Assessing Learning Progression of Energy Concepts Across Middle School Grades: The Knowledge Integration Perspective

HEE-SUN LEE
Department of Education, Tufts University, Medford, MA 02155, USA

OU LYDIA LIU
Educational Testing Service, Princeton, NJ 08541, USA

ABSTRACT: We use a construct-based assessment approach to measure learning progression of energy concepts across physical, life, and earth science contexts in middle school grades. We model the knowledge integration construct in six levels in terms of the numbers of ideas and links used in student-generated explanations. For this study, we selected 10 items addressing energy source, transformation, and conservation from published standardized tests and administered them to a status quo sample of 2688 middle school students taught by 29 teachers in 12 schools across 5 states. Results based on a Rasch partial credit model analysis indicate that conservation items are associated with the highest knowledge integration levels, followed by transformation and source items. Comparisons across three middle school grades and across physical, life, and earth science contexts reveal that the mean knowledge integration level of eighth-grade students is significantly higher than that of sixth- or seventh-grade students, and that the mean knowledge integration level of students who took a physical science course is significantly higher than that of students who took a life or earth science course. We discuss implications for research on learning progressions. © 2009 Wiley Periodicals, Inc. *Sci Ed* **94:**665–688, 2010

INTRODUCTION

The importance of developing continuous and coherent understanding of science throughout the secondary school years has been well recognized (American Association for the Advancement of Science [AAAS], 1999, 2007). Big ideas such as energy, evolution, and matter are central in the organization and principles of science (AAAS, 1993; National Research Council [NRC], 1996) and are considered unifying because they provide "connections between and among traditional scientific disciplines and are fundamental and comprehensive" (NRC, 1996, p. 115). The longitudinal development of big ideas has been promoted through learning progressions defined as "descriptions of the successively more sophisticated ways of thinking about a topic that can follow one another as children learn about and investigate a topic over a broad span of time" (NRC, 2007, p. 219).

Because student learning of big ideas occurs across science contexts over a long period of time (NRC, 2007), the science education community is in need of assessments that can validly and reliably measure the conceptual development of big ideas with increased sophistication (Pellegrino, Chudowsky, & Glaser, 2001). In this study, we use an assessment approach intended for measuring a construct called knowledge integration. To lay out differences in sophistication associated with the development of energy understanding, we apply knowledge integration theory (Linn & Eylon, 2006). We define the knowledge integration construct as the knowledge and ability to generate and connect scientifically normative ideas in explaining a scientific phenomenon or justifying a claim made for a scientific problem (Liu, Lee, Hofstetter, & Linn, 2008). We characterize the development of student understanding as the progress made in the direction of eliciting more and more scientifically relevant and normative ideas as well as making more and more scientifically elaborated connections among the normative and relevant ideas elicited.

In this paper, we address two research questions: (1) what levels of knowledge integration middle school students demonstrate on energy source, transformation, and conservation items and (2) how their knowledge integration levels differ by science course as well as by grade level. For this study, we assessed knowledge integration levels of status quo 2,688 middle school students taught by 29 teachers in 12 schools. We designed three item blocks for physical, life, and earth sciences and included 10 items addressing energy source, transformation, and conservation concepts. We selected the 10 energy items from the released Trends in International Math and Science Study (TIMSS) and National Assessment of Educational Progress (NAEP) item sets and modified them into a two-tier format: multiple-choice followed by open-ended explanation. We applied an item response theory analysis based on the Rasch partial credit model to validate a learning progression of energy concepts on the knowledge integration construct.

In the literature review, we characterize learning progression for energy concepts based on three perspectives representing science, standards, and students' conceptual resources. We then describe how the development of energy understanding across various science contexts can be captured on the knowledge integration construct. In the methods section, we describe subjects, item selection and design, item block construction, and data collection and analysis procedures. We describe findings and discuss implications for research on learning progressions.

ENERGY UNDERSTANDING

To be effective, assessments for learning progressions should recognize various conceptual resources students bring in to the science classroom as starting points, identify end points considering "societal expectations (values)" (NRC, 2007, p. 110), and characterize

reasonable intermediate learning performances. The purpose of the following review is to justify the selection and sequencing of energy concepts used in this study.

Scientific Perspective

Throughout history, the advancement of physical science has been made through scientists' search for "constancies or regularities in a world of change, in an attempt to find order in a world of apparent chaos" (Cohen, 1974, p. xiii). Through this search, four conservation principles have been developed: the conservation of momentum in the seventeenth century, charge in the mid-eighteenth century, matter in the late eighteenth century, and energy in the mid-nineteenth century (Cohen, 1974). Unlike the other conservation principles linked to concrete observations and experiences of scientists, energy conservation was developed as a purely theoretical and philosophical thought (Lindsay, 1975). Lijnse (1990) argued that "energy is part of a coherent, internally consistent theory that is constructed discontinuously from life-world experiences" (p. 571). Feynman summarized:

> We can calculate a number for each different kind of energy. . . .When we add all the numbers together, from all the different forms of energy, it always gives the same total. . . there is a number such that whenever you calculate it, it does not change. (Lijnse, 1990, p. 573)

Moreover, mathematical formulization is needed to describe energy conservation in its essence (Sexl, 1981).

Energy as a conserved quantity is defined at the system level and is built upon many supporting concepts such as source, transfer, flow, transformation, work, force, dissipation, and entropy (Mclldowie, 1995). Within a system, energy sources, carriers, and receivers can be identified, and energy flow from one part of the system to another can be observed through a series of changes (Schmid, 1982). The description of energy flow is frequently used in many biological and technological applications (Ametller & Pinto, 2002; Lin & Hu, 2003). For example, energy flows from sources such as the sun and moves through a number of carriers to eventual receivers, such as through producers to consumers in food chains. Changes that occur in parts of a system can be described as energy transformation processes. In thermodynamics, all changes are described in terms of energy transfer (Kaper & Goedhart, 2002). While energy transformation processes are inferred, the amount of energy transferred from one system to another can be measured through work (Elise, 1988). As the internal energy of a system cannot be completely changed into work, the concept of entropy (i.e., wasted heat) is needed—i.e., "when multiplied by temperature, a measure of the amount of energy no longer capable of conversion into useful work" (Chaisson, 2001, p. 17).

Standards Analysis

From the scientific perspective, a conceptual end point for the development of energy understanding is to apply the energy conservation concept to various science contexts with mathematical descriptions. However, decades of research on energy understanding has not agreed on how this level of understanding should be achieved. In science textbooks, energy is commonly defined as capacity to do work and work is defined as the product of displacement and force. This definition appears when kinematics is introduced in upper secondary school grades. Warren (1983) insisted that energy should not be taught until work and force concepts are mastered. Nonetheless, the word energy is often and liberally used even for elementary school science. The National Science Education Standards (NSES;

NRC, 1996) acknowledged that elementary students have "intuitive notions of energy—for example, energy is needed to get things done; humans get energy from food" (p. 126).

Energy transfer and transformation are often used as intermediary concepts before conservation is introduced. In the physical science part of the NSES, energy transfer is introduced in grades 5–8 in connection with various energy sources: "Energy is a property of many substances and is associated with heat, light, electricity, mechanical motion, sound, nuclei, and the nature of a chemical; Energy is transferred in many ways" (NRC, 1996, p. 155). In the 9–12 grade span, conservation of energy is introduced with an increase in disorder (entropy) along with interactions of energy with matter, living systems, and the earth system.

Although energy transfer is thought to provide a better scientific account needed in energy conservation (Chisholm, 1992; Kaper & Goedhart, 2002), energy transformation is also adopted for secondary school science (Becu-Robinault & Tiberghien, 1998; Papadouris, Constantinou, & Kyratsi, 2008). Energy transformation is "a precondition for the concept of conservation of energy was to make precise the various forms and manifestations of energy, to analyze their interconvertibility, and to establish quantitative measure of energy" (Cohen, 1974, p. xiii). The *Atlas of Scientific Literacy* (AAAS, 2007) included a map for the development of the energy transformation concept across K-12 grades. Large-scale standardized tests like the NAEP (National Assessment Governing Board, 2004) classified energy form, source, and transformation as the main science content strands. Liu and McKeough (2005) identified 27 TIMSS items that targeted various energy concepts such as work, source, form, degradation, transfer and conservation and claimed that item difficulty is associated with students' cognitive development.

Students' Conceptual Resources

Because energy is already part of students' everyday language and experience (Lijnse, 1990; Trumper, 1990), the development of energy understanding in the direction of energy conservation is challenging (Driver & Warrington, 1985). In diagnosing student understanding of energy, researchers asked students to rate a set of energy statements using various Likert scales (Barak, Gorodetsky, & Chipman, 1997; Kruger, Palacio, & Summers, 1992), to associate "energy" with other familiar words (Goldring & Osborne, 1994), to define the meaning of energy in their own terms (Solomon, 1985), to give examples of energy (Duit, 1984; Trumper, 1990), to select and describe pictures that show energy (Bliss & Ogborn, 1985; Trumper, 1991; Watts, 1983), or to draw a concept map (Liu, Ebenezerr, & Fraser, 2002). Findings indicate that when asked to generate their own ideas, students often consider energy as human-related, depository, activity-related, or as an ingredient, product, function, or fluid-like substance (Watts, 1983).

Other diagnostic studies posed scientific situations that require the application of energy concepts: gravitational free fall, projectile motion, and an object placed on a "U"- or a "∩"-shaped rail (Duit, 1984); various devices that convert one form of energy to another form such as windmills and electronic fans (Driver & Warrington, 1985; Duit, 1984); energy transfer diagrams (Ametller & Pinto, 2002; Styliamidou, Ormerod, & Ogborn, 2002); energy flow in the food chain, photosynthesis, and respiration (Lin & Hu, 2003); and chemical reactions (Papadouris et al., 2008). Results of these studies indicate that students' understanding of energy conservation is not transferable. Although appropriate and relevant, the energy conservation principle is not spontaneously used by students even after instruction of energy conservation and even when students recall that energy cannot be created or destroyed (Driver & Warrington, 1985).

A lot of research has been conducted in thermodynamic contexts such as conduction and insulation (Clough & Driver, 1985; Duit & Kesidou, 1988) and heat and temperature

differentiation (Erickson, 1979, 1980; Harrison, Grayson, & Treagust, 1999; Lewis & Linn, 1994). These studies indicate that students do not differentiate between heat and temperature concepts, consider heat having a substance-like quantity, associate different sensations with different temperatures, and have difficulty understanding heat equilibrium and latent heat.

In sum, the development of energy understanding involves understanding many aspects of energy such as energy source, transfer, transformation, and conservation. To be scientifically complete and sophisticated, understanding should be based on energy as a conserved quantity. Students' overall understanding can progress toward energy conservation by identifying energy sources in a system and connecting various forms of energy and energy transfer processes to changes occurring in the system. In addition, students should be able to recognize and use energy concepts across mechanical, thermodynamic, biological, chemical, and technological applications. In this study, we use the knowledge integration assessment approach to investigate whether this energy concept sequence can be supported when student responses to items addressing energy source, transformation, and conservation concepts are analyzed across physical, biological, and earth science contexts.

THE KNOWLEDGE INTEGRATION ASSESSMENT APPROACH

In the measurement community, an underlying ability that leads to consistent observable responses across a set of related items or tasks is called a "construct." Constructs are latent because they cannot be observed directly and thus must be inferred from responses. Constructs establish "direct probes and modeling of the processes underlying test responses, an [assessment] approach becoming both more accessible and more powerful with continuing developments in cognitive psychology" (Messick, 1989, p. 17). Establishing construct validity is iterative, and "studies of performance differences over time, across groups and settings, and in response to experimental treatments and manipulations" (Messick, 1989, p. 17) are needed to further refine construct definition, item and test design, scoring rubrics, and score interpretation (Wilson, 2005).

Wilson (2005) described construct modeling as setting up a continuum on which respondents and their item responses can be ordered and placed. This construct-modeling approach (Wilson, 2009) has been gaining popularity in learning progression research that requires descriptions of student performances at various levels. It has been used in determining item designs (Briggs, Alonzo, Schwab, & Wilson, 2006), grouping students at various levels of proficiency on a proposed construct (Steedle & Shavelson, 2009), and following students' increasing competency throughout an intervention period (Songer, Kelcey, & Gotwals, 2009). Descriptions of potential learning progressions can be found in several content areas such as matter (Smith, Wiser, Anderson, & Krajcik, 2006), modern genetics (Duncan, Rogat, & Yarden, 2009), and biological evolution (Catley, Lehrer, & Reiser, 2005), all of which require empirical verifications of the hypothesized constructs.

In this study, we define the knowledge integration construct as students' knowledge and ability to elicit and connect scientifically normative and relevant ideas in explaining a scientific phenomenon or justifying their claim in a scientific problem. In our previous research, we established the knowledge integration construct as a unidimensional variable based on a Rasch partial credit model analysis of 201 multiple-choice and explanation items (Liu et al., 2008). Establishing unidimensionality of a construct provides a psychometric foundation in comparing student performances estimated from different test versions, interpreting students' test scores as estimates of their ability on the construct, and examining individual item difficulties according to the construct. This presents unprecedented advantages to researchers for updating test content, reducing testing time, examining individual

TABLE 1
Knowledge Integration Construct Levels

Knowledge Integration Level	Score	Student Characteristics
Complex-link	5	Students elicit and connect three or more normative and relevant ideas in a given science context.
Full-link	4	Students elicit and connect two normative and relevant ideas in a given science context.
Partial-link	3	Students elicit normative and relevant ideas in a given science context.
No-link	2	Students elicit nonnormative ideas or make invalid connections between nonnormative ideas or between normative and nonnormative ideas in a given science context.
Irrelevant	1	Students do not elicit ideas relevant to a given science context.
No information	0	Students do not provide any response.

items, and obtaining more accurate estimates of student knowledge and ability than using total scores (Bond & Fox, 2007).

According to knowledge integration theory, students can develop science understanding by eliciting prior ideas, adding new normative ideas, and comparing and contrasting the new and the old ideas. Through these knowledge integration processes, students can make links among many relevant ideas from both normative science and everyday experiences, leading to more integrated understanding (Linn, 2006). In representing various stages of knowledge integration, we define the knowledge integration construct in six progressively more sophisticated levels of reasoning. (See Table 1.) The first and second knowledge integration levels represent "blank" and "irrelevant" responses, respectively. At the "no-link" level, students elicit nonnormative ideas or make links using nonnormative ideas. "Partial-link" responses use relevant and normative ideas that are not connected. "Full-link" responses include one scientifically valid and fully elaborated link between two scientifically normative and relevant ideas. "Complex-link" responses include two or more scientifically valid and elaborated links among three or more normative and relevant ideas. In this rubric, all alternative ideas students have are assigned to the no-link level. Unlike other learning progression models where the levels are determined by students' making fewer and fewer common conceptual errors (Alonzo & Steedle, 2009; Steedle & Shavelson, 2009), the knowledge integration construct recognizes students' elicitation of normative ideas and their more and more sophisticated use of the elicited ideas by establishing links in explanations.

Based on this overall construct description, a rubric is developed for each scientific problem to score student-generated explanations. Since students' responses to items are rewarded for the knowledge integration levels they exhibit, the total scores from a test can be interpreted as representing their knowledge integration level. Moreover, the knowledge integration construct provides consistency and comparability among knowledge integration scores assigned to items featuring different scientific contexts. In the Methods section, we provide a few knowledge integration scoring examples.

The knowledge integration view of conceptual sophistication resonates well with the views expressed in *Benchmarks for Scientific Literacy* as "two or more concepts at one

level may converge at the next level to form a more complex idea; or a concept may, at the next level, connect to two or more others" (AAAS, 1993, p. 315) as well as by Baxter and Glaser (1998), saying

> key among these differences [between people who have learned to be competent in solving problems and performing complex tasks and beginners who are less proficient] is *integrated knowledge*, knowledge that allows students to think and make inferences with what they know, and *usable knowledge*, knowledge that is utilized in appropriate situations. (p. 38)

Because we measure the knowledge integration construct with student-generated explanations, it is quite possible that students can be measured having lower levels of knowledge integration than they know about a given scientific phenomenon if they do not explicitly elicit ideas and elaborate the connections among the ideas at the time of testing. This is intended because students' knowledge and ability to *both* elicit and connect ideas is measured on the knowledge integration construct, not simply what students know.

In this study, we hypothesize that, overall, items that require elicitation of single ideas and thus lower levels of knowledge integration are easier to solve than those that require connections among multiple ideas. We list knowledge integration requirements for energy source, transformation, and conservation items as follows:

- Energy source items ask students to identify the source of energy. For example, the source of energy for the earth's water cycle is the sun's radiation. The source of energy for the person pushing a bicycle is the food the person has eaten.
- Energy transformation items require students to recognize that one form of energy converts to another form, causing a desired or unexpected change. Consideration of a system is not required to recognize energy transformation processes, but students need to have ideas associated with particular forms of energy and make a link between one form of energy and another.
- Energy conservation items require students to have all the energy transformation ideas that occur within a closed system and connect the ideas to predict changes based on the energy conservation principle.

Any science context featured in an energy item can be described with energy source, transformation, and conservation concepts, even though the item asks only for a particular energy concept. For instance, a multiple-choice item asks the source of energy for the water cycle. When asked to elaborate why students choose a particular answer, they can bring in many other ideas related to the water cycle and energy concepts other than source. This distinction between multiple-choice and explanation items becomes important when we discuss how these two item types contribute to the measurement of students' overall knowledge integration levels in the Results section.

According to the knowledge integration construct definition, students with higher knowledge integration levels are expected to (1) solve energy items across science contexts by selecting more correct multiple-choice answers, (2) write more integrated explanations in each item according to the knowledge integration scoring rubric, and (3) write more integrated explanations across items. In this knowledge integration assessment approach, explanations are weighted five times more (scored 0 to 5) than multiple-choice answers (scored 0 to 1).

METHODS

Participants

We tested a total of 2688 students taught by 29 science teachers in 12 middle school schools across 5 states in the United States at the end of one school year. Among these students, 73.3% were from California, 11.2% Virginia, 8.5% Arizona, 5.4% Massachusetts, and 1.5% North Carolina. These middle schools represented public school districts with varying degrees of language, socioeconomic status, and academic standing. According to the state-mandated test results, three schools were considered high performing, four medium, and five low compared to the middle schools in their states. Among these students, 49.3% were male, 48.8% female, and 1.9% did not provide gender information; 14.0% were in the sixth grade, 47.6% in the seventh grade, and 38.4% in the eighth grade. We classify these students as a status quo sample because they did not receive any particular intervention designed to improve their understanding of energy concepts. The access to the students was obtained through a large-scale curriculum efficacy trial that offered 6 one-week long inquiry-based science curriculum modules (Lee, Linn, Varma, & Liu, in press). During the school year, these teachers implemented one or two of the modules and administered tests at the end of the year. The energy items were included in these tests. The modules the teachers implemented did not address energy but offered an average of 20 explanation-writing opportunities. However, the knowledge integration scoring criteria used in this study were not taught to the students.

Item Design and Item Block Design

Table 2 lists items tested in this study with references. Seven multiple-choice (MC) items and two explanation items were selected from the item sets released by TIMSS in 1995, 1999, and 2003. One MC item called Green was from NAEP in 1990. Among the

TABLE 2
Item Specification

Item	Item Origin (Target Grade)	Item Reference	Original Format[a]	Item Format Modification	Energy	Science Area
Wcycle	TIMSS 95 (8)	I17[b]	MC	Added explanation	Source	Earth
Green	NAEP 90 (8)	–	MC	Added explanation	Source	Earth
Global	TIMSS 99 (8)	S022254[c]	MC	Added explanation	Conservation	Earth
Surface	TIMSS 95 (8)	J1[b]	MC	Added explanation	Transformation	Earth
Light	TIMSS 95 (8)	Y1[b]	MC + EXP	No change	Conservation	Physical
Sequence	TIMSS 99 (8)	S012022[c]	MC	Added explanation	Transformation	Physical
Keisha	TIMSS 95 (4)	N7[d]	MC	Added explanation	Source	Life
Corn	TIMSS 99 (8)	S022141[c]	EXP	Added multiple choices	Transformation	Life
Element	TIMSS 03 (8)	S032386[e]	MC	Added explanation	Conservation	Life
Aquarium	TIMSS 95 (8)	X2[b]	EXP	No change	Transformation	Life

Note. [a] MC = Multiple choice, EXP = Explanation.
[b]International Association for the Evaluation of Educational Achievement (IEA; 1995b).
[c]IEA (1999).
[d]IEA (1995a).
[e]IEA (2003).

TABLE 3
Coverage of Science Topics and Energy Concepts Across Five States[a]

Science Topic	California	Virginia	Arizona	Massachusetts	North Carolina
Sun as energy source in water cycle	6[b]	6, E[c]	6	E	6
Greenhouse effect	6	E	–	–	7
Global warming consequences	6	E	–	–	7
Earth landscape change due to plate tectonics and erosion	6	E	7	E	6
Energy use in electrical circuit	8	P	6	P	8
Energy transformation in chemicals and devices	8	P	8	P	8
Digestion and respiration	7	L	6	L	7
Food web interactions	7	L	7	L	6
Chemical element recycling	8	L	–	L	6
Roles of light and plants in maintaining ecosystem	7	6, L	6	L	6
Energy source	6	6, P	6	E, P	6
Energy transformation	6	6, P	6	P	6
Energy conservation	–	P	–	–	6

[a] California, Arizona, and North Carolina provide content standards for each grade level; Massachusetts provides those for each subject area for middle school grades; Virginia provides sixth-grade content standards and subject area specific content standards for seventh and eighth grades.

[b] 6: the topic is specified in sixth the state's content standards for sixth grade; 7 in seventh grade; 8 in eighth grade.

[c] E: the topic is specified in the state's content standards for middle school earth science; L for life science; P for physical science.

10 items chosen, two items, Light and Aquarium, were used as released by TIMSS. We added "Explain your choice" to the other seven MC items. We created an MC part to one explanation item called Corn. We therefore tested 9 two-tier items consisting of MC and explanation parts. (See Table 2.) The pairing of MC and explanation was necessary to measure knowledge integration levels (Sandoval, 2003; Yeh, 2001). As a result, we tested a total of 9 MC items and 10 explanation items. Three item pairs addressed energy source (Keisha, Wcycle, and Green), four energy transformation (Surface, Sequence, Corn, and Aquarium), and three energy conservation (Global, Light, and Element).

As shown in Table 2, the science contexts for the 10 energy items addressed a variety of middle school science topics such as the water cycle, global warming phenomena, the greenhouse effect, plate tectonics and erosion, the food web, element recycling, ecosystems, respiration, electrical circuits, and chemical reactions. Table 3 shows little consistency across states concerning when to introduce these topics. The states of Arizona and Massachusetts did not specifically mention climate change and related science, whereas the

other states addressed them through renewable and nonrenewable resources. All states addressed energy at the sixth-grade level focusing on energy source and transformation or transfer. However, the treatment of energy conservation was very different across states. California, Arizona, and Massachusetts standards did not mention energy conservation, whereas Virginia and North Carolina standards suggested the use of energy conservation in teaching middle school science topics.

We created three test versions consisting of

- physical science energy item block: Keisha, Wcycle, Sequence, and Light,
- life science energy item block: Keisha, Wcycle, Corn, Element, and Aquarium, and
- earth science energy item block: Keisha, Wcycle, Green, Global, and Surface.

The Keisha and Wcycle items were common across energy blocks and were used to equate students' knowledge integration levels estimated from each of the three item blocks. This test design method is called "item block design" and is commonly used in large-scale assessments for testing more items in limited time (Ferraro & Van de Kerckhove, 2006; National Center for Educational Statistics, 2007; Organization for Economic Cooperation and Development, 2007). An additional benefit of using the item block design was that we matched the item content to the science discipline area taught at a particular grade level. Students took one of the three item blocks depending upon which science discipline topics the teacher was teaching during the school year. The earth science energy block was taken by 334 sixth-grade students and 301 seventh-grade students, the life science energy block by 997 seventh-grade students, and the physical science energy block by 41 sixth-grade and 1074 eighth-grade students. Students took about 15–20 minutes to answer energy items in each item block.

Scoring

The MC parts of the energy item pairs were scored dichotomously, "1" for correct and "0" for incorrect answers. Student explanations were scored from 0 to 5 according to the knowledge integration rubric. We initially developed a knowledge integration scoring rubric for each explanation item based on the analysis of what ideas and links were needed to solve the item. Then, we applied the initial knowledge integration rubric to score 100 randomly chosen student explanations. After matching with typical student responses, we further refined the initial knowledge integration scoring rubric. We tested the revised rubric with additional 100 randomly chosen explanations until 0.90 or higher intercoder reliability was reached between the two coders. After the rubric was finalized, the rest of student explanations were coded by a single trained coder for each explanation item. Figures 1, 2, and 3 show three items related to Light, Keisha, and Wcycle with knowledge integration rubrics, respectively. Each knowledge integration rubric included knowledge integration levels, scores, criteria, and student explanation examples.

Rasch Analysis

Rasch Partial Credit Model. For the dichotomously scored MC items, we used the Rasch model (Rasch, 1960/1980). For polytomously scored explanation items, we used the Rasch partial credit model (Wright & Masters, 1982):

$$P_{nix}(\theta) = \frac{\exp\left[\sum_{j=0}^{x}(\theta_n - \delta_i - \tau_{ij})\right]}{\sum_{r=0}^{m_i}\left[\exp\sum_{j=0}^{r}(\theta_n - \delta_i - \tau_{ij})\right]} \tag{1}$$

Electrical energy is used to power a lamp. Is the amount of light
  energy produced more than, less than, or the same as the
  amount of electrical energy used?

Item context description

The amount of light energy produced is

_____ more than, _____ less than _____ the same as

the amount of electrical energy used.

Multiple-choice part

Explain your choice.

Explanation part

Ideas:
- Energy dissipation idea: Some energy is lost in the circuit or through heat.
- Energy transformation idea: Some energy is transformed to light energy or used to light the lamp.
- Energy conservation idea: The total amount of energy is conserved.

Link:
- Energy dissipation-transformation link: Electrical energy is used to both heat and light the lamp.

Knowledge Integration Level (score)	Criteria	Examples
Irrelevant: Off-task (1)	Wrote some text unrelated to the item.	• Because I think so.
No-link: Nonnormative ideas (2)	Elicited nonnormative ideas or restated the multiple-choice answer chosen.	• Light energy is less needed than electrical energy and because of that the electrical energy is greater. • Because the system would eventually overload if it was too much and it wouldn't work if it was less.
Partial-link: Normative ideas (3)	Elicited one of the energy dissipation, transformation, and conservation ideas	• The light produced does not use all of the electrical energy that goes into the light bulb. Some of the energy is used for the production of the light, but it doesn't actually cause the light to be brighter.
Full-link: Single link between two normative ideas (4)	Made a link between energy dissipation and transformation ideas	• Some of the electrical energy dissipates into heat and is absolutely lost. The remaining energy is converted to heat and used as light energy.
Complex-link: Two or more links among three or more normative ideas (5)	Made a link among energy dissipation, transformation, and conservation ideas.	• A lamp works by charging a filament, which produces both light and heat. Because of conservation of energy, the light energy must be less than the total energy used.

Figure 1. Light item (IEA, 1995b) and knowledge integration scoring rubric.

where $P_{nix}(\theta)$ stands for the probability of student n scoring x on item i. θ stands for the student location on the knowledge integration construct in this study. δ_i refers to the item difficulty. $\tau_{ij}(j = 0, 1, \ldots, m)$ is an additional step parameter associated with each score (j) for item i. For dichotomously scored items, the step parameter (τ_{ij}) is removed from the equation, so the equation becomes the simple Rasch model.

The computer software *ConQuest* was used to perform the Rasch analysis (Wu, Adams, & Wilson, 1997). *ConQuest* provides an overall item difficulty estimate on each item as well as a knowledge integration estimate on each student. The knowledge integration estimates

Keisha is pushing her bicycle up a hill. Where does
Keisha get the energy to push her bicycle?

    (a) From the food she has eaten
    (b) From the exercise she did earlier
    (c) From the ground she is walking on
    (d) From the bicycle she is pushing

Explain your choice.

Ideas:
- Digestion ideas: (1) process idea (Food is broken down), (2) location idea (digestive system including stomach, intestines, etc.), (3) product idea (glucose provides energy)
- Respiration idea: (1) process idea (molecules are converted to produce energy), (2) location idea (muscle cells), and (3) product idea (CO_2, H_2O, and energy)
- Energy source idea: Energy is stored in molecular bonds. Food has chemical energy.

Links:
- Energy transformation link: Chemical energy in food is converted to mechanical energy to push the bicycle.
- Digestion link: Any two ideas from the three digestion ideas are connected.
- Respiration link: Any two ideas from the three digestion ideas are connected.

Knowledge Integration Level (score)	Criteria	Examples
Irrelevant: Off-task (1)	Wrote some text unrelated to the item.	• Because I think so.
No-link: Nonnormative ideas (2)	Elicited nonnormative ideas or restated the multiple-choice answer.	• Food gives you energy so you can do activities.
Partial-link: Normative ideas (3)	Elicited any one of the ideas listed above.	• Sugars from food give you energy. • Food is digested and turns into energy.
Full-link: Single link between two normative ideas (4)	Made a link related to energy transformation, digestion, or respiration listed above.	• Food is broken down to sugars her muscles need.
Complex-link: Two or more links among three or more normative ideas (5)	Made two or more links related to energy transformation, digestion, and respiration listed above.	• She digests the food and it is broken down into sugars which her cells use as energy. • The food that she has eaten has been broken down by her digestive system, and the carbohydrates absorbed into her bloodstream. These are then used by cells to make chemical energy that the cells can use.

Figure 2. Keisha item (IEA, 1995a) and knowledge integration scoring rubric.

usually took values from −4.0 to 4.0. The student knowledge integration estimates and the item difficulty estimates were calibrated to be on the same scale. In Rasch models, the probability of answering correctly to an item depends on both student knowledge integration level and item difficulty. For example, on a dichotomously scored item, if a student's knowledge integration estimate is larger than an item's difficulty estimate, then the student has a larger than .5 probability answering correctly on that item. The relationship between the student's knowledge integration level and the item's difficulty is shown in a

The source of energy for the earth's water cycle is the
      (a) Wind
      (b) Sun's radiation
      (c) Earth's radiation
      (d) Sun's gravity

Explain your choice.

Main Ideas:
- Sun idea: Sun warms (heats) up earth.
- Evaporation idea: Water is evaporated (water changes from liquid to gas or water vapors).
- Condensation idea: Water changes from gas to liquid to form clouds.
- Precipitation idea: Different types of precipitation fall from sky.
- Energy idea: Energy is required in evaporation.
- Water cycle idea: Water cycle consists of evaporation, condensation, and precipitation.

Links:
- Sun–energy link: Sun's radiation energy is used for evaporation of water.
- Evaporation Link: Water molecules are broken apart from the surface of the liquid water.

Knowledge Integration Level	Criteria	Examples
Irrelevant: Off-task	Wrote some text unrelated to the item.	• Because I think so.
No-link: Nonnormative ideas	Elicited nonnormative ideas or restated the multiple-choice answer chosen.	• Sun's radiation gives light to the water cycle. • Wind moves the water through the air.
Partial-link: Normative ideas	Elicited any one idea listed above.	• Sun warms up water. • Sun evaporates water.
Full-link: Single link between two normative ideas	Used either sun–energy link or evaporation link.	• The water molecules are broken off from the surface of the water. • Sun heats up the water to evaporate.
Complex-link: Two or more links among three or more normative ideas	Used either sun–energy or evaporation link and added an additional idea or link listed above.	• The sun's radiation makes the water evaporate in to the air which goes through evaporation, condensation, then precipitation.

Figure 3. Wcycle item (IEA, 1999) and knowledge integration scoring rubric.

diagram called a Wright map (Wilson, 2005). In the Results section, we use this Wright map to show how energy concepts and middle school students were measured on the knowledge integration construct.

Item Block Equating.   To compare student performance obtained from the three energy item blocks, we used the nonequivalent-groups anchor test design (Kolen & Brennan, 2004). The two common items (Keisha and Wcycle item pairs) were used to link student performance across the three item blocks. We used the *mean/sigma* method (Marco, 1977) and transformed item difficulty estimates of these common items in the three blocks so that they had the same mean and standard deviation values. Based on the linear equation used in this transformation, student knowledge integration estimates obtained from each item block were recomputed to be on the same knowledge integration scale across three item blocks.

Science Education

RESULTS

To address what levels of knowledge integration students demonstrate across energy items, we first describe how distributions of item difficulty and student knowledge integration levels are shown on a Wright map. Then, we use the map to examine how students' knowledge integration levels compare across energy items and how MC parts compare with explanations on the knowledge integration scale. To address how students' knowledge integration levels differ by science course and by grade level, we compare student performances on the two common item pairs across the three energy item blocks as well as knowledge integration estimates from the Rasch analysis.

Descriptive Item Statistics

The entire energy item set had a Cronbach's alpha of .72. The item separation reliability was .98, suggesting that the responses from this sample of students could effectively differentiate the items according to the knowledge integration construct. Table 4 shows the percentages of students who chose correct answers on the MC parts of the items as well as those of students who received corresponding knowledge integration scores on their explanations. Overall, the percentage of students who chose a correct answer for the MC part decreased as the item requires a greater knowledge integration level with energy concepts, i.e., from energy source to transformation and to conservation. The highest correct percentage was found on the Green item (72%), whereas the lowest correct percentage was on the Light item (31%). Table 4 also shows that students tested in this study scored higher on the five MC items than published international average values.

Across 10 explanation items, 13.7% of all responses were blank, 9.5% irrelevant, 45.7% no-link, 22.5% partial-link, 7.6% full-link, and 1.3% complex-link. This means that 68.9%

TABLE 4
Descriptive Item Statistics

Item	N	Multiple choice (%, Correct)		Explanation (%, Knowledge Integration Levels[a])					
		TIMSS[b]	This Study	Blank	Irrelevant	No	Partial	Full	Complex
(a) Energy source items									
Green	635	–	72	9	19	45	19	9	2
Keisha	2688	–	70	9	4	72	13	1	0.2
Wcycle	2688	41	54	20	11	42	19	6	2
(b) Energy transformation items									
Sequence	1126	59	63	19	9	34	33	4	0.3
Surface	635	41	50	12	12	38	31	7	1
Corn	979	–	49	16	8	60	9	6	1
Aquarium[c]	979	–	–	4	4	31	39	20	3
(c) Energy conservation items									
Element	979	–	48	24	10	40	19	6	1
Global	635	33	43	6	9	46	22	14	2
Light	1126	8	31	18	9	49	21	3	0.2

[a] Knowledge integration levels: Irrelevant = Irrelevant link; No = No-link; Partial = Partial-link, Full = Full-link, Complex = Complex-link.
[b] TIMSS statistics provide eighth-grade international average values.
[c] The Aquarium item had only the explanation part.

of students' responses to explanation items were based on the absence of explanations relevant to the items or nonnormative ideas about energy concepts. Only 22.5% of the explanations included normative ideas and 8.9% were based on full and complex links. This indicates that students chose correct MC answers often without eliciting scientifically normative, relevant, and elaborated ideas or links. In particular, easy MC items, as indicated by higher correct percentages such as the Green item, were not necessarily associated with higher knowledge integration levels than other more difficult MC items. Overall, students' knowledge integration level with energy concepts across science contexts was not sophisticated.

Wright Map Analysis

Figure 4a shows a cumulative probability curve for the MC part of the Wcycle item that asked "The source of energy for the earth's water cycle is the _____" and four choices included wind, sun's radiation, earth's radiation, and sun's gravity. The x-axis represents the knowledge integration scale while the y-axis represents the probability of answering correctly on the MC part of the Wcycle item. The probability of choosing a correct answer, i.e., identifying the sun's radiation energy as a source of the water cycle, increased as their knowledge integration level increased. The knowledge integration estimate value on the x-axis that intersects with the 50% probability on the y-axis is called the item threshold. The item threshold value for the MC part of the Wcycle item pair was −0.52. Students with a knowledge integration estimate of −0.52 had a 50% chance of choosing a correct answer. Students with higher than −0.52 had a more than 50% chance of choosing a correct answer, whereas those with lower than −0.52 had a lower than 50% chance of doing so.
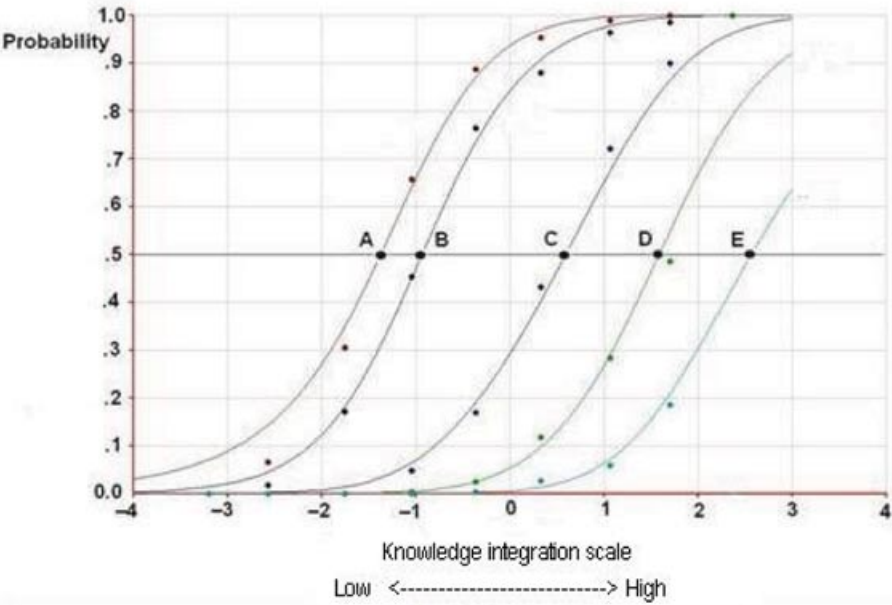
Figure 4b shows cumulative probability curves for the explanation part of the Wcycle item. The x-axis represents the knowledge integration scale, whereas the y-axis represents the probability of obtaining score $j+1$ or higher from j or lower (where $j = 0, 1, 2, 3$, and 4). Since the explanation part was scored from 0 to 5, five cumulative probability curves are represented. The far left cumulative curve indicates the probability of obtaining a score of 1 (irrelevant) or higher from a score of 0 (blank) on the Wcycle explanation item. The far right line indicates the probability of obtaining a score of 5 (complex-link) from a score of 4 (full-link) or lower. As shown in Figure 4b, it was very difficult to obtain a score 5 on the Wcycle item. Even students with very high knowledge integration estimates (i.e., values larger than 2.0 on the x-axis) had a higher probability of reaching the score 4 level than the score 5 level. Points A– E in Figure 4b represent five item thresholds from "A" to "E" that intersect with the .50 probability line.

The Wright map produced from the Rasch partial credit model analysis is shown in Figure 5. On the same knowledge integration scale, item threshold values of all MC and explanation items are shown. On this map, the vertical axis represents the knowledge integration scale. Student knowledge integration estimates and item thresholds share this same vertical axis. Higher positions on the vertical axis represent higher knowledge integration levels. Figure 5a shows that the distribution of students' knowledge integration estimates had an average of −0.29 on the scale of −4.0 to 4.0 with a standard deviation of 0.79. The student knowledge integration estimates ranged from −2.51 to 2.31. Figure 5b shows the distribution of item thresholds: one item threshold for each of nine MC items and five item thresholds for each of the 10 explanation items. The higher the position, the more difficult it was to achieve that score.

This Wright map provides evidence that the knowledge integration construct could be considered unidimensional despite variations in science content. First, the item threshold locations of the five knowledge integration levels were represented in each explanation item
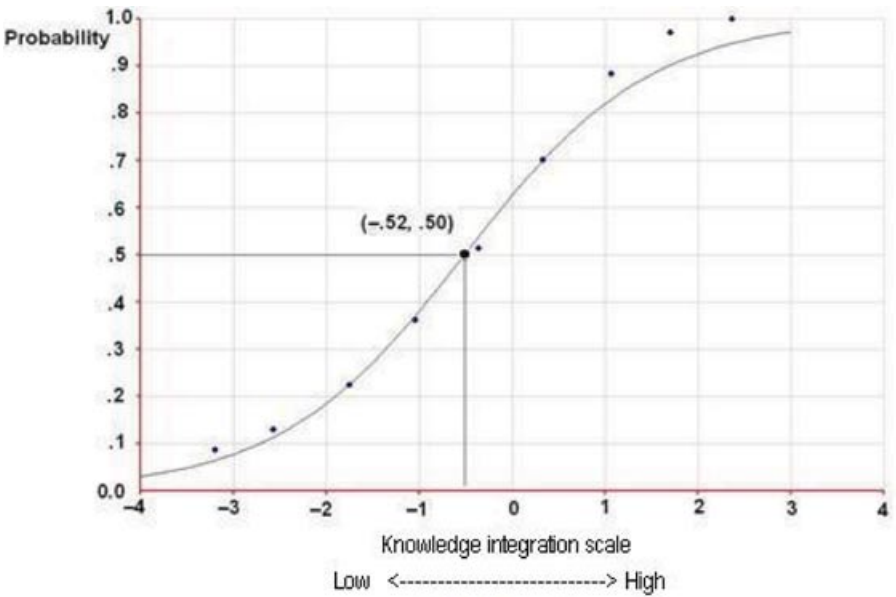
Figure 4. (a) A cumulative probability curve for the MC part of the Wcycle item and (b) five cumulative probability curves for the explanation part of the Wcycle item. The item threshold of the MC part is -0.52. Five item thresholds for the explanation part are shown for irrelevant (A), no-link (B), partial-link (C), full-link (D), and complex-link (E). These item thresholds are displayed on the Wright map shown in Figure 5. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]
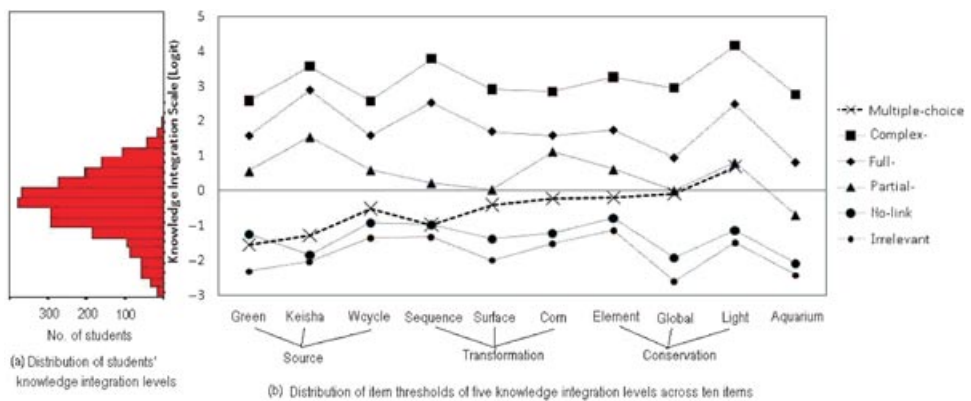
Figure 5. The Wright map shows the student distribution (a) and the item threshold locations (b) according to the common knowledge integration scale. Item thresholds of nine multiple-choice item parts addressing energy source, transformation, and conservation are plotted. Item threshold locations of the same knowledge integration level are connected across 10 explanation items. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

in the same order as hypothesized in Table 1. Second, item thresholds for each of the knowledge integration score (e.g., thresholds for "4") across 10 explanation items were clustered covering a similar range on the knowledge integration scale. The five clusters of item thresholds progress from irrelevant to complex-link. The "irrelevant" cluster ranged from −2.61 to −1.14, the "no-link" cluster from −2.09 to −0.79, the "partial-link" cluster from −0.70 to 1.54, the "full-link" cluster from 0.82 to 2.88, and the "complex-link" cluster from 2.57 to 4.16. The relatively short distances between irrelevant and no-link levels indicate that the blank (score "0") and the irrelevant (score "1") knowledge integration levels could be combined without loss of information on student performance across explanation items.

Figure 5b shows item difficulty locations for the nine MC items. The item difficulty locations of the MC items ranged from −1.55 (Green item addressing energy source) to 0.68 (Light item addressing energy conservation). Overall, energy source MC items were located at the lower part of the knowledge integration scale than the energy conservation MC items. Energy transformation MC items were located in between. This indicates that energy conservation MC items were most difficult for students to choose a correct answer, followed by energy transformation MC items. Energy source MC items were easiest. The same order of item difficulty among source, transformation, and conservation items were found in a study by Liu and McKeough (2005) who associated the item difficulty with students' overall cognitive ability differences. In this study, we relate the item difficulty to the interaction between the knowledge integration levels required by the items and those held by students. The item difficulty locations of the nine MC energy items overlapped with the irrelevant to partial-link levels of explanations. No MC items used in this study therefore could differentiate students at the full- and the complex-link knowledge integration levels.

Students' Knowledge Integration Levels With Energy Concepts

Since the earth, life, and physical science energy blocks were taken by students from mixed-grade levels, we compared student knowledge integration levels across science courses as well as across grade levels. The science course analysis addressed whether knowledge integration levels with energy concepts depended upon science contexts where the energy concepts were embedded. The grade-level analysis addressed whether knowledge integration levels with energy concepts progressed by age. However, the these two analyses were very closely related as the majority of sixth-grade students took the earth science

TABLE 5
Mean Comparison Across Three Science Courses

Performance (Score Description)	Earth Science (N = 635) M(SE)	Life Science (N = 979) M(SE)	Physical Science (N = 1074) M(SE)	F (2, 2685)[a]	Post hoc[b]
Keisha: Multiple choice (correct percentage)	67.1 (1.9)	81.9 (1.2)	61.5 (1.5)	55.0***	P ≪ E ≪ L
Keisha: Explanation (knowledge integration score)	1.83 (0.03)	2.00 (0.03)	1.98 (0.02)	11.7***	E ≪ L, P
Wcycle: Multiple choice (correct percentage)	44.5 (2.0)	58.0 (1.6)	58.6 (1.5)	19.0***	E ≪ L, P
Wcycle: Explanation (knowledge integration score)	1.80 (0.05)	1.69 (0.03)	2.08 (0.04)	27.4***	E, L ≪ P
Knowledge integration (knowledge integration Rasch estimates)	−0.41 (.03)	−0.28 (0.02)	−0.22 (0.02)	11.7***	E ≪ L, P

[a] $*p < .05$, $**p < .01$, $***p < .001$.
[b] P = physical science, L = life science, E = earth science. A statistically significant difference at the .05 level is represented as ≪.

energy block and all of eighth-grade students took the physical science energy block. Half of the seventh-grade students took the earth science energy block, whereas the other half took the life science energy block. Therefore, results on the age progression were intertwined with science courses learned at the particular grade level.

Table 5 lists average scores of student groups by science courses on the MC and explanation parts of the Keisha and Wcycle item pairs. The Keisha item shown in Figure 2 addressed food as the energy source that enables Keisha to push her bicycle up a hill. The Wcycle item shown in Figure 3 addressed the sun's radiation as the source of the water cycle on earth. A one-way ANOVA on all students' responses to these two item pairs was used to identify group differences, followed by Tukey's post hoc tests to examine which two groups were significantly different at the .05 level.

On the MC part of the Keisha item pair, the mean knowledge integration level of students who learned life science during the school year was significantly higher than that of students who learned earth or physical science. The mean of the physical science group was lowest among the three groups despite the fact that the group mainly consisted of the eighth-grade students. On the explanation part of the Keisha item pair, the mean of students who learned life or physical science was higher than that of students who learned earth science. On the MC part of the Wcycle item pair, the mean of students who learned earth science was significantly lower than that of students who learned life or physical science. On the explanation part, students who learned physical science performed significantly higher

TABLE 6
Mean Comparison Across Three Middle School Grade Students

Performance (score description)	Sixth ($N = 375$) M (SE)	Seventh ($N = 1280$) M (SE)	Eighth ($N = 1033$) M (SE)	$F(2, 2685)^a$	Post hoc[b]
Keisha: Multiple choice (correct %)	64.5 (2.5)	79.5 (1.1)	61.0 (1.5)	51.98***	6, 8 ≪ 7
Keisha: Explanation (knowledge integration score)	1.79 (0.04)	1.98 (0.02)	1.98 (0.02)	10.11***	6 ≪ 7, 8
Wcycle: Multiple choice (correct %)	44.5 (2.6)	55.0 (1.4)	59.0 (1.5)	11.65***	6 ≪ 7, 8
Wcycle: Explanation (knowledge integration score)	1.89 (0.07)	1.71 (0.03)	2.07 (0.04)	26.23***	7 ≪ 6 ≪ 8
Knowledge integration (knowledge integration Rasch estimates)	−.38 (0.05)	−.32 (0.02)	−.21 (0.02)	7.97***	6, 7 ≪ 8

[a] $^*p < .05$, $^{**}p < .01$, $^{***}p < .001$.
[b] 6 = Sixth grade, 7 = seventh grade, 8 = eighth grade. A statistically significant difference at the .05 level is represented as ≪.

than those who learned earth or life science. On the knowledge integration scale, physical and life science students performed significantly better than earth science students. These results indicate that the development of knowledge integration with energy concepts was associated with science topics learned during the year. In this study, the most effective was physical science and the least effective was earth science.

Table 6 lists student performance comparison results across three grade levels on the Keisha and Wcycle item pairs. On the Keisha MC item part, the mean of seventh-grade students was highest among the three groups and was significantly different from sixth- and eighth-grade students. On the Keisha explanation item part, the means of seventh-grade and eighth-grade students were significantly higher than that of sixth-grade students. On the Wcycle MC item part, the means of seventh- and eighth-grade students were significantly higher than that of sixth grade. On the explanation part of the item, the mean of eighth-grade students was highest, followed by that of sixth-grade students. The mean of seventh-grade students was lowest. When we used Rasch knowledge integration estimates, the age progression was shown more clearly than using the individual items. Table 6 shows that the mean knowledge integration estimates increased across three middle school grade levels. In particular, the mean of eighth-grade students was significantly higher than those of sixth- and seventh-grade students. This indicates that students' knowledge integration levels did not change much between sixth and seventh grades, despite energy relevance in some science topics and states' content standards introducing energy at the sixth grade (see Table 2).

DISCUSSION

In this study, we used a construct that represents students' knowledge and ability to generate and connect scientifically normative and relevant ideas and established the knowledge

integration scale drawn from the Rasch analysis of student responses to items addressing energy source, transformation, and conservation. Results indicate that students' overall knowledge integration levels with energy concepts are mediocre, that advanced energy concepts such as conservation are more difficult than identifying energy sources, and that the origin of this difficulty is in part related to the increased demand for integrating many scientifically relevant ideas.

These findings indicate that items addressing energy conservation require higher knowledge integration levels than those based on singular ideas such as identifying an energy source or associating energy forms. When students have ideas about energy sources and transformation processes in their conceptual repertoire, they can learn energy conservation in a more integrated manner by recognizing a system with various components, associating changes of the system with energy transformation processes, and analyzing behaviors of the system components that obey the energy conservation principle. This means that the learning progression of energy can be facilitated when students can generate and connect multiple ideas. Therefore, pursuing single correct ideas for each science problem may play a detrimental role in promoting understanding of highly integrated concepts such as energy conservation.

However, these results do not necessarily support an instructional sequence moving from energy source to transformation to conservation. According to our analysis of the current science content standards of five states, we notice two different treatments of energy: using energy conservation at the sixth-grade level as seen in Virginia and North Carolina, and not introducing conservation at all at the middle school level as seen in California, Massachusetts, and Arizona. An open question is whether energy understanding should be developed cumulatively from less integrated to more integrated energy concepts, simultaneously with all needed concepts to explain a system, or as an organizing framework looking for applications across science problems. Research on interventions that manifest these three models is needed to empirically determine when, in what sequence, and for how long various energy concepts can be taught. This study suggests that, to help students develop an understanding of energy, science curricula should address the relevant instructional sequence of energy concepts as well as encourage students to integrate ideas.

We compared differences in knowledge integration levels of students based on a large cross-sectional status quo sample of middle school students using individual MC items, individual explanation items, and student knowledge integration estimates. Comparison results on individual MC and explanation items by grade level do not always yield an increasing trend over the three middle school grades. As each item involves an energy concept and a science topic where the energy concepts are applied, it appears that whether students learned the science topic in the item also contributes to their knowledge integration with the energy concepts. Therefore, understanding of energy concepts should be developed along with that of science contexts where energy concepts become relevant.

In comparing performance differences across middle school grade levels, we define the development of energy understanding as being able to write explanations with an increased number of scientifically normative ideas and elaborated links between the ideas. On the basis of this definition, we applied a construct-modeling approach (Wilson, 2005) to interpret student performance scores in the direction of improved knowledge integration levels. By adopting the Rasch analysis with test-equating methods, we were able to create tests responsive to the needs of assessing science topics learned in a particular grade level and comparing student performance across three grade levels. Results of this study indicate that knowledge integration levels with energy concepts depend upon both science course learned during the year and grade level, though we are not claiming the distinctive contribution of each factor. Physical science appears to contribute the most to students'

knowledge integration, and the eighth-grade students tend to demonstrate higher knowledge integration levels with energy despite all states' emphasis on energy at the sixth-grade level. Interventions are needed to help students develop an integrated understanding of energy when they are learning earth or life science topics in the sixth and the seventh grades.

In longitudinal studies where the same students are followed across multiple years, the knowledge integration construct approach can be useful. We highlight some benefits of using the knowledge integration assessment approach by discussing challenges when individual items are used. First, proper sampling of energy items for repeated use throughout the intended longitudinal development period is challenging. As the eventual goal of energy understanding is for students to develop multifaceted understanding of energy, tests for this purpose can be lengthy. Moreover, it may not be appropriate to ask items that address science topics or concepts students are yet to know. Second, repeated use of the same items is problematic because change in student performance can be also linked to solving the same items over and over again. Third, the role of assessment in providing timely and instructionally sensitive feedback can be compromised when the same items have to be used. Fourth, individual items do not have reliability, and the performance change from the assessment cannot be generalized beyond the individual items.

On the other hand, the knowledge integration assessment approach allows the content and the length of tests reasonably adjusted to the intended instruction by eliminating and adding items. Therefore, assessment results can be used to provide feedback that reflects change in instructional foci. In comparing the ranges of knowledge integration levels targeted by MC items and by explanation items, we note that MC items have difficulty tapping on higher levels of knowledge integration. When tracking student progress toward higher levels of knowledge integration, the usefulness of tests that consist of mainly MC items is limited (Lee, Liu, & Linn, in press).

The generalization of findings in this study is limited. The study sample came from 12 schools that served diverse student populations in terms of language, socioeconomic status, and achievement measured by state-administered standardized tests. A similar study based on a randomized sample from the general middle school population can strengthen the results found in this study. We sampled a fraction of energy items released by TIMSS and NAEP. Therefore, different patterns may emerge if other energy items are used. We encourage researchers to investigate other energy items in assessing and promoting learning progressions for energy understanding. As we did not follow individual students over time, the question of whether students progress in energy understanding over time under the current science education system is not answered in this study. In addition, the use of student-generated explanations to measure the knowledge integration construct might have underestimated what students know and how they reason. The lack of students' epistemological commitment to formulating scientific explanations is well documented elsewhere (e.g., McNeill, Lizotte, Krajcik, & Marx, 2006; Sandoval, 2003). In this study, students' lack of epistemological commitment was also captured because students' own elicitation of normative and relevant ideas to the item was part of the knowledge integration construct.

CONCLUSION

Assessing students' development of understanding across science topics and disciplines through unifying ideas has been challenging. Big ideas tend to be abstract and parsimonious, and their appropriateness and usefulness cannot easily be appreciated by students without conceptual resources or epistemological commitment held by practicing scientists. Owing to developmental and experiential constraints of students, some aspects of big ideas can

be too difficult for students to learn and to be assessed. We applied knowledge integration theory to define a unidimensional construct on which learning progression of unifying concepts such as energy can be measured across science topics and grade levels.

REFERENCES

Alonzo, A. C., & Steedle, J. T. (2009). Developing and assessing a force and motion learning progression. Science Education, 93, 389–421.

American Association for the Advancement of Science. (1993). Benchmarks for science literacy. New York: Oxford University Press.

American Association for the Advancement of Science. (1999). Atlas of science literacy (Vol. 1). Washington, DC: AAAS Project 2061.

American Association for the Advancement of Science. (2007). Atlas of science literacy (Vol. 2). Washington, DC: AAAS Project 2061.

Ametller, T., & Pinto, R. (2002). Students' reading of innovative images of energy at secondary school level. International Journal of Science Education, 24(3), 285–312.

Barak, J., Gorodetsky, M., & Chipman, D. (1997). Understanding of energy in biology and vitalistic conceptions. International Journal of Science Education, 19(1), 21–30.

Baxter, G. P., & Glaser, R. (1998). Investigating the cognitive complexity of science assessments. Educational Measurement: Issues and Practice, 17(3), 37–45.

Becu-Robinault, K., & Tiberghien, A. (1998). Integrating experiments into the teaching of energy. International Journal of Science Education, 20(1), 99–114.

Bliss, J., & Ogborn, J. (1985). Children's choices of uses of energy. International Journal of Science Education, 7(2), 195–203.

Bond, T. G., & Fox, C. M. (2007). Applying the Rasch model: Fundamental measurement in the human sciences (2nd ed.). Mahwah, NJ: Erlbaum.

Briggs, D. C., Alonzo, A. C., Schwab, C., & Wilson, M. (2006). Diagnostic assessment with ordered multiple-choice items. Educational Assessment, 11(1), 33–63.

Catley, K., Lehrer, R., & Reiser, B. (2005). Tracing a prospective learning progression for developing understanding of evolution. The National Academies Committee on Test Design for K-12 Science Achievement.

Chaisson, E. J. (2001). Cosmic evolution: The rise of complexity in nature. Cambridge, MA: Harvard University Press.

Chisholm, D. (1992). Some energetic thoughts. Physics Education, 27, 215–220.

Clough, E. E., & Driver, R. (1985) Secondary students' conceptions of the conduction of heat: bringing together scientific and personal views. Physics Education, 20, 176–182.

Cohen, I. B. (1974). Foreword. In Y. Elkana (Ed.), The discovery of the conservation of energy (pp. xi–xiv). Cambridge, MA: Harvard University Press.

Driver, R., & Warrington, L. (1985). Students' use of the principle of energy conservation in problem situations. Physics Education, 20, 171–176.

Duit, R. (1984). Learning the energy concept in school-empirical results from the Philippines and West Germany. Physics Education, 19, 59–66.

Duit, R., & Kesidou, S. (1988). Students' understanding of basic ideas of the second law of thermodynamics, Research in Science Education, 18, 186–195.

Duncan, R. G., Rogat, A. D., & Yarden, A. (2009). A learning progression for deepening students' understandings of modern genetics across the 5th–10th grades. Journal of Research in Science Teaching, 46(6), 655–674.

Elise, M. (1988). Transferring not transforming energy. School Science Review, 69, 427–437.

Erickson, G. L. (1979). Children's conceptions of heat and temperature. Science Education, 63, 221–230.

Erickson, G. L. (1980). Children's viewpoints of heat: A second look. Science Education, 64, 323–336.

Ferraro, D., & Van de Kerckhove, W. (2006). Trends in International Mathematics and Science Study (TIMSS) 2003 Nonresponse Bias Analysis (NCES 2007–044). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved November 10, 2008, from http://nces.ed.gov/pubsearch.

Goldring, H., & Osborne, J. (1994). Students' difficulties with energy and related concepts. Physics Education, 29, 26–32.

Harrison, A. G., Grayson, D. J., & Treagust, D. F. (1999). Investigating a grade 11 student's evolving conceptions of heat and temperature. Journal of Research in Science Teaching, 36, 55–87.

International Association for the Evaluation of Educational Achievement. (1995a). TIMSS science items: Released set for population 1 (third and fourth grades). Chestnut Hill, MA: Boston College.

International Association for the Evaluation of Educational Achievement. (1995b). TIMSS 1999 science items: Released set for population 2 (seventh and eighth grades). Chestnut Hill, MA: Boston College.

International Association for the Evaluation of Educational Achievement. (1999). TIMSS 2003 science items: Released set eighth grade. Chestnut Hill, MA: Boston College.

International Association for the Evaluation of Educational Achievement. (2003). TIMSS science items: Released set for eighth grade. Chestnut Hill, MA: Boston College.

Kaper, W. H., & Goedhart, M. J. (2002). "Forms of energy," an intermediary language on the road to thermodynamics? Part I. International Journal of Science Education, 24(1), 81–95.

Kolen, M. J., & Brennan, R. L. (2004). Test equating, scaling, and linking: Methods and practices. New York: Springer-Verlag.

Kruger, C., Palacio, D., & Summers, M. (1992). Surveys of English primary teachers' conceptions of force, energy, and materials. Science Education, 76(4), 339–351.

Lee, H.-S., Linn, M. C., Varma, K., & Liu, O. L. (in press). How does technology-enhanced inquiry instruction with visualizations impact classroom learning? Journal of Research in Science Teaching.

Lee, H.-S., Liu, O. L., & Linn, M. C. (in press). Validating measurement of knowledge integration in science using multiple-choice and explanation items. Applied Measurement in Education.

Lewis, E. L., & Linn, M. C. (1994). Heat energy and temperature concepts of adolescents, adults, and experts: Implications for curricular improvements. Journal of Research in Science Teaching, 31(6), 657–677.

Lijnse, P. (1990). Energy between the life-world of pupils and the world of physics. Science Education, 74(5), 571–583.

Lin, C.-Y., & Hu, R. (2003). Students' understanding of energy flow and matter cycling in the context of the food chain, photosynthesis, and respiration. International Journal of Science Education, 25(12), 1529–1544.

Lindsay, R. B. (Ed.). (1975). Energy: historical development of the concept. New York: Dowden, Hutchinson, & Ross.

Linn, M. C. (2006). The knowledge integration perspective on learning and instruction. In K. Sawyer (Ed.), The Cambridge handbook of the learning sciences (pp. 243–264). New York: Cambridge University Press.

Linn, M. C., & Eylon, B. -S. (2006). Science education: Integrating views of learning and instruction. In P. A. Alexander & P. H. Winne (Eds.), Handbook of educational psychology (pp. 511–544). Mahwah, NJ: Erlbaum.

Liu, O. L., Lee, H. S., Hofstetter, C., & Linn, M. C. (2008). Assessing knowledge integration in science: Construct, measures, and evidence. Educational Assessment, 13(1), 1–23.

Liu, X., Ebenezer, J., & Fraser, D. M. (2002). Structural characteristics of university engineering students' conceptions of energy. Journal of Research in Science Teaching, 39(5), 423–441.

Liu, X., & McKeough, A. (2005). Developmental growth in students' concept of energy: Analysis of selected items from the TIMSS database. Journal of Research in Science Teaching, 42(5), 493–517.

Marco, G. L. (1977). Item characteristics curve solutions to three intractable testing problems. Journal of Educational Measurement, 14, 139–160.

McIldowie, E. (1995). Energy transfer—Where did we go wrong? Physics Education, 30(4), 228–230.

McNeill, K. L., Lizotte, D. J., Krajcik, J., & Marx, R. W. (2006). Supporting students' construction of scientific explanations by fading scaffolds in instructional materials. Journal of the Learning Sciences, 15(2), 153–191.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), Educational measurement (pp. 13–103). Phoenix, AZ: The Oryx Press.

National Assessment Governing Board. (2004). Science framework for the 2005 National Assessment of Educational Progress. Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.

National Center for Education Statistics (2007). Mapping 2005 state proficiency standards onto the NAEP scales (NCES 2007-482). U.S. Department of Education. Washington, DC: Author.

National Research Council. (1996). National Science Education Standards. Washington, DC: National Academy Press.

National Research Council. (2007). Taking science to school. Washington, DC: National Academic Press.

Organisation for Economic Cooperation and Development (2007). PISA 2006 science competencies for tomorrow's world. Paris: Author.

Papadouris, N., Constantinou, C. P., & Kyratsi, T. (2008). Students' use of the energy model to account for changes in physical systems. Journal of Research in Science Teaching, 45, 444–469.

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). Knowing what students know. Washington, DC: National Academic Press.

Rasch, G. (1980). Probabilistic models for some intelligence and attainment tests (Copenhagen, Danish Institute for Educational Research), expanded edition (1980) with foreword and afterword by B. D. Wright. Chicago: University of Chicago Press. (Original work published 1960)

Sandoval, W. A. (2003). Conceptual and epistemic aspects of students' scientific explanations. Journal of the Learning Sciences, 12(1), 5–51.

Schmid, G. B. (1982). Energy and its carriers. Physics Education, 17, 212–218.

Sexl, R. U. (1981). Some observations concerning the teaching of the energy concept. European Journal of Science Education, 3, 285–289.

Smith, C., Wiser, M., Anderson, C. W., & Krajcik, J. (2006). Implications of research on children's learning for standards and assessment: A proposed learning progression for matter and atomic-molecular theory. Measurement: Interdisciplinary Research and Perspectives, 14(1&2), 1–98.

Solomon, J. (1985). Teaching the conservation of energy. Physics Education, 20, 165–170.

Songer, N. B., Kelcey B., & Gotwals, A. W. (2009). How and when does complex reasoning occur? Empirically driven development of a learning progression focused on complex reasoning about biodiversity. Journal of Research in Science Teaching, 46(6), 610–631.

Steedle, J. T., & Shavelson, R. J. (2009). Supporting valid interpretations of learning progression level diagnoses. Journal of Research in Science Teaching, 46(6), 699–715.

Stylianidou, F., Ormerod, F., & Ogborn, J. (2002). Analysis of science textbook pictures about energy and pupils' readings of them. International Journal of Science Education, 24(3), 257–283.

Trumper, R. (1990). Energy and a constructivist way of teaching. Physics Education, 21, 208–212.

Trumper, R. (1991). Being constructive: An alternative approach to the teaching of the energy concept—Part two. International Journal of Science Education, 13(1), 1–10.

Warren, J. W. (1983). Energy and its carriers: A critical analysis. Physics Education, 18, 209–212.

Watts, D. M. (1983). Some alternative views of energy. Physics Education, 18, 213–217.

Wilson, M. (2005). Constructing measures: An item response modeling approach. Mahwah, NJ: Erlbaum.

Wilson, M. (2009). Measuring progressions: Assessment structures underlying a learning progression. Journal of Research in Science Teaching, 46(6), 716–730.

Wright, B. D., & Masters, G. N. (1982). Rating scale analysis. Chicago: Mesa Press.

Wu, M., Adams, R., & Wilson, M. (1997). ConQuest: Generalized item response modeling software. Melbourne, Australia: ACER Press.

Yeh, S. S. (2001). Tests worth teaching to: Constructing state-mandated tests that emphasize critical thinking. Educational Researcher, 30(9), 12–17.