Alternate Route (Second Career); Physics Certification</td>
<td>10 years</td>
<td>Rural school in the western US</td>
<td>11th Grade</td>
<td>College Prep Physics</td>
</tr>
<tr>
<td>Teacher D</td>
<td>Chemistry and Biochemistry Major; Non-Traditional Physics Education program</td>
<td>Alternate Route (Second Career); Physics Certification</td>
<td>6 years</td>
<td>Large school in both an urban and suburban district in the eastern US</td>
<td>10th Grade</td>
<td>College Prep. Physics</td>
</tr>
</tbody>
</table>

Along with variations in their training, each teacher’s practice varied as measured by the CKT-E assessment, artifact protocol, observation protocol, and student
assessments. The ranks of each teacher highlighted here with regard to the 32 teachers in the study for each of the measures are shown in table 2.

Table 2

Case Study Teachers’ Ranks on CKT-E Assessments, Measures of Practice, and Student Assessments

<table>
<thead>
<tr>
<th>Name</th>
<th>CKT-E Rank</th>
<th>Artifact Rank</th>
<th>Observation Rank</th>
<th>Students’ General Science Pre-Test Rank</th>
<th>Students’ Energy Post-Test Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher A</td>
<td>23</td>
<td>11</td>
<td>24</td>
<td>21</td>
<td>18</td>
</tr>
<tr>
<td>Teacher B</td>
<td>8</td>
<td>1</td>
<td>17</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Teacher C</td>
<td>4</td>
<td>28</td>
<td>4</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>Teacher D</td>
<td>28</td>
<td>24</td>
<td>20</td>
<td>20</td>
<td>22</td>
</tr>
</tbody>
</table>

Data analysis

For each case, we first reviewed the scores on each of the instruments developed for the study to determine the relation between the teachers’ CKT scores and the demand of their assignments and assessments. We then analyzed teachers’ interviews, videos of their lessons and observations scores and to help us describe consistencies and discrepancies between their CKT-E score and the quality and demand of their classroom artifacts. For example, in the case of Teacher A (described in detail below), we were able to determine how their teacher preparation provided them with the skills and resources needed to select artifacts that enabled them to engage their students with high quality, demanding assignments and to use the artifacts appropriately during instruction.

Following the data analysis, we employed an explanation building approach (Yin, 2009). In the explanation-building approach, the researcher first identifies patterns in the data and develops explanations for those patterns. Then they revise both over time in light of new data. In this study, we started with an explanation based on the results of a previous study relating the quality and demand of teachers’ assignments and assessments and their
CKT-E scores. In that study we found a significant, positive correlation between CKT-E and the quality and demand of teachers’ assignments and assessments (Zisk et al., 2019a). From the results of this study, we discovered a pattern that the quality and demand of teachers’ assignments and assessments were closely tied and developed the explanation that teachers with higher CKT-E had more knowledge and resources available to them that enabled them to develop and select higher quality assignments and assessments. Through the analysis of the cases however, we were able to determine that the relationship between CKT-E and the quality of classroom artifacts is not always clear cut. We found that a teacher’s preparation plays an important role in how they use artifacts and what artifacts they select for instruction and how teachers’ content knowledge for teaching also plays a role in their decisions concerning the emphasis that they place on assignments and assessments during instruction.

**Cases**

In this section, we will describe the four cases outlined above. We will first describe the teachers’ backgrounds and then review their CKT assessments, artifacts, observations, and student assessment scores.

**Teacher A**

Teacher A was a fourth-year teacher at the time of the study. As an undergraduate, they were a chemistry major, and their training consisted of six physics courses and several math courses. Following their undergraduate training, they attended a master’s level teacher preparation program at a major state university that is focused solely on teaching physics/physical science, and became certified to teach both physics...
and chemistry. They taught physics to ninth grade students and all of their courses were at the college prep level.

**CKT Assessment**

On the CKT-E assessment, Teacher A’s score placed them in the bottom quartile of the 31 teachers in the intensive study scoring -.25 standard deviations below the mean. On the items on the assessment that focused strictly on content knowledge, they correctly responded to half. They did however respond correctly to the item that asked to identify the system of four different scenarios based on the types of energy present in the system and the changes in energy, indicating knowledge of systems and how to determine what types of energy are present in a system.

**Pre-Instruction Interview**

During the pre-instruction interview, Teacher A discussed a variety of both in-depth goals related to energy, and more superficial goals. Teacher A spent a large portion of the interview describing how they believed that it was important for students to learn the names of the different types of energies (kinetic, gravitational potential energy, and elastic potential energy) and described how they wanted the students to be able to describe energy using the different types (Energy Target C1).

“I want them to be able to use the word energy with their appropriate names in front of it. For example, I don’t want them to almost ever say energy by itself. I want them to say gravitational energy this or that, kinetic energy this or that.

I want them to define energy in terms of the individual energies not energy as just one term. I want to fix, I guess, the terms when people, general public use just energy everywhere. I want them to be able to say it’s not just energy, but specific whatever kind you’re talking about.”

Teacher A further went on to describe how they often let the students name the energy types with names that they feel better described the type of energy, such as *speed*
energy for kinetic energy, or height energy for gravitational potential energy. Along with describing the different types of energy, Teacher A also spent a portion of the interview talking about the work as a means of changing the amount of energy in a system (Energy Target D2). Specifically, they mentioned that the students learn about “…work first. The first thing in the unit is work. When we get to the activities, we use the idea of work to develop what happens after you do work to a system. How much energy do you start off with, how much do you end up with and where is it coming from?” They also described how it is important for students to understand how positive and negative work affect the system in different ways.

Teacher A also mentioned how they focus on the energy of the system several times throughout the interview, specifically how work changes the energy of the system, and they also briefly touched on how the chosen system tells you if work is being done and they stated that they starts “without any work being done including everything in the system” (Energy Target B1).

Teacher A further discussed work as a means of helping students to derive equations (Energy Target E2). Specifically, they described in detail how they were able to use the concepts of work and Newton’s second law to derive the equation for kinetic energy. As shown in the excerpt from their interview below, they worked to help the students walk through how to derive the kinetic energy equation based on information they have already learned.

“So, you push a car that’s moving initially and then push it to make it go faster. You start off with some amount of kinetic energy. You end with more and we have to specify that you’re moving on a very frictionless surface because otherwise you have internal involved.
We know that this force is making it faster than before so we use Newton’s second law to say that acceleration is coming from the extra push. The change in velocity is there, so from original to the end.

Then we use one of the kinematics equations to have an expression for acceleration that relates to the distance that it was pushed for, and the initial and these final velocities that were involved.

So, use the kinematics equation to find the acceleration expression, take it and put in Newton’s second law, okay? Then change it in the second law in terms of work. Newton’s second law involves a force, work equation also involves a force and put them next to each other, I guess. We actually just did it today during our meeting.

You end up with something that looks like work equals ½ final velocity squared over 2 minus ½ and times initial velocity over 2. We stop at that and then go back to our normal bar chart stuff for energy and where you’ll see that if you’re just looking at the bar chart, work is going to be equal to final kinetic energy minus initial kinetic energy.”

Along with discussing their goals for instruction, throughout their interview, Teacher A frequently referred to their training when describing how they sequenced the unit and how they selected the activities that they used while teaching. For example, when describing how they start the unit, they immediately describe how they learned to start the unit with work and naming the types of energy from the instructor of all of the courses in their master’s program.

“This comes straight from Instructor\textsuperscript{2}, I want them to be able to use the word energy with their appropriate types in front of it. For example, I don’t want them to almost ever say energy by itself.”

Teacher A also frequently cited the curriculum that was used during their training as a source of the assignments and activities they used in class. On ten different occasions during the interview, they cited the curriculum when describing their sequence of instruction. For example, when talking about work they stated that they pulled the assignment directly from the curriculum.

\textsuperscript{2} Instructor has been used to protect anonymity
“Same from curriculum\textsuperscript{3}, there’s activities that, so we learn work first. The first thing in the unit is work. When we get to the activities, we use the idea of work to develop what happens after you do work to a system. How much energy do you start off with, how much do you end up with and where is it coming from?

Just from curriculum module we’ll do lifting up something from one height to the other and look at what the formula or equation for work ends up looking like for that scenario.”

Additionally, when asked why they sequenced the unit in the way that they did, they stated the following:

“It basically just comes back to my training. I think I haven’t moved from that because I haven’t had any different training since then, or neither did I have any real colleague collaboration to come up with a different way of doing energy. I’m the only regular physics teacher here, so I’ve never had even the chance to collaborate and come up with a different way of doing.

I’ve just been doing curriculum mostly. I know I took physics in high school, 3 of them and 1, 2 and AP and then college some more, but none of it clicked until I did things this way, the way they are in curriculum Module.”

Observations

In the classroom, Teacher A created a learning environment where the students were seated together in groups and worked together to complete their work each day. At each table, the students worked around whiteboards on various problems and tasks. In the lessons, students could be seen working on the same curriculum that Teacher A cited throughout their interview. As they were working, Teacher A circulated through the groups, often engaging in questioning based on what the students were working on.

While they spent time working with groups, the teacher’s questions and conversations were often short, and did not dive deeply into content. A typical exchange can be seen in the excerpt from a lesson:

Student - we did it twice with two different masses because I didn't have the mass for the first one so I had to look it up.

\textsuperscript{3} Curriculum has been used to protect anonymity
Teacher - yeah that looks a little too small, this should be meters. Oh this is better than that.

Student - we got 16 when we did that S: for this part do I use the mass of the bullet and the gun

Teacher - it is still the bullet.

In this exchange, Teacher A noticed that a group of students did not have the correct answer to a problem they were working on and worked with them to diagnose their difficulty. As typical of most exchanges in their classroom, in-depth content was not discussed and they did not question students beyond the exact assignment that the students were working on at the moment.

Using the observation protocol, we found that their lessons often showed evidence of a clear and explicit purpose and coherence between ideas. For example, in one lesson, they explicitly told the students that “today we are going to continue to use the data and experiments to find the equations for gravitational and kinetic energy.” They often included short statements such as these at the beginning of their classes. Nevertheless, Teacher A’s scores placed them in the bottom quartile of the teachers in the study as their lessons consistently scored low on the dimensions related to depth of content, connections between energy and other physics topics and on discourse related to energy ideas.

In addition, Teacher A’s lessons often scored rather poorly on dimensions related to classroom management. The students were often off topic and frequently expressed that they were not motivated to learn the content.

**Assignments and Assessments**

While Teacher A performed in the lowest quartile of study participants on the CKT assessment, their assignments and assessments scored in the top third of the
teachers in the study. In their interview, Teacher A stated that they either developed their assignments and assessments on their own, using items found from different sources, or utilized assignments from the curriculum to which they were exposed throughout their training.

In analyzing their final assessment, we found that the items consisted of a mix of multiple-choice items, as well as open-ended items. The multiple-choice items on the assessment did not include typical definition-type problems and “plug and chug” math items but, instead, contained more demanding items that required students to analyze different situations and compare the energies present, or to develop an equation based on the energies in the situation (figure 7 below).

**Figure 7**

*Sample Multiple-Choice Assessment Items – Teacher A*
Along with the multiple-choice items, their assessment also includes open-ended items that required students to analyze different phenomena and develop energy equations to accurately represent the phenomena using a specific system, such as in the items shown in figure 8 below.

**Figure 8**

Q4. A block starts from rest at the top of a smooth ramp. It slides without friction from the top to halfway down. The earth, block, and ramp are chosen as the system objects. Which of the following quantities is greater?

1. The kinetic energy of the system when the block is at the top of the ramp
2. The gravitational potential energy of the system when the block is at the halfway point of the ramp
3. The two quantities are about equal
4. There is not enough information to say which quantity is greater

Q14. Adventurous Theo was testing out a stunt cannon. He compressed the spring inside the cannon and placed himself in the barrel. He then released the spring and was pushed up. Which of the following Equations accurately represents the situation where Theo and the cannon are the system?

a) \( U_{ki} = U_{kf} + U_{gf} \)

b) \( W = U_{kf} + U_{gf} \)

c) \( U_{ki} + (-W) = U_{kf} \)

d) \( U_{ki} + (-W) = U_{kf} + U_{gf} \)
Looking at the content of the items on the assessment and the knowledge and skills required to successfully respond to the items, Teacher A’s final assessment addressed the same targets that they stated in their pre-instruction interview as goals for the unit. Thirteen of the items on the assessment required students to have knowledge of energy types and the physical quantities that correspond to each type (Energy Target C1), which they mentioned several times throughout their interview. Six items required students to have the knowledge that work can change the energy of a system (Energy Target D2), and several items also required the knowledge of how the choice of a system
could influence the types of energy present and whether work was done or not (Energy Targets B1 and B2). The types of items on the assessment, the range of targets addressed and the consistency between Teacher A’s goals and the content of their assessment indicated that they were able to develop in-depth and demanding assessments.

Along with demanding assessments, Teacher A’s assignments also scored near the top on the artifact demand protocol as compared to other teachers in the study. Much like the items on their assessment, the assignments required students to analyze different phenomena, in some cases to define their system, and then to create energy equations based on the given situation, such as in the items shown in figure 9 below from an “end of class” assignment. As with their final assessment, their assignments also targeted the same concepts and ideas that they mentioned as goals for their energy unit.

**Figure 9**

*Sample Assignment Item – Teacher A*

| 1) After falling 18 m, a 0.057 kg tennis ball has a speed of 12 m/s (the ball’s initial speed is zero). Determine the average resistive force of the air opposing the ball’s motion. Specify the system and the initial and final states. |
| 2) A frisbee gets stuck in a tree. You want to get it out by throwing a 1.0kg rock straight up at the frisbee. If the rock’s speed as it reaches the frisbee is 4.0 m/s, what was the speed as it left your hand 2.8 meters below the frisbee. Consider only the rock as your system and specify the initial and final states. |

*Student Energy Assessment*
Teacher A’s students performed in the bottom quartile of the classrooms in the study on the general science pre-test. On the energy assessment, their students scored at the median, indicating that they performed comparatively better on the energy assessment than they did on the general science pre-test.

**Summary**

Teacher A’s performance on the CKT assessment placed them near the bottom of the teachers in the intensive study, though their interview indicated that they had rather wide knowledge of the important energy targets and concepts, as well as a grasp of how they organized the unit. This indicates that while they were aware of the important concepts that students should learn during a unit on energy, such as the concept of a system and its role in energy analysis and are skilled in mathematical derivations of different types of energy, they lack robust knowledge of content knowledge for teaching energy as valued by the CKT-E assessment. This lack of content knowledge for teaching energy can also be seen in the observations of their classroom, as they scored in the bottom quartile with regard to their observation scores. Despite their poor performance on the CKT-E assessment and low scores on the observation protocol, their artifacts placed them in the top third of teachers in the study and their students performed rather well on the energy assessment in comparison to other classrooms in the study, particularly when taking into account their performance on the general science pre-test.

**Discussion of Case**

While Teacher A had low scores on the CKT-E assessment, they were able to provide their students with high quality assignments and assessments and their students were able to demonstrate a strong knowledge of energy content. We can hypothesize that
this discrepancy comes from the disconnect of two types of content knowledge for teaching introduced by Etkina and colleagues - CKT-D and CKT-P (Etkina et al., 2018). While Teacher A demonstrated relatively strong CKT-D knowledge in their interview and CKT-E assessment, their CKT-P knowledge was formulaic and based mostly on the curriculum approach that they were skilled in due to the physics teacher preparation program that they went through. Often, their questioning and discussions with students did not reflect the knowledge they exhibited on the CKT-E assessment and in the interview.

It is their knowledge and faith in a curriculum that they were trained on extensively and one that they found to be effective for their own learning resulted in their students being exposed to high quality and demanding assignments and assessments. Teacher A relied heavily on their training, which is typical of teachers, particularly in their first few years of teaching (Friedrichsen, 2009; Grossman, 1991), and in their training they gained a knowledge of the important learning targets (CKT-D) and a high quality curriculum to help students learn the targets (CKT-P). However, their lack of strong disciplinary knowledge beyond what they are taught limited their CKT-E.

Overall, in Teacher A’s case, their weakness with regard to personal knowledge for teaching energy was compensated for by their knowledge of the high quality curriculum that they spent time learning over their two years in their teacher education program.

Teacher B

Teacher B was a third-year teacher at the time of the study. As an undergraduate, they majored in physics, and in addition to their physics courses, they also took over five courses in mathematics. After their undergraduate training, they attended the same
physics teacher preparation program as Teacher A and became certified to teach physics. At the time of the study, they taught physics to juniors and seniors at the college prep., honors and Advanced Placement level.

**CKT Assessment**

On the CKT-E assessment, Teacher B’s score placed them in the top quartile of the 31 teachers in the intensive study with the score 1.13 standard deviations above the mean. The CKT-E assessment responses showed that Teacher B had a strong knowledge of energy content as on the items that focused strictly on content knowledge, they responded correctly to all of them. Teacher B’s responses to the content questions indicated that they had a strong knowledge of systems and energy conservation and constancy. For example, when responding to an item that asked to describe the change in energy of a system containing a bicycle, rider, air, Earth and the pavement, when the bicycle is being ridden up the hill at constant speed, they correctly indicated that the energy of the system would stay constant, and explained that, “Every object relevant to the process is in the system so the total energy of the system should remain constant.” This response showed that they understood that when Earth is included in the system, along with all other objects that are relevant to the observed process, no work is done on the system, and the energy in the system remains constant.

**Pre-Instruction Interview**

During the pre-instruction interview, Teacher B mentioned a large variety of in-depth goals corresponding to all the target categories within the energy targets developed for this study (only 6 out of 31 teachers in the study mentioned all target categories in their goals). They mentioned more in-depth goals in their interview than any teacher in
the study. In describing what they believe to be the most important goal of the energy unit, Teacher B spent a significant amount of time discussing the idea of a system, the difference between energy conservation and constancy, and how the choice of system determines if the energy of the system is constant or just conserved (work is done on the system) (Energy Targets B1, B3, and D1). For example, when discussing their specific learning goals for the energy unit, they started off by stating the following:

“So, when we talk about energy, I really focus heavily on the concept of energy is conserved versus energy is constant for a system. I find that especially with really students of any level the biggest challenge they have is understanding what system you’re focused on and that you have to choose the system that it’s not an automatic thing. That if there are no interactions or the interactions all balance each other out external to the system, there’s no change in the energy so it’s constant.

A lot of times I hear from students well in this particular case energy is conserved and then I say well that doesn’t make any sense because energy is always conserved. So, is energy constant here, is that what you mean? If they can get through that concept, everything else is just math. It’s the different types of energies, it’s the equations for them, but if they can really understand what it means to have a closed system and know energy enters or leaves that system, that’s one of the biggest concepts that you can learn in science.”

With this statement, Teacher B describes the importance of defining a system in energy, and that they believe that it is important to distinguish between energy conservation and constancy. They further go on to discuss the importance of a system in all sciences stating, “I mean systems are everywhere, it’s in chemistry, it’s in biology, thermal dynamics yeah it’s everywhere.” To Teacher A, the understanding of a system is the key to understanding energy.

Along with discussing the importance of a system, Teacher B also spent a significant time discussing the idea of energy conservation as a goal (Energy Target D1):

“They really do think that you can just get energy from nowhere or from work and be destroyed and go to nothing. The whole idea of something being a
conserved quantity means it goes somewhere else. It doesn’t disappear it just leaves the system or enters the system from somewhere and you know where that place is and you can calculate it and know how much there is, as opposed to constant which is it doesn’t go anywhere.

It might change to different types of energy inside your system but it’s still all there. Yeah, I just think with the way that we use the terminology with energy and conserving energy in real life when you’re not in a physics context is very different.”

Here, Teacher B discussed the idea of energy conservation. While colloquially, energy conservation is synonymous with energy not changing, Teacher B indicated that it was important to discuss that energy conservation did not mean the energy of a system does not change, but instead it means that you can account for the change in energy.

Along with discussing energy conservation and the choice of system, Teacher B also mentioned that students should develop the idea that work can change the energy of a system and also develop a sense that energy is a scalar quantity (Targets C1, D2 and E3):

“Then when we finally are ready to use the word energy, I’ll say to them okay you come up with your own definition for it and inevitably I get energy is the ability to do work and then I’ll, all right well what’s work? Work is the change in energy of the system. So, okay well that’s a circular definition so try again. We have, usually it takes like half a class period where they’re just constantly discussing, well how do we make a definition for this if it’s circular?

Of course, they end up with work is the change in energy of the system and then energy is, it’s some kind of physical quantity. It’s not a vector, it’s a scalar, there are different types of it. Some types are related to movement, some types are related to gravity or to other kinds of forces and we just kind of have to leave it at that.”

In the interview, Teacher B also described that after discussing the different types of energy that the students would then derive the equation for work (Targets D1 and E2) and then, “derive all of the equations for them based on conservation of energy.”
The final major goal that Teacher B described in the interview was the importance of choosing the correct system for a given analysis (Targets B2 and B3):

“I think it’s important because if you don’t understand what system you’re looking at you don’t know in the case of energy, what types of energy the system has at any given moment. If the entire universe is your system then all of the energy already exists in the universe and it will never change. It will stay constant.

It might transform from one type to another, but if you just focus on let’s say this classroom and then it gets very hot outside and then it starts to warm up the inside of the classroom, the outside external to the system is heating the inside of the classroom.”

Teacher B went on to state:

“In all the derivations that I have my students do for kinetic energy or gravitational potential energy, the only way of deriving those equations is by excluding certain objects from your system and using work to calculate them. So, there’s a really big difference between a net change in energy into or out of the system and if it’s already there to begin with.

If it’s already there to begin with and you don’t know the equation for it, how are you going to calculate it? Only if you know there’s this force exerted over this distance, here’s the equation for this force over this, that’s where you can actually come up with like mgh, right? That’s the force that earth exerts on something multiplied by the distance that it falls if it’s falling and you can use that to derive gravitational potential energy.

Of course, if it was never said out loud I didn’t know until I became a teacher myself and then I had to really understand that, because usually I would think of the system as just the ball or whatever the object is.

But if that’s just your only system and earth is not in that then earth would exert a force on it and then work would be done, and then you can’t say that there’s work done by the earth and there’s also gravitational potential energy because then you’re double counting how much energy there is.”

Through this statement, Teacher B discusses that when conducting energy analyses, you need to be careful to choose the correct system. Depending on the goal of the analysis, different systems may be more advantageous to use than others.

Along with discussing their goals for instruction, Teacher B also provided information as to how the class is structured. They noted in the interview that the class is
mostly comprised of students working in groups and mentioned that working as a, “whole class is not usual for him, that’s definitely out of the norm.” Teacher B also made several references to activities they learned in the teacher preparation program.

**Observations**

The observations of Teacher B’s classroom and instruction scored at the middle of the teachers in the study, and their depth of content scores (the depth of content scores were based on the presence and quality of explanations and the connections among the content made by the teacher) were also near the middle of all of the classrooms studied. However, the scores for classroom discourse were in the top third of the classrooms studied.

In Teacher B’s classroom the students sat individually for portions of the class, but generally sat in groups working together on problems and laboratory investigations. When the students were sitting at their desks, Teacher B typically conducted mini-lectures and also reviewed problems with the students. In these segments of the lessons, they also took the time to address any difficulties they noticed when the students were working on problems in groups. For example, in the exchange shown below, the students worked on a problem in which they were trying to determine if Earth does work on an egg depending on if Earth is in the system. Teacher B noticed that the students calculated both the gravitational potential energy of the egg-Earth system and accounted for work being done by Earth as the egg was dropped, resulting in double the amount of kinetic energy being calculated for the egg. This “double-counting” is a common difficulty that students have in analyzing the gravitational potential energy of a system.
Teacher - Let’s break this down more, a three-part bar chart and I want you to fill this in for yourself, here it’s on the floor, here it’s being lifted, and in the third part it’s falling, maybe halfway.

Student - There’s no energy in the beginning. Then it has energy when I lift it.

Teacher - We already did this one, there’s no energy there. When you lift it we said you and Earth both do equal amounts of work, now that it’s been lifted it’s still not moving, no likelihood for it to break the egg, not including Earth in the system.

Student - The egg gains energy when you lift it.

Teacher - If we focus on the earth as part of your system, this is what Ethan was talking about before, but if you’re focused on the egg and only the egg, the “in” between bar is energy given or taken away, so you can’t double count here, if you want to say there is some likelihood up here it’s because of Earth, it’s something to do with the egg and Earth, not just the egg. It’s just because of the system I’m using, the egg and only the egg not Earth, now it’s falling moving so there will be some amount of energy.

In their pre-interview, Teacher B mentioned that they are aware that system choice and double counting could be an issue for students. In this exchange, the teacher takes time to address the entire class regarding choosing the correct system and double counting. Exchanges such as this are common in their classroom and they are used to address difficulties and ideas that many students in the class may have.

Along with whole-class instruction, the students in Teacher B’s classroom spent a significant amount of time working in groups. In these groups, they worked together to solve problems, and Teacher B took the time to work with each group, addressing their questions and assessing them as they were working. In these exchanges, Teacher B did not tell the students the correct answer, but instead provided them with carefully thought through questions and engaged them in discussions that could potentially help them solve the problem by themselves. It was when the students were working in small groups that a large majority of the discourse in the classroom occurred, both among the students and
between students and Teacher B. For example, exchanges such as the two below were common.

*Exchange 1*

Student - work is change in energy right

Teacher - what is this .3 squared times 70? I don’t understand

Student - actually it is KE, right?

Teacher - You’re not following like any of the problem-solving steps. You're not writing any of your numbers with the variables. You're not drawing a picture and labeling it. You’re not doing an energy bar chart

Student - what is the point of having the cars…does that take all…

Teacher - what does that mean? The person in the car

Student - *inaudible*

Teacher - what is your system

Student - *inaudible*

Teacher - including the piston or not including the piston?

Student - No. Not including

Student 2 - Not including

Teacher - Okay, if you are including the piston then can it do work on the system?

Student - No its inside the system.

Teacher - No. No work is done by the piston because it is in the system. But if you exclude it from the system then it does do work.

Student - Then we could use the equation for work.

*Exchange 2*

Student - I am not sure where to go from here
Teacher - Where is your energy bar chart? (Pointing at the bar chart) This minus this equals zero and you know the equations for each of these.

Student - what did I do for this?

Teacher - think about what we did yesterday. You need to think about the mass of Earth and the mass of the rocket.

Student: inaudible

Teacher - You need to make sure your estimates are reasonable. How tall is a floor,

Student: 2 meters or 3

Teacher - Don't make it 10 and then every floor would have giant ceilings.

Student - That would be cool.

Student - That is why we have Google; I have all the numbers.

In both of these exchanges, the students were working on problems in their groups and Teacher B offered them assistance in the form of questioning. The teacher never directly gave the students the answers but questioned them in a way that they were able to complete the problems on their own. Through interactions such as these, Teacher B was able to help students develop concepts, complete problems, and work through any difficulties they might have had without ever directly telling them the answer.

Assignments and Assessments

While Teacher B’s observations scored at the middle of all the teachers in the study, their assignments and assessments scored the highest. Their assignments and assessments scored high for the depth of analysis required to solve the problems, the depth of mathematics and the use of representations. In each of the artifacts submitted by
Teacher B, the students were required to complete multi-step problems that could not easily be solved by simply plugging values into a known equation.

In analyzing their assignments and assessments, we found that each of the assignments presented the students with open-ended questions that required students to define their system, and use multiple representations in order to solve the problems. These problems directly addressed the important goals that Teacher B mentioned for the students in their pre-interview. For example, in the problem below (Figure 10), the first one from a quiz and the second from the final unit assessment, the students are asked to define their own system and develop energy bar charts for the phenomena based on their chosen system before solving. Teacher B also requires the students to state their assumptions when solving many problems, such as in the problem shown in Figure 10.
Along with problems such as these, Teacher B also included more conceptual problems throughout their assignments and assessments that did not require mathematics to solve, but instead required the students to deeply understand energy concepts. For
example, in the item below (Figure 11), the students had to explain why it is important to conserve energy if energy is a conserved quantity. In order to answer this question, students need to have an understand of the difference between energy conservation and energy constancy and also have an understanding that when energy changes forms, in can be converted into a type of energy that is no longer useful, such as when you bounce a ball on the floor and the floor warms up slightly. Teacher B mentioned this understanding as an important goal for their students within the energy unit.

**Figure 11**

*Sample Open-Ended Assessment item, Conserved Quantity – Teacher B*

> 3. Why are people so concerned with saving energy if energy is a conserved quantity? C1 2

Another conceptual question is shown in figure 12 below. In this question, the students are asked to decide whether the floor does work on you when you are walking up the stairs and justify their answer. It is a very complicated problem as the floor does not move in this situation and thus cannot do work on the person. Where does the energy come from? It is also important to note that even though the problem did not require the students to include an energy bar chart, more than half the students included one in their answer.
Throughout all their artifacts, Teacher B presented the students with problems that required them to have knowledge of systems and to distinguish between energy conservation and energy constancy. The teacher expected the students to use multiple representations when solving the problems.

**Student Assessment**

Teacher B’s students performed in the top third of the classrooms in the study on both the general science pre-test and the energy assessment. On the pre-test, their students scored the fifth highest, while on the energy assessment they scored the fourth highest.

**Summary of the Data**

Both Teacher B’s strong performance on the CKT-E assessment and the wide range of in-depth goals they set for their students during the energy unit indicated that they possessed a strong knowledge of energy concepts and how to help their students learn those concepts. Their interview also indicated that they had a strong knowledge of common student difficulties within the unit, as well as a knowledge of how the unit should progress and how students’ knowledge develops throughout it. Their observation scores were near the middle of the teachers within the study, though their discourse...
scores were in the top third of teachers in the study. Their written assignments and assessments were rated as the highest quality in the study and required the students to have knowledge of systems and to use multiple representations. All of the problems presented to the students were open-ended and required multiple steps to solve. The steps were not spelled out for the students. At the end of the unit, Teacher B’s students performed well on the post assessment, scoring the fourth highest out of all the classrooms in the study.

When we examine all of Teacher B’s scores as a whole, the observation score stands out as much lower (relatively) than the rest of the scores. It is possible that Teacher B’s observation score might have been influenced by the nature of their instruction and the nature of the observation instrument. The dimension of the observation protocol that lowered the average observation score for Teacher B was the depth of content of the lessons. When measuring the depth of content, the observation protocol valued explicit teacher explanations of the content and the connections among different concepts made by the teacher. As the students in Teacher B’s class spent the majority of class time working in groups, there were significant portions of the lessons without any explanations made by the teacher. Despite this score, when we consider the data in its entirety (including the assignments and assessments that the teacher designed and used during instruction and student post-test scores), it is clear that the level of the content in Teacher B’s classroom is very high. For example, when we look at the example item in figure 11, Teacher B is not only asking the students the meaning of a conserved quantity, but also asking the students to explain the significance of energy being a conserved quantity and also asking them to explain the difference between energy
constancy and energy conservation. This is indicative of the high level of conceptual understanding that Teacher B expects of their students.

**Discussion of Case**

Based on the data collected for Teacher B we find coherence between their personal knowledge of the content, their content knowledge for teaching as judged by the CKT-E assessment, their ability to help students learn through observed classroom interactions, and the written artifacts that they used during instruction and for final assessment of student learning. The CKT-E test, the interview, lesson observations and artifacts all point to the high levels of CKT-D and CKT-P possessed by Teacher B. Practice based conceptions of CKT state that teachers use their knowledge of the content (CKT-D) and how to teach the subject (CKT-P) to enact high quality teaching practices (Gitomer et al., 2014; Hill et al., 2008). In the case of Teacher B, they used their CKT-E to engage the students in high quality learning of physics, in which they collaborated with each other, and were continuously held accountable by the teacher for their learning. During whole-class instruction, Teacher B worked to address common student difficulties that their students encountered while working on problems. The teacher used collaborative group work to help students develop deep conceptual understanding and problems solving skills while working together. While students were working in groups, Teacher B used their CKT-D and CKT-P to listen to and assess students, and to help move them forward without providing them with the answers. Teacher B used carefully chosen questions that helped the students move along. We found clear coherence between Teacher B’s personal knowledge for teaching energy and the goals that they formulated for the energy unit, and clear coherence between the goals and the assessment items that
they used to find whether the students had achieved those goals. Teacher B’s questions required students to demonstrate deep understanding of the fundamental energy concepts (such as the concepts outlined by the energy targets), as opposed to the knowledge of how to solve simple problems using equations for work and different forms of energy.

**Teacher C**

Teacher C was one of the more experienced physics teachers in the study, teaching physics for ten years. Prior to teaching, they trained and worked as an engineer and their undergraduate education consisted of a mix of physics, science, and engineering courses. They obtained their physics teaching certification through an alternate certification and then obtained their master’s in education in a science education focused program. They taught both college prep level physics and honors physics to juniors and seniors and for the purpose of this study, they were observed teaching an honors course.

**Content Knowledge for Teaching Assessment**

Teacher C performed well on the Content Knowledge for Teaching Energy assessment. Among the 32 teachers who took the assessment as part of the intensive study, their score ranked him fourth, scoring 1.67 standard deviations above the mean. Their performance on the CKT-E assessment showed that they had a strong knowledge of energy content as they correctly responded to all seven items on the assessment that focused strictly on energy knowledge. Their responses to several items on the assessment indicated that they had strong knowledge of energy conservation and systems. For example, when responding to an item that asked to describe the change in energy of a system containing a bicycle, air and the pavement the bicycle is being ridden up the hill at a constant speed, Teacher C correctly indicated that the energy of the system would
increase, and explained that, “Every object relevant to the process is in the system so the total energy of the system should remain constant.” This response shows that they understand that when Earth is included in the system, along with all other objects that are relevant to the observed process, no work is done on the system, and the energy in the system remains constant.

**Pre-instruction Interview**

Along with performing well on the CKT-E assessment, during their pre-instruction interview Teacher C described a wide range of in-depth goals for their students during the energy unit (for the purpose of this study, we classified a goal as in-depth if it met one of the student energy targets identified for the intensive study). Among the teachers in the study, Teacher C’s goals for their students were more robust than most, as they stated goals that met 16 energy targets, while most teachers indicated goals that met between 6 and 7 targets, and the highest number mentioned by any teacher was 17. Teacher C placed a significant emphasis on the idea that work is the way the energy is transferred to or from a system mechanically (Energy Target D2). For example, as shown in a quote from the interview, Teacher C described how they build off the common definition of work that students come into their class with and moves towards the physics concept of work as a product of the force exerted on an object and its displacement and finally to the idea that work changes the amount of energy in a system.

“I think just even divorced from the physics idea of work, I think they get the idea that I lifted something, and I think that takes energy. They don’t explicitly say it, but they think about it. So, there is the, I have lifted something, I go to the gym and I spend a lot of energy, and I burn calories. I just lifted something here, but it was pretty easy, because of those pulleys. So, I start thinking about work that way… And then I like that I’ve got this definition of work as, the product of work and displacement. And then from that, I can work on ‘I am transferring energy to other objects, by doing work on them.’”
In their interview, Teacher C also described how they have the students use the concept of work to then build the equations for different types of energy (Targets C1 and E2). For example, when describing how students develop a definition for gravitational potential energy and its equation, by using the idea of work and lifting an object to develop both the definition and equation.

“We then look at doing work in lifting an object. Eventually they come up with “I do a certain amount of work, and an object changes height proportional to the amount of the work I do”. And we end up coming with a definition of gravitational energy, as the energy that an object gains that comes from a person lifting it to a different height than it previously was at.”

Teacher C then continued:

“I talk about the idea that having a definition of mgh, the H is only meaning in the context of it being relative to some reference height. The standard reference height we have is sea level. Then, that g in there comes from the force gravity acting on the object, which specifies the minimum force I need to apply to be able to lift the object.

So, when I’m talking about an object with gravitational energy, I give it energy when I lift it, because gravity is opposing my force. It’s proportional to the height that I lift it, because I’m separating that object from its initial position. So, I’m moving it to some higher position.

So, I talk about gravitational energy being the energy that an object has, in relation to the earth. So, at that point I talk about the earth, and the object being a system of objects, and I am adding energy to that system.”

Teacher C also placed significant emphasis on the importance of system choice in energy analyses (Target B1). In describing their unit on energy Teacher C stated, “But then once we get into conservation of energy, we start to look at the different ways you can configure the objects in a system, and advantages and disadvantages of those.”

Teacher C spoke at length about how the choice of system determines if work is being done on the system or not and also the difference in energy analysis when Earth is
included in the system and when it is not as evidenced in the quote above. In the quote below, they described interactions with students that help them to develop the idea.

“And while we’re going through that, some people will inevitably be saying like, well, I didn’t include this object in the system. So, we’ll just say okay well, you work through your scenario. We’ll work through our scenario, then we’ll compare and then we talk about the advantages and disadvantages of each.

Teacher C then continued:

Then for the other scenario where a group had included all the different possible objects in their system, we’ll say that, in this scenario it’s kind of nice. You didn’t have to have any work done on the system and energy was conserved.”

Teacher C’s interview also provided insight into their instruction. They stated that they wanted students to work in small groups so that they could share ideas, work on concepts together, and actually get to carry out experiments, such as pulling on a spring or pushing a bowling ball. They also discussed the importance of conferencing with the students, stating, “I want to be able to talk to a bunch of different kids, and ask them why they thought something happened in the demonstration, and did they have a different takeaway than I expected them to have.” To Teacher C, the time spent talking with students allowed them help the students develop knowledge of the important concepts, to assess their progress and to see where they need additional support, and they also stated that above all else, “Talking in small groups is the most effective thing I can do,” with regards to assessing their students.

Observations

Observations of Teacher C’s classroom and instruction scored fourth highest out of all the teachers in the study and the depth of content of their lessons and discourse between the teacher and the students scored the highest out of all those in the study.
During each lesson, Teacher C’s students would come into the room and immediately begin working together on a prompt that the teacher provided to them at the beginning of class. For example, in one lesson, focused on gravitational potential energy, the students were presented with a drawing of different objects at different heights, so at different distances from the edge of a surface, and were asked to compare the gravitational potential energy of the different object-Earth systems. The teacher would then call groups one by one up to the front of the room and conference with them, asking them about the day's assignment. In the small groups, Teacher C focused on the in-depth concepts that they stated as goals for the energy unit. This type of instruction matches their description in the pre-instruction interview.

For example, in the following exchange, Teacher C is helping the students to develop the idea of gravitational potential energy and its relation to height.

Teacher - I have a question for you guys so I have a ball on the ground with a mass of 5kg and I am going to call the ground zero height does this object have any GPE

Student - no because the height is zero

Teacher - so I can calculate it and I have zero joules

Student - yeah

Teacher - so now I lift the bowling ball up 1 meter, does it have GE now

Student - yeah

Teacher - how much S: 5 kg?

Student - 1 meter, it would be 50

Teacher - 5 x 1 x 10.

Student - we were thinking because it was .5 before
Teacher - yeah, so I just push it a little to the right and stick it on the shelf does it have GE now

Student - no

Student - yeah, it's not moving
Teacher - it wasn't moving when it was in my hand either

Student - it has height though.

Teacher - the gravitational energy doesn't have to do with the fact that it is moving or not

Student - no

Teacher - I moved it along the shelf to the back corner of the shelf

Student - it's the same thing.

Teacher – why?

Student - you just said it. It is still going to be at a height of 1 meter, just because you are moving it to a different location, it doesn't matter.

Teacher - okay so you don’t care that it is less likely to fall

Student - it's in the corner

Teacher - yeah here it is at the end of the shelf

Student - but it's still at one meter.

Teacher - so you guys say that it is at 1 meter that is how much GE it has, it doesn't matter that it is moving or not moving or perched on the edge of the table or whether it is tucked back in the corner

Student - yeah

Teacher - so umm so bowling ball back on the ground again, and I walk the ball up the hill and set it on the ground here. Does it have GE at the top of the hill

Student - and this is the ground

Teacher - why did you draw a horizontal line here?

Student - that is where the zero is.
Teacher - so I am laying on the ground at the bottom of the hill or laying on the ground at the top of the hill, are you distinguishing between them?

Student - it's just like the shelf, it's just there is nothing like to hold the hill up

Student - it is still being raised a height

Student - it has a height

Teacher - so somebody did work to raise it this distance maybe two meters so if my shelf was made of dirt it will still have the energy.

Student - wait. Okay so we said work is the same as GE but what we are talking about when you say we move it up a hill is that a different distance if we move it up vertically or horizontally?

Teacher - no because if I move it up a hill I just lift it up and once it is at that height

Student - it's like the table, there is nothing between the legs T: but it is still at a height of 1 meter, so what happens if the ball is on the ground and it is at a height of zero and I calculate Eg. But I did notice, way over here the hill went down and the ball rolled down the hill. Does it have neg Eg then?

Student - no because it is at zero.

Teacher - but I already called it zero.

In this exchange, Teacher C is clearly helping students to relate height and gravitational energy. Through this discussion the teacher is assessing the students’ knowledge of the relationship between height and potential energy, and they address the idea of a relative zero point as the student discusses how they set the bottom of the hill as the zero point and so the potential energy of the system increases when as the ball moves up the hill. Another student then expresses that they do not think the potential energy of the system increased because it is still on the ground. Teacher C is then able to address this student’s idea and move them toward a more complete understanding. In this conversation, Teacher C also addresses a common difficulty that some students have
regarding confusing the potential for an object to fall (if it is close to the edge of a shelf or farther away) and gravitational potential energy. Through interactions such as these, Teacher C is able to help students develop concepts, assess their knowledge and address any difficulties they may have.

**Assignments and Assessments**

While Teacher C demonstrated a strong knowledge of energy concepts and how to teach energy, they indicated in their pre-observation interview that they do not rely on assignments or assessments to assess students’ true understanding of the content. In fact, in talking about their final assessment, they stated “The goal would be you were awake and generally focused in class and you should get a 50% on it. If you were awake and focused and kind of leading your lab group through some of the activities, or at least thinking actively and asking me questions you should get 75% on it.”

Analyzing Teacher C’s assignments and assessments we found that the final assessment for the energy unit was the 2nd least demanding out of all final assessments of the teachers in the study. The items on the assessment were multiple choice and simple “plug and chug” math problems and basic conceptual questions. Additionally, the problems on the assessment did not match the goals for instruction that Teacher C stated in the interview. For example, though they stated that an important goal for their students was to have the students understand that work can change the energy of a system, the majority of the questions related to work on their assessment were similar to the questions in figure 13. Similarly, though Teacher C discussed the importance of system choice in their interview, very few items on the assessment even addressed the system choice or even mentioned the word “system”. In analyzing the three other assignments
Teacher C provided for the study (two quizzes and 1 lab assignment) this finding was repeated as they were rated as the 3rd least demanding among artifacts used by the teachers in the study.

**Figure 13**

*Sample Assessment Items – Teacher C*

1. Determine if the work done by the force of gravity is positive or negative or zero in the following situation:
   
   A soccer ball is kicked by a soccer player horizontally.
   
   a. Zero  
   b. Positive  
   c. Negative  

2. A 20.0-newton force is used to push a 5.00-kilogram cart a distance of 2.00 meters. What is the work done on the cart?
   
   a. 40.0 J  
   b. 100.0 J  
   c. 150.0 J  
   d. 200.0 J  
   e. 240.0 J  

**Student Assessment**

Teacher C’s students performed in the top third of the classrooms in the study on both the general science pre-test and the energy assessment. On the pre-test, their students scored the ninth highest, while on the energy assessment they scored the seventh highest.
Summary

We find coherence among most sources of data related to Teacher C. Both Teacher C’s excellent performance on the CKT-E assessment and the wide range of in-depth goals they set for their students during the energy unit indicate that they have strong knowledge of energy concepts and how to help their students learn these concepts. The interview responses also indicated that Teacher C was well aware of how the unit evolves, and how the students work to develop the concepts within it. Teacher C’s observation scores were among the highest in the study and it was clear from both the observations and the interview that Teacher C valued student interactions and conversations in class and used verbal assessment during lessons. At the end of the unit, Teacher C’s students performed well on the energy post assessment, scoring the 7th highest out of all of the classrooms in the study.

Only one score for Teacher C was low. The written assignments and assessments used by Teacher C were among the least demanding in the study.

Discussion of the Case

Based on the data collected for Teacher C we find a discrepancy between their personal knowledge of the content and how to help students learn it when observed in oral interactions and the content knowledge for teaching energy as judged by the written artifacts that they used during instruction and for final assessment of student learning. The CKT-E test, the interview, and lesson observations point to the high levels of CKT-D and CKT-P possessed by Teacher C. The students in their classroom engaged in high quality learning of physics, they collaborated with each other, and were continuously held accountable by the teacher for their learning. And yet Teacher C’s knowledge was not
reflected in the assignments and assessments they created for instruction. In this case, relying on just teacher knowledge without gaining into their beliefs related to teaching and assessment would yield an incomplete picture of their true teaching practice (Charalambos, 2015). Though most often, assignments, assessments and quizzes are seen as the typical way to assess students’ knowledge (Salend, 2009), to Teacher C, traditional assessments were merely a way to see if the students were paying attention throughout the unit, and as such, they do not offer insight into their knowledge or ability to teach energy. The low scoring artifacts of Teacher C do not necessarily indicate that they are unable to create in-depth and robust assignments and assessments, especially as they indicated in their interview that they did not place emphasis on paper and pencil tasks. Instead, in Teacher C’s case, the real instruction and assessment of students’ knowledge occurred during the conferences and meetings with the students. In these meetings, Teacher C relied on their strong CKT-D and CKT-P to listen to the students, evaluate their knowledge and respond to them in the moment, which is often more difficult than creating and scoring a paper and pencil assessment as it requires a deeper knowledge to be able to quickly respond to student ideas as they are heard. Their beliefs regarding assessment, along with their CKT-E, allowed them to create a high-quality learning environment for their students without using higher-quality “traditional” assessments.

**Teacher D**

Teacher D was a seventh-year teacher at the time of the study, teaching physics for six of those years. As an undergraduate, they were a chemistry and a biochemistry major, and their training consisted of four physics courses and five math courses. Prior to teaching, they worked as a chemist for five years. They obtained their physics teacher
certification through an alternate certification program that provides teachers who have been certified in other content areas with a physics certification after being trained to teach using a specific curriculum. Taught physics at both the college prep and honors level.

**CKT Assessment**

On the CKT-E assessment, Teacher D’s score placed them in the bottom quartile of the 31 teachers in the intensive study scoring -.46 standard deviations below the mean. On the items on the assessment that focused strictly on content knowledge, they correctly responded to half. On the two questions that required knowledge of systems, Teacher D responded incorrectly.

**Pre-Instruction Interview**

During the pre-instruction interview, Teacher D discussed both in-depth goals related to energy, and more superficial goals. They started off the interview by discussing that their students learn about work first and that applying a force over a distance can add energy. Specifically, they stated:

“So, the unit starts actually with the discussion on work. What work is, because they know they understand forces now, from earlier in the year. So, we actually start with how a force applied over a distance can add work or energy.

Then we go from there to I setup a problem of, I think it’s like a weightlifter. How much work does the weightlifter do in lifting the barbell above his head, something simple like that?”

In this statement, they addressed the idea that work can change (add) energy to a system (Target D2) and later mentioned that work could cause a system to lose energy as well. Additionally, at a later point in the interview they mentioned that work is done on a system (as opposed to being done on an object).
Despite mentioning systems in relation to work and mentioning that the energy of a system is conserved, Teacher D made no other mention of systems in the interview other than talking about the different types of energy present in a system.

Teacher D spent most of the interview discussing the different types of energy and how students were able to identify what types of energy were present (Energy Target C1). For example, when talking about gravitational potential energy, they stated, “just being able to identify, this object has a height and therefore has gravitational potential energy, seems really, really simple.” They made similar statements regarding kinetic and elastic potential energy.

Along with defining the different types of energy, Teacher D also mentioned that the students would derive the equations for each energy type, but then states that, “We conclude that day with understanding the work done, by the force of gravity. Then equate that to gravitational potential energy. Then I give them the equation.” This statement contradicts their statement that they wanted students to derive the equation, as they stated that they provided it to them.

During the interview, Teacher D also spent a significant amount of time discussing how they wanted students to be able to solve problems using the equations for different types of energy and work. For example, when they stated that they provided the students with the equation for gravitational potential energy, they followed that statement by saying “Then we use it to solve several problems.” They went on to say that the students use a personal response system to respond to questions:

“Throughout all of their problem solving for gravitational potential energy, as they are getting comfortable with the equation and the variables in the units, they are responding numerically. It’s not a multiple choice. They sit, solve the problem and take the number into their remote.”
Statements such as this one indicate that the students spend a portion of their class time working on traditional, “plug and chug” math problems.

Teacher D’s interview also provided some insight into their training and the curriculum they used. They mentioned that they were a chemistry teacher and were asked by their school to start a physics program.

“I was not a physics teacher before that. I was a chemistry teacher before that. Both of my undergraduate degrees are in chemistry.

But where I was teaching wanted to bring physics into the school. So, they asked some of us to go and become certified in physics, and so I did.”

Continuing on, they mentioned that when discussing the materials for the unit, “The whole presentation comes from something called the training program.” This statement indicated that, at least for the energy unit, they relied heavily on the materials they learned to use in their training program.

**Observations**

In the classroom, Teacher D’s students were sitting in rows, facing the front of the room. The majority of the class consisted of presentations and lectures using pre-made slides from the curriculum that they were trained to use during their certification program. During each class, they would present information on the slides for several minutes and then stop the class and have the students respond to a set of questions using a personal response system. An example of a typical segment of the class can be seen in the exchange below.

Teacher - So if the force and distance are perpendicular, then the work done is 0. They must be parallel only to the distance. When perpendicular the work is 0. Okay. One quick note on units. The units for work and energy. Newton meters is

---

4 Training program used to protect anonymity.
the unit for energy and work and it's given a special name because of this guy James Joule, so our units are Joules and they have the symbol J, so on your reference paper you have several variables and units. F and D, so units for force are N, units for distance are m and what is the unit for energy?

Student - J

Teacher - Another note, if you are given a number and its kJ and its kN what does kilo mean? 1000. So if I have a number that is 1000J we need to convert this so take your number and multiply it

The teacher then talks about using the personal response system.

Teacher - The question is behind me, read it make a list of givens, we have 5 minutes lets see if we can get through 5 questions

Teacher - If a positive 24N force is applied and moves 3m what is the work done?

In this example, the teacher is lecturing to the students about work and its units.

They are focusing strictly on the definition for work and not working to develop a deeper understanding. These exchanges are typical throughout their teaching.

In analyzing their lessons using the observation protocol, their scores placed them in the bottom half of the teachers in the study. Their lessons consistently scored low on the dimensions related to depth of content, and those related to student discourse. For each of the dimensions that measured student talk and ideas present in the classroom, they scored in the bottom quartile. Additionally, their score for use of representations while teaching was the lowest of the teachers in the study. Their total observation scores were enhanced by evidence of a clear and explicit purpose in several lessons and coherence between ideas, along with their management of time and behavior in the classroom, as students were often sitting at their desks listening to the lecture.

Assignments and Assessments

In analyzing the assignments and assessments that Teacher D used during instruction, both their assignments and assessments scored in the bottom quartile of the
teachers in the study. In their interview, they stated that they exclusively used the assignments and assessments developed for the curriculum they used during their training.

In analyzing their final assessment, we found that the items consisted of multiple-choice items and the items that had one right answer. The multiple-choice items on the assessment included typical definition-type problems and “plug and chug” math items (figure 14 below). The first item is a typical definition-type problem, but requires the students to have knowledge of centripetal force and work. Adding centripetal force to the problem enhances the difficulty but does not require a deep understanding of energy content. The second item in the example is typical of a “plug and chug” item, but it also requires the students to have a knowledge of the definition of work to know that if there is no displacement, no work is done.

Figure 14

Sample Multiple-Choice Assessment Items – Teacher D

2. Does the centripetal force acting on an object do work on the object?
   A) Yes, since a force acts and the object moves, and work is force times distance.
   B) Yes, since it takes energy to turn an object.
   C) No, because the object has constant speed.
   D) No, because the force and the displacement of the object are perpendicular.

3. A 50 N object was lifted 2.0 m vertically and is being held there. How much work is being done in holding the box in this position?
   A) More than 100 J
   B) 100 J
   C) Less than 100 J, but more than 0 J
   D) 0 J
There were also a few items that required students to analyze a situation and compare two values (figure 15 below). In the first item shown, the students need to know the definitions of kinetic energy and potential energy and also understand what is happening to an object as it is being thrown up to answer, but overall, the situation that is being analyzed is rather simple. Similarly, in the second item, the students need to know the equation for elastic potential energy to answer it correctly.

**Figure 15**

*Sample Multiple-Choice Assessment Items, Comparing Values – Teacher D*

<table>
<thead>
<tr>
<th>A stone is thrown straight up. Compare its kinetic energy KE, to its potential energy GPE as it is going up.</th>
</tr>
</thead>
<tbody>
<tr>
<td>A) KE increases and GPE decreases.</td>
</tr>
<tr>
<td>B) KE decreases and GPE decreases.</td>
</tr>
<tr>
<td>C) KE increases and GPE increases.</td>
</tr>
<tr>
<td>D) KE decreases and GPE increases.</td>
</tr>
</tbody>
</table>

How does the work required to stretch a spring 3m compare with the work required to stretch it 1m?

<table>
<thead>
<tr>
<th>A) same amount of work</th>
</tr>
</thead>
<tbody>
<tr>
<td>B) 3 times the work</td>
</tr>
<tr>
<td>C) 6 times the work</td>
</tr>
<tr>
<td>D) 9 times the work</td>
</tr>
</tbody>
</table>

Similar to the multiple-choice problems, the open-ended item on the assessment did not require a deep understanding of energy concepts (Figure 16 below). While the phenomenon appears complicated with a spring, loop, and ramp, the problem is essentially a “plug and chug” problem due to the guidance with the questions. To be successful though the students needed to have an understanding of energy constancy in an isolated system in order to know that the total energy stays the same throughout.
Overall, the assessment only required the students to know the definitions and equations for the different types of energy (Target C1) and the equation for work and that work requires displacement (Target D2, but at a surface level), and have an understanding of energy constancy (Target B1, but at a surface level). This matches the goals Teacher D mentioned in their pre-instruction interview, with the exception that the assessment did not require students to understand that work changes the energy of a system.
Teacher D’s assignments were similar to their assessment and scored in the bottom quartile of teachers in the study with regard to quality and demand. One of the assignments they submitted was a lab quiz that assessed students’ knowledge of what they did in a spring-launcher lab that verified constancy of energy. The other assignments they submitted asked students to solve various plug and chug items similar to those on their assessment.
Student Energy Assessment

Teacher D’s students performed in the bottom quartile of the classrooms in the study on the general science pre-test and the energy assessment. On the pre-test they scored the fourth lowest of the classrooms in the study and on the energy assessment, they scored the second lowest. This indicates that their students performed similarly on both assessments.

Summary

Teacher D’s performance on the CKT assessment placed them near the bottom of the teachers in the intensive study, and their interview indicated that they had a rather narrow understanding of the important concepts within energy. This indicates that they have a lack of both the concepts that students should learn during a unit on energy and the knowledge of how to help their students develop these concepts. This lack of knowledge for teaching energy can also be seen in the observations of their classroom, as they scored in the bottom quartile with regard to their observation scores. Their classroom had a traditional setting with little collaboration between the students and assignments that required to find the right answer for well-defined problems. Similarly, their assignments and assessments did not require a deep understanding of energy concepts to complete and focused mostly on definitions and simple math problems. Their students also performed near the bottom of the classrooms in the study on both the pre-assessment and energy assessment.

Discussion of Case

Data collected for Teacher D indicated low level of CKT-D and CKT-P knowledge as evidenced by the results of the CKT-E assessment, classroom observations,
and artifacts used by the teacher during the energy unit. Teacher D relied heavily on their training, which is common when analyzing teacher practice (Friedrichsen, 2009; Grossman, 1991), and the curriculum that Teacher D was experienced in and used during the lessons did not have physics mistakes, it did not engage students in the active exploration of the material. The students were expected to learn from lectures and slide shows and then solve simple problems. This is exactly what teacher D did. This resulted in the students not learning the concepts deeply and moving down in rank on their energy assessment compared to pre-instruction general science assessment.

**Discussion**

This study was based on the premise that a teacher’s CKT influences all aspects of teaching practice (Ball et al., 2008). As such, the goal of the study was to investigate this premise by closely examining the consistency between teacher’s Content Knowledge for Teaching Energy and multiple measures of classroom practice, as well as student achievement. In doing so, we addressed two research questions, first,

- To what extent do teachers demonstrate consistency among different measures of their content knowledge for teaching energy and student achievement, and second,
- How does the framework of Content Knowledge for Teaching explain consistencies and discrepancies among different measures and student achievement?

We investigated these questions through case studies of four teachers who were a part of an intensive study of their CKT-E and their teaching practice related to energy.
We first examined teachers’ consistency across CKT-E assessment scores, measures of classroom practice and assignments and assessments, and student achievement related to energy. In previous work, we found significant, positive correlations between teachers’ CKT-E and classroom practice (as measured through an observation protocol), and CKT-E and the quality and demand of assignments and assessments, (Bell et al., 2017; Zisk et al., 2019a). In the cases studies presented here, two of the four teachers studied exhibited strong consistency among their CKT-E and each measures of classroom practice and student achievement and the other two teachers exhibited some discrepancies.

Teacher B scored highly on the CKT-E assessment, and their high CKT-E was reflected in their classroom teaching and the active nature of student learning within their classroom, assignments and assessments, and the performance of their students. Conversely, Teacher D scored low on the CKT-E assessment and this low score was also consistent with each of the measure of practice and student achievement. These results support the claim that CKT should be reflected in all aspects of practice and fit the results of previous correlational studies between CKT and measures of practice and student outcomes (Baumert et al., 2010; Carlisle et al., 2009; Gitomer et al., 2014; Hill et al., 2005; Hill et al., 2008; Keller et al., 2017; Kersting et al., 2012; Park et al., 2011; Phelps et al., 2012).

Along with two teachers showing consistency across CKT-E assessment scores, classroom practice and student achievement, two other teachers did not show consistency across all measures. In the case of Teacher C, their CKT-E assessment score was high, their classroom practice strong and their students showed good understanding of energy
as measured by the post-test. However, their written classroom assignments, and their written assessments in particular, were of low quality and demand. In this case, we can explain the discrepancy as the teacher made a personal choice to not place an emphasis on paper and pencil assessments and instead assess students through oral discussions.

Teacher A also exhibited a discrepancy between their CKT-E, teaching practice and student achievement. Their CKT-E assessment score was low, as were their scores based on classroom observations. Their observations scored particularly low on the dimensions related to explanations of the content and questioning. However, their pre-instruction interview indicated that Teacher A had solid knowledge of the important concepts within the domain of energy. In addition, the videos of Teacher A’s lessons showed the students actively engaged in cooperative learning during every lesson. The teacher made the students aware of the purpose of each day’s lesson. Classroom artifacts exhibited high quality and demand and were based on researched, high-quality curriculum materials. Finally, their students performed well on the energy post assessment, indicating that the students had robust understanding of energy concepts.

In the case of Teacher A, based on their CKT-E assessment scores, we would not expect that their assignments and assessments would be of high quality or that their students would perform well on the energy assessment. However, in analyzing their pre-instruction interview we found Teacher A pointing frequently to their teacher preparation program, the knowledge they gained in that program, and the curriculum that they learned to use during their training. In this case, Teacher A’s strong knowledge of the curriculum and corresponding teaching methods, such as providing students with the purpose of the lesson and having them work cooperatively, allowed them to select the
appropriate instructional tasks and created a supportive classroom environment, which could have helped their students develop a strong knowledge of energy concepts.

The results of previous CKT studies and the cases of Teacher B and Teacher D offer support for the assertion that a teacher’s CKT has a strong influence on all aspects of their teaching practice. However, in some cases, teachers perform differently than expected based on just the assessment of their knowledge for teaching. In these cases, the teachers use their CKT-D, CKT-P and curricular knowledge to carry out effective instruction to help students learn.

One discrepancy we did not see in the intensive study was a case in which a teacher exhibited strong CKT-E, but had low practice scores (Teacher C and one other teacher in the intensive study not highlighted here had low assignment and assessment scores, but in the case of Teacher C, they chose not to use rigorous paper and pencil assessments, and in the second case, the teacher was required to use certain assignments and assessments by their school) and low student achievement. If a teacher has a high CKT and a strong motivation to teach (as would the teachers who volunteered to sign up for an intensive study of their practice), they use their knowledge to carry out effective teaching practice. While that practice may not look the same in every case, teachers with high CKT can use their knowledge to ensure that their students learn..

After examining the coherence among CKT, measures of practice, and student achievement, we set out to attempt to explain the coherence and any discrepancies between teachers’ performance on the measures through the construct of CKT. To do so, we analyzed each case through the constructs of CKT-D (disciplinary CKT) and CKT-P (pedagogical CKT). CKT-D concerns a teacher’s knowledge for the important concepts
and ideas within a content area, but does not require a knowledge of student learning. CKT-P builds off of CKT-D by requiring not only disciplinary knowledge, but also knowledge of student learning. For example, CKT-P allows teachers to choose the correct instructional sequences and to attend to student ideas and difficulties.

In cases when a teacher exhibits consistent performance across all measures, the explanation is rather straightforward. Teacher B, for example, exhibited both high CKT-D and CKT-P on the CKT-E assessment, and through their pre-instruction interview which indicated they had a strong knowledge of the important energy targets, knowledge of student ideas, and knowledge of how to help students develop knowledge of the energy targets. This knowledge was again employed in the classroom as evidence through their classroom observations and the analysis of their assignments and assessments. In the classroom, students were working together to develop knowledge of the important energy targets, and Teacher B acted as the facilitator only stepping in to ask questions when needed and spending little time lecturing to the class. Their high CKT allowed them to carry out effective teaching practice, which led to high student achievement.

Conversely, teachers such as Teacher D who lack both CKT-D and CKT-P are unable to carry out effective teaching practice in order to help their students learn. Teacher D, for example, lacked CKT-D and CKT-P as shown by their CKT-E assessment and interview, but did have a knowledge of the curriculum that they used for instruction, though the knowledge was limited to how topics were sequenced throughout the unit. Their lack of CKT resulted in lecture-based lessons that relied solely on the slides provided by the curriculum and did not provide students with the opportunity to work
through the concepts on their own. A teacher’s low CKT results in their inability to carry out tasks of teaching effectively, which in turn results in low student achievement.

Inconsistencies between a CKT, teaching practice, and student achievement may also be explained through the lens of teachers’ CKT-D and CKT-P (Etkina et al., 2018). We would not expect a teacher with high CKT to perform poorly in the classroom and, in turn, to have students who achieve on a low level. While in the intensive study we did not see any teachers that fit such a profile, there were instances in which teachers with high CKT performed poorly on individual measures of practice, such as Teacher C, whose assignments and assessments were of low quality and demand. In Teacher C’s case, it was their personal belief that written assignments and assessments were not necessary, and that they could better assess their students through discussion. This teacher highlights the importance of understanding teacher beliefs when evaluating teacher knowledge (Charalambos, 2015). Their belief that paper and pencil assessments were not the most effective way of assessing students led them to develop a system of assessment through discussion. In cases such as this, teachers can use their high CKT-D and CKT-P to carry out effective instruction, even when it might not necessarily be seen as conventional high-quality instruction.

A teacher’s CKT-D and CKT-P can also be used to explain cases in which teachers with low CKT perform well on some aspects of classroom practice and have high student achievement. In cases similar to Teacher A, strong knowledge of the discipline (CKT-D) or the important student learning targets in the content was not sufficient to enact high-quality instruction. Such teachers lack the ability to craft instruction to help students develop knowledge of the learning targets, particularly when
faced with a student question or idea during a class. However, a strong knowledge of a specific curriculum and the teaching methods that are central to that curriculum can lead to a teacher with lower CKT having students who achieve at high levels. For example, Teacher A, made the students aware of the goal and purpose of each lesson and provided them with the opportunity to work through high quality activities from the curriculum they were trained on in cooperative groups. This allowed them to work together to create their own knowledge. This strong knowledge of the curriculum that Teacher A learned to use allowed them to create high quality learning experiences for their students but did not allow them to address novel teaching situations, such as those on the CKT-E assessment. Similarly, we would not expect to see explorations of content beyond the scope of the curriculum in their class. Teacher A’s CKT-D and CKT-P related to the curriculum they learned in the teacher preparation program set them apart from teachers such as Teacher D, who had some knowledge of the curriculum they were teaching, but lacked CKT-D, which did not allow them to carry out effective teaching practices. Both teachers relied heavily on their training in the classroom as is typical among teachers (Friedrichsen, 2009; Grossman, 1991), but only Teacher A had sufficient CKT-E to implement the curriculum effectively.

**Implications**

The results of this study offer implications that span across research into teacher knowledge, teacher preparation, and the evaluation of teaching practice. With regard to research in teacher knowledge, this study supports the assertion that, as defined, the construct of Content Knowledge for Teaching strongly influences teaching practice, and in turn, student achievement. Those with robust CKT can effectively carry out teaching
practice and support their students in developing knowledge of the important learning targets within a content area, while those who lack CKT cannot. Therefore, more effort needs to be placed into developing assessments of Content Knowledge for Teaching for individual content areas within a domain, such as forces and motion, D.C. circuits, optics, and other content areas within Physics.

The results here also point to the idea that those with a strong knowledge for teaching a particular curriculum, even when not having a more general CKT for teaching a particular content area, can also teach effectively, if they are aware of the important learning targets and how to support those targets using the curriculum that they are familiar with. Teachers with high curricular knowledge may not perform well on CKT assessments, but still have the knowledge to help their students learn. As such, when considering teachers’ knowledge for teaching, teacher educators need to pay attention to not only their content knowledge for teaching particular content, but also their knowledge for teaching using a particular curriculum, as high levels of CKT and high levels of curricular knowledge can both possibly lead to quality teaching practice and high student achievement.

That a strong knowledge of a high-quality curriculum can lead to effective instruction and student achievement also has implications for teacher education. Teacher preparation programs are often short in duration (less than 2 years) and tend to focus on a survey of many teaching methods and resources, as opposed to focusing on one curriculum. However, as shown in this study, a strong knowledge of a curriculum can lead to effective teaching. As such, teacher preparation programs might consider devoting significant time to allowing pre-service teachers to learn how to implement a specific
curriculum in detail and how such curriculum is used to help students learn day by day in the classroom.

The work here also has implications for the evaluation of teaching practice. The evaluation of teaching practice often focuses on one aspect of teaching practice, often classroom observations, but as illustrated in this study, to get an accurate picture of teaching practice, multiple measures should be used. Teacher A did not score highly on the classroom observation measure, but scored highly on the assignment and assessment protocol, and their students performed well. Similarly, if Teacher C was judged by their assessments, the rigorous nature of their instruction and how they truly assesses their students would be missed. True pictures of teaching quality must take into account how a teacher performs across all essential tasks of teaching in order to make judgements about teaching quality.

**Limitations**

The limitations of this qualitative study arise mainly from the sample size. As we conducted case studies, the sample had to be limited to the teachers who were used to illustrate typical patterns found in the intensive study data. Further work into investigating similar patterns of coherence and discrepancies need to be conducted.

The sample was also limited in the scope of the teachers who participated. Though the teachers in the CKT-E intensive study ranged in their levels of CKT-E, as a whole their performance on the CKT-E assessment was higher than the mean level of performance among all of those who took the assessment. The mean CKT-E estimate of those in the intensive study was approximately .5 standard deviations higher than the mean level of performance of all of those who took the test.
Additionally, as the qualitative study was limited to four teachers, the number of curricula used by the teachers in the study was limited. Of the four teachers, only 3 different curricula were used, limiting the scope of different instructional techniques and philosophies used.

Finally, there were limitations with regard to data collection. In the study, we only asked teachers to submit their final unit assessment and three other assignments or assessments. This provided us with only a small sample of the assignments and assessments teachers used during instruction. Additionally, the method used for recording lessons provided us with the ability to clearly capture audio of the teacher speaking and audio when the teacher was working with students one on one or in small groups. However, discussion between students when they were working in small groups without the teacher was not captured. This did not allow us to capture the depth of their explanations and understanding when the teacher was not working directly with them.
References


Chapter 5: Summary

The papers in this dissertation investigate the link between teachers’ content knowledge for teaching energy, the quality of their instruction, and the achievement of their students. In each of the papers, we investigated the quality of instruction through the quality and demand of the assignments and assessments teachers use during instruction.

Below I provide a short summary of the finding of each paper.

The first paper in this dissertation described the creating of a protocol intended to measure the quality and intellectual demand of the assignments and assessments teachers use during instruction. In this study, we determined that the quality and demand of assignments and assessments can be measured reliably and the artifacts used by the teachers in the study represented a range of quality and intellectual demand. When investigating the relationship between teachers’ CKT-E and the quality and demand of their artifacts, we found positive, significant relationships between teachers’ CKT-E and the depth of content addressed on their assignments and assessments, the overall demand of their energy unit assessments, and the level of analysis required on the assignments that they provided to their students during instruction. This indicates that teachers with more robust knowledge for teaching energy tend to design and select higher quality and more demanding assignments and assessments than those with less robust CKT-E.

The second paper in this dissertation further investigated the relationship between teachers’ CKT-E and the quality of instruction by analyzing the content of teachers’ goals for instruction, the content that they assessed on their unit assessments, and the coherence between their unit goals and the content of their assessments. We found that all teachers tended to state the same breadth of content in their unit goals, but teachers with higher CKT-E tended to state more in-depth goals, and those with lower CKT-E mentioned
more superficial goals. Similarly, teachers with higher CKT-E tended to address more goals on their unit assessment and do so at a more in-depth level than those with lower CKT-E. Finally, when analyzing the coherence between teachers’ goals and the content of their assessments, we found that teachers with lower CKT-E tended to assess more content on their unit assessments than indicated by their unit goals. This study provides further evidence that teachers with more robust content knowledge for teaching provide their students with higher quality instruction, as they help students develop knowledge of more in-depth content.

In the final paper in this dissertation, I used four case studies to describe how a teacher’s content knowledge for teaching energy relates to teaching practice, and how CKT-E and teaching practice mediates student achievement. I studied four teachers, two with high CKT-E and two with low CKT-E based on the CKT-E assessment. Teachers with high CKT-E carried out high quality instruction, even when that instruction did not look typical. For example, one teacher with high CKT-E used written assessments that were of low quality and demand, and instead assessed his students through small group discussions. During these discussions the teacher was able to assess students’ knowledge and ask follow-up questions in the moment. For the teachers with low CKT-E, we found that while one teacher was unable to carry out high quality instruction, the second teacher overcame their low CKT-E and was able to provide their students with high quality instruction due to their strong knowledge of a high-quality curriculum and the teaching methods used in it.

Taken as a whole, these studies each advance the theoretical construct of content knowledge for teaching by providing evidence that content knowledge for teaching is
related to teaching quality as measured by the quality and demand of teachers’ assignments and assessments, the quality of classroom interactions, and the amount of student learning. The results of these studies support the assertion that CKT enables teachers to carry out key tasks of teaching with the goal of helping students learn.