JOURNAL OF RESEARCH IN SCIENCE TEACHING

Towards a Learning Progression of Energy

Knut Neumann,[1] Tobias Viering,[1] William J. Boone,[2] and Hans E. Fischer[3]

[1]Leibniz-Institute for Science and Mathematics Education (IPN) at the University of Kiel,
Olshausenstraße 62, 24098 Kiel, Germany
[2]Miami University, Oxford, Ohio
[3]University Duisburg-Essen, Essen, Germany

Received 15 June 2011; Accepted 3 October 2012

Abstract: This article presents an empirical study on an initial learning progression of energy, a concept of central importance to the understanding of science. Learning progressions have been suggested as one vehicle to support the systematic and successful teaching of core science concepts. Ideally, a learning progression will provide teachers with a framework to assess students' level of understanding of a core concept and to guide students towards a more sophisticated level of understanding. Taking existing research as a point of departure, developing a learning progression involves recurring cycles of empirical validation and theoretical refinement. In this article, we report about our efforts in working towards a learning progression of energy. First, we derived an initial learning progression by utilizing existing curriculum, research on students' understanding, and development of students' understanding of energy. Second, we used these sources of guidance to develop a robust measurement instrument, the Energy Concept Assessment (ECA), based on multiple choice questions. Third, we utilized this instrument to assess the understanding of N ¼ 1,856 students from three different grade levels, Grades 6, 8, and 10. Findings provided evidence that students from Grade 6 mostly obtain an understanding of energy forms and energy sources. Students of Grade 8 additionally demonstrate an understanding of energy transfer and transformation, whereas only students of Grade 10, and then only some of these students, achieve a deeper understanding of energy conservation. We discuss the implications of our findings against the background of existing research on students understanding of energy. Finally, further steps in working towards a learning progression of energy are identified. © 2012 Wiley Periodicals, Inc. J Res Sci Teach © 2012 Wiley Periodicals, Inc. J Res Sci Teach

Keywords: learning progression; energy; assessment; conceptual change; conceptual growth; Rasch measurement

Zusammenfassung: Der vorliegende Artikel beschreibt den ersten Schritt in der Entwicklung und Validierung einer Learning Progression für das Energiekonzept; einem Konzept, das zentral für die Entwicklung eines tiefergehenden Verständnisses der Naturwissenschaften ist. Learning Progressions sollen das das systematische und erfolgreiche Unterrichten zentraler naturwissenschaftlicher Konzepte unterstützen. Idealerweise sollen Learning Progressions Lehrkräften eine Rahmen bieten, den Entwicklungsstand ihrer Schülerinnen und Schüler hinsichtlich des Verständnisses zentraler naturwissenschaftlicher Konzepte einzuschätzen und Unterricht so zu gestalten, dass er die Entwicklung eines elaborierten Verständnisses befördert. Die Entwicklung einer Learning Progression beginnt mit der theoretischen Begründung einer vorläufigen Learning Progression, gefolgt von iterativen Zyklen empirischer Validierung und Überarbeitung. In diesem Artikel berichten wir über unsere Arbeiten

Additional Supporting Information may be found in the online version of this article.
Contract grant sponsor: Deutsche Forschungsgemeinschaft (DFG); Contract grant number: NE-1368/2-1.
Correspondence to: K. Neumann; E-mail: neumann@ipn.uni-kiel.de
DOI 10.1002/tea.21061

zur Entwicklung einer Learning Progression für das Energiekonzept. Im Rahmen dieser Arbeiten wurde zunächst ausgehend von vorliegenden Befunden zum Verständnis und der Entwicklung des Verständnisses von Energie eine vorläufige Learning Progression begründet. Im zweiten Schritt wurde auf Grundlage der Learning Progression ein entsprechendes Instrument auf Basis von Multiple-Choice-Aufgaben entwickelt – das Energy Concept Assessment (ECA). Im dritten und letzten Schritt wurde das Instrument eingesetzt, um das Verständnis von Energie bei N ¼ 1856 Schülerinnen und Schülern der Jahrgänge 6, 8 und 10 zu erfassen. Die Ergebnisse unserer Untersuchung legen nahe, dass Schülerinnen und Schüler aus Jahrgang 6 im Wesentlichen über ein Verständnis von Energieformen und –quellen verfügen. Schülerinnen und Schüler aus Jahrgang 8 zeigen darüber hinaus ein Verständnis von Energieumwandlung und –transport. Ein Verständnis von Energieerhaltung ist nur von Schülerinnen und Schüler aus Jahrgang 10 und dann auch nur von einem Teil dieser Schülerinnen und Schüler zu erwarten. Vor dem Hintergrund dieser Ergebnisse und der bisherigen Forschung zum Energieverständnis, diskutiert der Artikel weitere Schritte für die die Entwicklung einer Learning Progression für das Energiekonzept.

Keywords: learning progression; energie; kompetenzentwicklung; conceptual change; schülervorstellungen; basiskonzept

Over the past decades, scientific literacy has become the overarching aim of science education (Roberts, 2007). That is, science education is expected to provide students with the ability to interact with a world shaped by science and technology. However, given the extensiveness and complexity of scientific knowledge, students cannot be expected to acquire all of this knowledge. Research suggests students are challenged even when they are presented with a limited number of scientific topics for which they are expected to develop a deeper understanding. Given these goals and the realities of learning, science education research has to identify which knowledge is central to participating and acting meaningfully in a scientifically and technologically influenced world. And science education research has to provide a reliable foundation for systematic teaching which demonstrably will allow students to develop sound understanding of such knowledge.

In their report to the National Research Council, Duschl, Schweingruber, and Shouse (2007) suggest focusing on core concepts in the teaching of science and providing ‘‘learning progressions’’ as a basis for the systematic teaching of these core concepts. Within this framework, core concepts are understood as ‘‘foundational in terms of connection to many related scientific concepts’’ (Duschl et al., 2007, p. 5). Learning progressions are intended as a means to align content, instruction and assessment in order to provide students with the opportunity to develop a deeper understanding of the particular concept (Stevens, Delgado, & Krajcik, 2010, p. 688). As such, development of a learning progression does not only include the description of increasingly sophisticated levels of understanding a core concept, but also the description of (a) how students’ level of understanding may be assessed and (b) typical instruction which would foster a more sophisticated level of understanding. Developing a learning progression is an iterative process of empirical validation and theoretical refinement. However, the process starts with the description of an initial learning progression.

One core concept of science that students need to understand, not only in order to be able to explain physical phenomena, but also to understand our technological world, is the concept of energy (Driver & Millar, 1986). For example, given the world’s need for energy it is important that students understand that, although all energy is conserved (from a physics perspective), the usage of energy renders energy no longer useful for human endeavors. In this article, we therefore detail our efforts to describe and validate an initial learning progression of energy.

Theoretical Background

In this section, we discuss the process of developing a learning progression. Subsequently, we present a review of literature concerning students’ understanding of energy.

Based on this review, we propose an initial learning progression of energy for the middle school grades.

Developing and Validating Learning Progressions

To date, a variety of learning progression definitions have been presented in the literature (Duschl et al., 2007; Smith, Wiser, Anderson, Krajcik, & Coppola, 2006; Stevens et al., 2010; Wilson & Bertenthal, 2006). Generally, there seems to be consensus, that learning progressions describe how students’ understanding of scientific concepts or practices develops across multiple grades (Duncan & Hmelo-Silver, 2009). Duschl et al. (2007) suggested four main characteristics of a learning progression. First, learning progressions are based on current research. Second, learning progressions describe intertwined strands of scientific proficiency. That is, learning progressions not only involve knowledge, but also the abilities and skills required to solve real-life problems (Schwarz et al., 2009; Smith et al., 2006; Songer, Kelcey, & Gotwals, 2009). Third, learning progressions are delineated by an upper and lower anchor. The upper anchor is defined by the level of understanding expected from students once they have mastered the learning progression. Students' tentative understanding of a particular idea or concept upon entering the learning progression defines the lower anchor. Fourth, a learning progression is characterized by intermediate levels of understanding that describe students' progression from the lowest level of understanding (the lower anchor) towards the most advanced level (the upper anchor). Krajcik, Drago, Sutherland, and Merritt (2012) suggest that a learning progression should moreover provide (1) detailed information on learning performances (i.e., what kind of tasks students' will be able to solve on each level of the learning progression) and (2) detailed information regarding the teaching which would take place at each level (i.e., what kind of teaching would foster students progression to the next, more sophisticated level of understanding).

Development of a learning progression starts by defining the upper and lower anchor, followed by description of intermediate steps (cf. Stevens et al., 2010). Description of the upper anchor is usually informed by policy documents such as standards or analysis of the domain (Duncan & Hmelo-Silver, 2009; for an example see Berland & McNeill, 2010) and is informed by science education theory (Mohan, Chen, & Anderson, 2009). The lower anchor is often described based upon research on students understanding or outcomes defined for a preceding stage of schooling (e.g., primary school standards may be used to define the lower anchor for a middle school learning progression). The intermediate steps borrow heavily from previous research on students' learning (Duncan & Hmelo-Silver, 2009), analysis of the domain or both (Berland & McNeill, 2010). This process leads to an initial learning progression.

This initial learning progression requires empirical validation. Two principal approaches to validate an initial learning progression have been identified (cf. Duncan & Hmelo-Silver, 2009). Both approaches require instructional interventions as well as measurement instruments aligned with the learning progression. The first approach starts with the development of an instructional intervention suitable to foster students' progression through a portion of the learning progression (e.g., Nordine, Krajcik, & Fortus, 2010). For this purpose, the initial learning progression needs to be detailed through instructional components and learning outcomes in order to guide curriculum and material development (Krajcik et al., 2012). Evaluation of instructional intervention allows for investigating whether students indeed progress through the learning progression as hypothesized. If students are found to progress through the learning progression as hypothesized, validation of the learning progression continues with the development of an instructional intervention which addresses the next component of the learning progression. The second approach is initiated with the development of a measurement instrument, suitable to investigate students' progression through the learning progression as a whole (Mohan et al., 2009). The instrument may be used to investigate how the current curriculum impacts students learning. This requires the existing curriculum to be to some degree aligned with the learning progression, which can be the case when the same research base has been used both to inform the design of the current curriculum and the learning progression. The investigation of students learning will have to include repeated measurement of students understanding. The time frame for these repeated measurements being determined by the curriculum (i.e., if the curriculum spans Grades 6–9, students could be tested in Grades 6, 7, 8, and 9). Given the time and costs associated with longitudinal studies, it seems reasonable to initially conduct a cross sectional study (the time frame being the same as above, i.e., Grades 6, 7, 8, and 9). This approach facilitates information about typical students’ progression in mastering the learning progression under the existing curriculum. In both approaches if students are found to not progress through the learning progression as hypothesized, the alignment of the curriculum and the measurement instrument with the learning progression, as well as the learning progression itself need to be reconsidered and revised where appropriate. Then the process of empirical validation needs to be commenced once more.

In summary, development of a learning progression is an iterative cycle of empirical validation and theoretical refinement. It is important to note, that validating a learning progression does not mean to demonstrate that every single student will be shown to move through the learning progression as hypothesized (Duncan & Hmelo-Silver, 2009; Stevens et al., 2010). Instead validating a learning progression relates to obtaining evidence about students’ progression with respect to the learning progression and to using this information to iteratively align curriculum, instruction and assessment to foster students’ progression in mastering the particular domain in the best possible way.

Students’ Understanding of the Energy Concept

The energy concept is of central importance to understanding the biological, chemical, physical, and technological world (Driver & Millar, 1986). One reason why energy is such an important concept is that it is a conserved quantity (Feynman, 2011). That is, whenever energy is transferred from one place to another, or converted from one form into another, the overall amount of energy is conserved. Only when energy is transformed or transferred is some of the energy degraded. The main characteristics of energy, from a scientific perspective, are (1) energy comes in different forms, (2) energy can be transferred or transformed from one form into another, (3) whenever energy is transformed or transferred some of it is degraded, (4) the overall amount of energy remains conserved (Duit, 1986). Driver and Millar (1986) argued that this understanding of energy is a prerequisite for understanding the relevance of energy for society. As a consequence of this argument there is a particular consensus among science education researchers that students should obtain a particular understanding of energy with respects to the four characteristics listed above (e.g., Duit, 1984; Solomon, 1986; cf. Doménech et al., 2007). This consensus has found its way into numerous policy documents, although as would be expected, in different combinations and with different emphasis (American Association for the Advancement of Science, 2001; National Research Council, 1996, 2012; Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK], 2005).

Extensive research on students’ understanding of the energy concept [for an overview see reviews by Duit (1986) or Vosniadou (2008)] suggests that students have not mastered an understanding of energy as laid out in policy documents. Research suggests not only that students enter formal schooling with a variety of different energy conceptions stemming from everyday experience and language (e.g., Duit, 1981a; Lijnse, 1990; Solomon, 1983a, 1983b; Trumper, 1993; Watts, 1983b), but also that students have difficulty differentiating between energy and other scientific concepts such as force (Viennot, 1979; Watts, 1983a), power (Goldring & Osborne, 1994; Watts & Gilbert, 1983), or temperature (Erickson & Tiberghien, 1985; Lewis & Linn, 1994). Students have also been found to have difficulty understanding energy degradation (e.g., Black & Solomon, 1983) and to lack a true understanding of the conservation of energy (e.g., Boyes & Stanisstreet, 1990; Driver & Warrington, 1985).

Development of Students Understanding of the Energy Concept. Based on a review of the research on students conceptions of energy, Driver, Squires, Rushworth, and Wood-Robinson (1994) were among the first to suggest a sequence along which students were hypothesized to progress, namely: awareness of personal energeticness, extending energeticness to other living things, awareness of nonliving things spontaneously being able to do things, extending energeticness to some nonliving things that possess energy, awareness of stored energy in elastic materials, awareness of gravitational potential energy, being able to tell the energy story, that is, describing events in energy terms, awareness of energy conservation, that is, describing events in quantitative terms, and awareness of energy degradation, that is, recognition that things run down, and efficiency. Liu and McKeough (2005) utilizing Driver et al.'s (1994) work hypothesized that students progression in understanding the energy concept is characterized by five distinct, hierarchically ordered conceptions: perceiving energy as activities or abilities to do work (Activity/Work), identifying different energy sources and forms (Form/Source), understanding the nature and processes of energy transfer (Transfer), recognizing energy degradation (Degradation), and realizing energy conservation (Conservation). Based on an analysis of TIMSS data these authors argued students of greater age were able to solve items requiring a more elaborate conception of the energy concept (also see Liu & Ruiz, 2008). The hierarchy of conceptions of the energy concept proposed by Liu and McKeough (2005) was also observed by Dawson-Tunik (2006) and was also confirmed in a recent study by Lee and Liu (2010).

Development of Students' Understanding of Individual Conceptions. So far, we have discussed the varied conceptions of energy students may hold and the potential sequence of conceptions along which students progress in developing an increasingly sophisticated understanding of the energy concept. However, the question remains—how do students develop an understanding of individual conceptions. Bransford, Brown, and Cocking (2000) suggest that the development of conceptual understanding is related to the acquisition of a well-organized and structured knowledge base (cf. Stevens et al., 2010). The mechanism of knowledge acquisition has also been utilized to investigate, and better understand how students' understanding of concepts might progress (Aebli, 1980; Bruner, 1966; Case, 1985; Commons, 2007; Fischer, 1980; Gagne & White, 1978; Piaget, 1972). Much of this research has built upon the idea that learning corresponds to the development of increasingly more complex cognitive operations (e.g., Commons, 2007) or the development of an increasingly more complex knowledge base (e.g., Aebli, 1980). In this notion, an individual's ability or knowledge base can be thought of as a structure of elements. More complex structures emerge from combinations of less complex structures as a consequence of learning. Researchers have argued that breadth and depth (i.e., the complexity) of an individual's knowledge base may be considered a measure of conceptual understanding (Alao & Guthrie, 1999). From this point of view, novices can be considered to start out with a rather fragmented knowledge base, with few connections between individual knowledge elements. Through acquiring new knowledge

elements and establishing connections between existing and new elements, they develop a well-structured and organized knowledge base. They become an expert (Bransford et al., 2000; cf. Stevens et al., 2010).

With respect to the field of science, researchers recently have made use of the idea of an increasingly complex knowledge base to predict the difficulty of science tasks (Bernholt & Parchmann, 2011; Kauertz & Fischer, 2006; Liu, Lee, Hofstetter, & Linn, 2008). For instance, Kauertz and Fischer (2006) have suggested six levels of complexity to describe the knowledge required to solve physics tasks: one fact, several facts, one relation, several unconnected relations, several connected relations, and conceptual understanding. Tasks that require the knowledge of facts (e.g., that wood floats) were considered to be the easiest, whereas those tasks requiring the knowledge of a system of intertwined relations between a series of facts, were thought to require a conceptual understanding and therefore were considered to be the most difficult items. In a subsequent study, Kauertz and Fischer (2006) authored a set of tasks for each complexity level and administered these tasks to a sample of 10th graders. Results showed that indeed difficulty was correlated with the complexity level of an item. Bernholt and Parchmann (2011) utilized a slightly different scale to author tasks requiring a particularly complex knowledge base to be solved: everyday knowledge, factual knowledge, knowledge about processes, knowledge about linear relations, knowledge about multivariate interdependencies; whereas tasks requiring everyday knowledge were considered the easiest, tasks requiring knowledge about multivariate interdependencies were considered the most difficult ones. Results of a study in which the test items were administered to students from Grades 6 to 10, demonstrated a correlation between the complexity of the knowledge base required to solve an item and the item's difficulty. Liu et al. (2008) suggested that the depth of students' reasoning (i.e., the quality of the link students can establish between scientific ideas) may be used as a measure for students' conceptual understanding. Rating students' open-ended explanations of their answer choice when solving TIMSS items, Liu et al. (2008) found that a higher complexity corresponds to a deeper understanding.

In summary, research suggests the complexity of students' knowledge about a particular concept corresponds to the level of students' conceptual understanding of the concept. Previous research suggests four principle levels of complexity of students knowledge: a fragmented knowledge base, where students possess singular pieces of knowledge unconnected to each other (facts), a knowledge base, where simple connections have been established between the individual knowledge pieces (mappings), a knowledge base, where more qualified connections exist (relations), and a knowledge base which embraces complex intertwined connections which the individual knowledge pieces together to form particular structures (concepts).

An Initial Learning Progression of Energy

Our description of an initial learning progression of energy was informed by the German National Education Standards (NES) for Physics (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK], 2005), existing science education theory about what a comprehensive understanding of energy means (e.g., Duit, 1986) as well as existing research about students understanding of energy and the development of students understanding of energy (e.g., Liu & McKeough, 2005). The German NES view energy as a core concept, that will allow students to develop a sound understanding of physics. The standards expect students to develop an understanding of energy, which includes an understanding of energy forms and sources, energy transfer and transformation, energy degradation, and energy conservation (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK], 2005).

This is well in line with science education theory (e.g., Duit, 2007). Consequently, using these resources, the upper anchor of our energy progression is defined to correspond to students having fully developed an understanding of the four aforementioned aspects of energy listed above.

Research on students understanding of energy suggests that students enter formal schooling with a variety of energy misconceptions (e.g., Duit, 1981a; Lijnse, 1990; Solomon, 1983a, 1983b; Trumper, 1993; Watts, 1983b). Research also suggests that after a first unit of formal schooling students may be expected to have developed an understanding of energy that corresponds to understanding that energy may come in different forms and from different sources (Liu & McKeough, 2005; Liu & Ruiz, 2008). As such, we defined the lower anchor of our learning progression to be that students understand how energy comes in different forms and from different sources.

The definition of intermediate learning progression levels (between the upper and lower anchor) was informed by two strands of research: (1) research on how students develop conceptual understanding by developing an increasingly complex knowledge base and (2) research on how students understanding of the energy concept changes over time. Using research such as that conducted by Liu and McKeough (2005), Lee and Liu (2010), or Nordine et al. (2010) we hypothesized a learning progression in which students progress in their learning about the energy concept from an understanding that energy comes in different forms, and from different sources, toward an understanding of energy conservation along the following sequence of conceptions: energy forms and sources, energy transformation and transfer, energy degradation, and energy conservation. In an effort to detail how students progress in their understanding from one level to the next we drew upon research on how students develop conceptual understanding through acquiring increasingly complex knowledge. Based on Kauertz and Fischer's (2006) and Bernholt and Parchmann's (2011) work, we hypothesized that students develop an understanding of each of the four energy conceptions by forming an increasingly complex knowledge base about each conception along the following steps: facts, mappings, relations and concepts. Figure 1 depicts the resulting initial energy learning progression.

This initial learning progression embraces four major levels and four minor levels for each of the four major levels. We assume that the typical student enters formal schooling with a variety of misconceptions. Our initial learning progression envisions that students will first learn about energy forms. They will learn that energy comes in different forms (Level 1: facts/forms) and that these forms are represented through physical measures such as velocity in the case of kinetic energy (Level 2: mappings/forms). Students will learn about the relation between the respective forms of energy and related physical measures (Level 3: relations/forms) and throughout this process understand that energy is a somewhat abstract quantity that is assigned to different forms based on observed measures (Level 4: concept/forms). Learning continues with students obtaining factual knowledge about energy transformation (Level 5: facts/transformations), establish connections between facts in course of ongoing instruction, and develop a conceptual understanding of energy transformations (Level 8: concept/transformations). A similar process will take place for energy degradation and conservation when students progress towards a comprehensive understanding of the energy concept (Level 16: concept/conservation). In such a sequence one expects that students who have learned that energy comes in different forms and how these forms relate to physical measures (Level 2: mappings/forms) will sometimes be able to recognize in observing a particular phenomenon that when energy decreases in one form, it increases in another (cf. Nordine et al., 2010). Thus some overlap between levels 1–4, 5–6, 7–11, and 12–16 is predicted.

Figure 1. Initial learning progression of energy (visualization according to Wilson, 2009).

The aim of the study is to provide details of an initial cycle of empirical validation and subsequent theoretical revision of the initial energy learning progression through development of a measurement instrument. The specific research questions are:

(1) To what extent is the developed measurement instrument suitable for assessing students understanding of energy with respect to the hypothesized learning progression?
(2) To what extent does the hypothesized learning progression describe students' progression in understanding the energy concept?

Methods

Based on the initial learning progression, a measurement instrument, the Energy Concept Assessment (ECA), was constructed. The instrument was composed of a set of multiple choice items developed to measure the complexity of students understanding with respect to four energy conceptions: (i) forms and sources, (ii) transfer and transformation, (iii) degradation, and (iv) conservation. The ECA was administered to a sample of N ¼ 1,856 students of Grades 6–10. We used Rasch analysis to obtain linear measures of person ability and item difficulty. This analysis facilitated an assessment of students' progression in understanding energy with regard to the hypothesized learning progression.

Instrument Development

An iterative multistep procedure was utilized to develop the ECA. Items were authored, piloted, and revised. A second round of pilot data were collected and evaluated, and the instrument items again revised to inform the creation of a final version of the ECA.

Item Authoring. The process used to guide the authoring of items aimed for the best possible match between the proposed learning progression and developed items. That is, for all levels of the learning progression, items were authored to help determine whether students had mastered a particular level. Item authoring guidelines were used to guide the process of item construction.

As a first step in item development, a context was chosen. Learning progressions describe how students progress in their understanding of core concepts. More specifically, how students are able to use their knowledge to describe real-world scenarios. Therefore, when authoring items to assess students understanding of the energy concept, items were designed as a function context; that is, of real-world scenarios which require students to use their understanding of energy to explain those scenarios. For example, a skateboarder riding a half pipe (see Supporting Information Table 1 for a full overview of the contexts used in item development).

Following this step, four multiple choice items of varying complexity were authored for each of the four conceptions of energy. This step of the process was informed by a detailed description of the initial learning progression (see Supporting Information Table 2, cf. Figure 1). In the first step of this process, for each of the four conceptions, one multiple-choice item was designed. Each of these items was composed of a brief description of a scenario in the chosen context (e.g., the skater standing at the top of the half-pipe about to drop into it), a respective picture, a question, and a set of four answering options (cf. Figure 2). One option, the attractor, was correct. This option was designed to require a conceptual understanding of the respective conception of energy (i.e., forms and sources, transfer and transformation, degradation, conservation) to be correctly solved. In case of the context ‘‘skateboarder riding a half-pipe’’ conceptual understanding would require students (1) to correctly identify that a skateboarder at the top of the half-pipe may be assigned gravitational energy (forms and sources); (2) to correctly describe how energy is converted from gravitational energy into kinetic energy and vice versa when the skateboarder goes down and up in the half-pipe (transfer and transformation); (3) to explain why the skateboarder eventually stops moving (degradation); and (4) to explain what happens to energy initially held by the skateboarder (conservation). The other three answer options, the distractors, were incorrect. These answer options were designed based on a list of misconceptions compiled from a review of literature on students’ conceptions about energy (see Supporting Information Table 3). Based on the scenario described in the item, and the conception the item was designed to address, applicable misconceptions were identified from the list of misconceptions. These misconceptions were used to author distractors. This way for given context, one item was designed for each of the four conceptions of energy. These items were designed to require conceptual understanding of the respective conception. As a consequence, these items were considered to be of highest complexity (concept).

In the next step, for each of the four items, three items of lower complexity were developed by successively adding more information to the highest complexity item (cf. Figure 2). That is, to reduce item complexity to the level of ‘‘relations,’’ one piece of information (information A, Figure 2) was added. This additional information concerned a central fact necessary to solve the items that students—despite possessing a particularly complex knowledge base—might have missed. In order to reduce item complexity to the level of ‘‘mappings’’ a second piece of information (information B, Figure 2), was added to items. This was a central piece of information on the level of a relation between two facts expected to help students with a less complex knowledge base solve the item. In order to reduce the complexity of an item to the lowest level, a third piece of information (information C, Figure 2), another relation that should allow students with only factual knowledge to correctly solve the item, was added to the items.

Figure 2. Sample Item from the Energy Concept Assessment (ECA). The item is part of the context ‘‘Skater riding a half-pipe’’ and it is designed to assess students understanding of ‘‘Energy Forms.’’ The complexity of the item may vary according to the information provided: concept (no information), relations (A), mappings (A þ B), facts (A þ B þ C).

To summarize, the additions that were made step by step to items, began at the highest complexity level, that level named ‘‘concept’’ involved the addition of no information for respondents presented with the item. The next level of item complexity concerned items which are named ‘‘relations.’’ For these items one piece of information, a fact, was added to the concept level items. Items which were one step below the complexity of relations items were named ‘‘mappings’’ items. These items consisted of two pieces of information having been added to the item—one fact and one relation. For items (named ‘‘facts’’) of the lowest complexity level, three pieces of information, one fact and two relations were added to the original concept level form of the item.

Piloting—1st Stage. Following development of an initial set of items for five contexts (cf. Supporting Information Table 1), a first stage of pilot testing was conducted to further inform item and test development. For this purpose 64 items that best represented the learning progression were selected from the developed set of 80 items. Since items of the same conception, but different complexities were not independent of each other, the 64 items were distributed throughout a set of 16 booklets with four independent items each. The 16 booklets were administered to a sample of N ¼ 72 students of Grades 8 and 10. The students were from selected German schools which would not be used to provide the student sample for the main study. Each student was administered one booklet randomly chosen from the total number of 16 booklets. Also each student was asked to complete a questionnaire for each item in order to provide more insight into the overall functioning of the items. The provided questionnaire was based on a questionnaire created by the American Association for the Advancement of Science (2007) to assess item quality during the process of item development. Students were asked (1) to note if there was something in the item they found confusing, (2) to circle all words present in the item they did not understand, (3) to indicate the difficulty of the item using a four step Likert scale (very easy, easy, difficult, and very difficult), (4) how easy the text was to read using the same four step Likert scale, and (5) to evaluate item answering options with regard to (i) difficulty (using a two step scale of ‘‘easy’’ and ‘‘difficult’’), (ii) clarity (using a two step scale of ‘‘confusing’’ and ‘‘not confusing’’), and (iii) similarity (using a two step scale of ‘‘similar to each other’’ and ‘‘different from each other’’). The questionnaire also asked students whether the picture presented in the item was helpful (yes/no) and whether they guessed when answering the item (yes/no).

Analysis of students’ answers to items and to the questionnaire revealed several findings which were used to inform item revisions: While students considered items quite easy to read (Q4: 37.8% very easy, 42.0% easy, 14.6% difficulty, 3.1% very difficult), students’ evaluation of the answering options (Questionnaire item Q5) suggested that students considered options confusing and difficult when they were very similar. For example, when one option was

‘‘Kinetic energy is converted into potential energy’’ and another option was ‘‘Potential energy is converted into kinetic energy.’’ Results of this data collection were used to make test item options less similar.

Piloting—2nd Stage. Following the revision of items as informed by the initial pilot, items were developed for an additional 12 contexts (see Supporting Information Table 1). Then a second stage of piloting was conducted. This time piloting was used to investigate how the items function with respect to assessing different levels of students’ understanding with respect to the hypothesized learning progression. In order to collect item data that would facilitate such an analysis, a smaller number of items were administered to a larger number of students. Two items representing each of the 16 levels of the learning progression were selected. These items were distributed across two test booklets of 20 items each. A total of eight items were presented in both booklets to allow for test equating (Wright & Stone, 1979). The two test booklet forms were administered to a sample of N ¼ 395 (Grades 7–11) students from volunteering schools in which students had received instruction with regard to energy during the preceding school year. The schools were different from the ones participating in the 1st stage pilot.

Analysis of collected data, suggested that items of higher conceptions of energy were found to be more difficult than items of lower conceptions (Neumann, Viering, & Fischer, 2010; Viering, Fischer, & Neumann, 2010). Only three items of the conception ‘‘forms and sources’’ were found to be more difficult than expected compared to items of the other conceptions. All three of these items were related to the identification of the source of a particular form of energy. For example, students were asked to consider the source of the energy of a bike being ridden. Analysis of these items suggested, that the items implicitly required an understanding of energy transformation. For example, with regard to the test item including a bicycle rider, the correct answer often chosen by the students was that the source of energy is the food the rider consumed. However, one incorrect often selected answering option was that the source of energy was the rider’s legs (Neumann et al., 2010) As a result all items concerning energy forms and sources were revised so that answering options would not require an implicit understanding of energy transformations or transport. Following this revision, all items were provided to two external reviewers. Reviewers were asked to provide feedback on the tasks. Reviewers’ comments included suggestions concerning how to rephrase particular parts of items. Some reviewer comments also identified alternative answering options that might be correct in very unique circumstances. Based on the reviewer feedback items were once again revised.

Following a final round of item revision, a set of 120 items were chosen from the total of 272 potential items and distributed across 12 test booklets. The 120 items were chosen based on the following criteria: (1) representation of the learning progression, (2) equal distribution of items across the 16 possible combinations of conception and complexity, (3) the constraint that no two items of the same context and conception must be administered to the same student, and (4) overall item quality. The final set of 120 items represented 13 contexts (for an overview of item distribution across contexts, conceptions and complexity see the technical handbook on the ECA provided as online Supporting Information. Four contexts were not included in the final instrument, as the items from these contexts did not fully fulfill the above criteria). To create the 12 test booklets, a first step was the distribution of the 120 items into 12 blocks (named A-L), with 10 items per block. Each test booklet was composed of two blocks such that two adjacent booklets shared a common block (AB, BC, . . . LA). This organization of items in this manner and the use of Rasch analysis allows the difficulty of all items to be expressed on the same metric and allows the test performance of all respondents to be expressed on the same metric (for details on test linking and equating see Kolen & Brennan, 2004).

Data Collection and Data Entry

Data were collected from N ¼ 1856 students from Grades 6 to 10 in the state of North-Rhine-Westphalia, Germany. Students attended one of eight “Gymnasium,” the most academic of German school tracks. Most Gymnasium students will continue their education and earn a university degree. The state of North-Rhine-Westphalia was selected for two reasons: Firstly, it is the most populous German state. Secondly, the physics curriculum in North-Rhine-Westphalia (Kultusministerium des Landes Nordrhein-Westfalen, 1993) places a particular emphasis on the teaching of energy from Grades 5 to 10.

The North-Rhine Westphalia energy curriculum for Grades 5–10 at Gymnasiae is designed as a spiral curriculum. Moreover, the curriculum is organized based on the four conceptions of energy. During three, two grade periods (5/6, 7/8, and 9/10) schools are free to choose when to cover energy (e.g., some schools may choose to cover energy in Grades 6, 8, and 10 while other schools may cover energy in Grades 5, 8, and 9). In Grades 5/6 all four conceptions are introduced to students, with a focus on transfer and transformation as well as degradation; conservation is covered nonquantitatively. In Grades 7/8 and 9/10 the concept of energy is revisited in different contexts. In Grades 9/10 quantitative considerations are added. Students after completion of Grade 6, after completion of Grade 8 and after completion of Grade 10 usually will have received the same amount of instruction concerning the four conceptions of energy. Therefore, the large sample utilized for the final data collection was composed of school classes of Grades 6, 8, and 10. Five classes of Grade 7 which, according to the schools, had received no teaching on energy in Grades 5 and 6, but in Grade 7 had been taught energy concepts, were also included. One class of Grade 9 which had already had completely covered the topic of “energy” was also included in the sample. The data from Grades 7 and 9 were included in the sample and evaluated as if data were of Grade 6 or 10, respectively. Supporting Information Table 4 provides details regarding the sample as a function of school and grade. Students were allocated 45 minutes to complete the ECA.

Data Analysis

Following data entry, Rasch analysis was used to evaluate instrument functioning, and obtain linear measures for further statistical analysis with respect to the validation of the learning progression. Rasch analysis is based on a mathematical model which facilitates the computation of linear measures for item difficulty and person ability (for details see for example Bond & Fox, 2007). Linear measures must be computed for parametric tests, for raw scores are potentially nonlinear and therefore violate assumptions of parametric tests (for details see Field, 2009). An additional aspect of the model is that measures of students’ ability and item difficulty are expressed on the same (interval) scale. Expressing item difficulty and person ability on the same scale allows for the interpreting of student achievement as a function of item content. In our study person measures can be interpreted in light of item conception and complexity. Use of the Rasch model moreover provides the possibility to compare person measures of individuals who were administered different test booklets. As long as a subset of items is identical across two booklets, the performance of students administered the two booklets can be compared. Our analysis was informed by the extensive literature in the field of Rasch measurement (e.g., Bond & Fox, 2007) and the application of Rasch measurement in science education (e.g., Liu & Boone, 2006; Liu, 2010). For all Rasch analysis conducted for this study Conquest 2.0 (Wu, Adams, & Wilson, 2007) was utilized. The statistical software package R 2.10.1 was used for statistical analysis of the data set.

Instrument Functioning. Rasch analysis provides one technique by which the psychometric functioning of an instrument can be monitored. This typically includes the exploration of item misfit, summary statistics regarding the instrument and the distribution of items and persons across the latent trait (e.g., Boone & Scantlebury, 2006; Liu & McKeough, 2005; Neumann, Neumann, & Nehm, 2011). Item fit statistics commonly used to investigate item misfit include MNSQ oufit and MNSQ infit as well as respective standardized values (ZSTD). Since the MNSQ outfit is sensitive to outliers, often MNSQ infit is given priority when analyzing item misfit. Typically different acceptability intervals for MNSQ fit and ZSTD fit values are employed depending on the sample size per item (e.g., Bond & Fox, 2007). In our analysis moderate cut-off levels, that is, MNSQ infit acceptability values of between 0.8 and 1.2 and ZSTD values of below 2 were applied. In addition to item fit statistics, we computed traditional item discrimination and identified items with a discrimination below 0.2 which is a common cut-off criteria (e.g., Adams, 2002). Generally these items can be viewed as not sufficiently discriminating between students.

Summary statistics provide several indices which also can be utilized to identify instrument function such as test reliability. In the context of Rasch analysis, typically at least two measures of reliability are considered: item separation and person separation reliability (Wu et al., 2007). Additional information about the instrument functioning may be obtained from the distribution of item difficulty and person ability. Wright Maps which present a plot of both persons and items also provide extensive information about how well persons and items

are distributed along the latent trait. This allows, for example, for the identification of person-item mismatch patterns (Bond & Fox, 2007) and facilitates further assessment of construct validity.

Learning Progression Validation. Measures of item difficulty and person ability obtained from Rasch analysis were used for further statistical analysis to obtain information for added refinement of the measurement instrument and detailing of the learning progression. For instance, statistical analysis consisted of utilizing item difficulty measures for an analysis of item difficulty as a function of ‘‘conception’’ and ‘‘complexity.’’ We hypothesized that students will first develop a successively more complex understanding of energy forms and sources, and then students will proceed to a mastery of energy transfer and transformation. This is then followed by students understanding energy degradation and finally students master the topic of energy conservation. If this theory holds true, items of higher conceptions should generally be more difficult than items of lower conception. For each of the conceptions, items of higher complexity should be more difficult than items of lower complexity. Also, students of higher grades should be more able (have a higher likelihood of correctly answering items) than students of lower grades. Linear Rasch person measures were used for the computation and comparison of students’ ability as a function of school grade. Analysis of variances (ANOVA) was used to investigate differences in the person measures and item measures with respect to school grade and conception/complexity. In order to investigate whether item difficulty indeed increases with conception/complexity and whether students’ ability increased with their grade, Kendall’s t correlation coefficient was utilized. Student’s t-test was utilized for comparison of item difficulties for different conceptions or complexity levels.

Results

The data obtained with the ECA were analyzed in two steps. In the first step, we investigated whether the ECA was functioning psychometrically as intended. In the second step, we investigated whether the ECA indeed documented a progression of student understanding as hypothesized.

Instrument Functioning

In order to investigate the psychometric functioning of the ECA we reviewed (1) item fit statistics in order to identify misfitting items, (2) summary statistics to explore the overall characteristics of the instrument, and (3) the Wright Map to ensure items and persons were well distributed across the latent trait.

Item Fit Statistics. Applying the moderate cut-off levels for MNSQ fit values of between 0.8 and 1.2 and respective ZSTD values of below 2, we found one of the 118 items with valid data to have an MNSQ fit value of more than 1.2. No items were found to have an MNSQ fit value of less than 0.8. Four items were found to have a ZSTD fit value of more than 2.0. This suggests that these items, albeit having exhibited good fit, deviate from the item response function. Investigating item discrimination, another 10 items were found to have a discrimination below 0.2. A low discrimination indicates that the chance of answering these items correctly does not sufficiently depend on students' actual ability. That is, in addition to the identified misfitting items, it is questionable whether these 10 items sufficiently define the latent trait. We excluded these 15 items and performed another Rasch analysis of the remaining 103 items, which suggested one more item might misfit (the items ZSTD fit was >2.0). After excluding this item we obtained a set of 102 items all with adequate fit. This set of items was used for all further analysis.

Summary Statistics. Using the set of 102 items with adequate fit, overall statistics of the instruments were examined starting with the instrument's reliability. The overall item separation reliability of the remaining 102 items was found to be 0.98. This indicates that item parameters could be well separated. However, item separation reliability is usually high with large sample sizes. Given the large sample of our study we further investigated person separation reliability (more specifically WLE person separation reliability). Person separation reliability has been suggested by Wright and Stone (1979) as a measure of test reliability that can much be interpreted as reliability measures used in classical test theory such as Cronbach's α (for details see Field, 2009). (WLE) person separation reliability for our data set was determined to be 0.61. Although this value may be considered acceptable (Kline, 2000), it also suggests that there may be some limitations in obtaining reliable person ability measures using the ECA.

As it is typically the case with Rasch analysis, mean item difficulty is set to zero logits (Wu et al., 2007). The mean standard error of item difficulty estimates were determined to be 0.11 logits. This is a considerably small error given the item difficulty span of between −1.78 and 1.38. Mean person ability was −0.66 logits with a minimum of −4.13 logits and a maximum of 3.81 logits. The average person ability being lower than the average item difficulty suggests the ECA was slightly more difficult than the sample as a whole was able in average. However, this was expected as only students from Grade 10 were expected to be able solve all or most of the items. Students from Grade 6 were expected to only be able to solve items regarding energy forms and sources, that is, the easier items. This particularly low mean standard error in item difficulty is due partially to the fact that there are more persons answering items than there are items answered per person (Rasch, 1960).

Wright Map. To further investigate the distribution of items as a function of student abilities, a Wright Map was constructed. Wright Maps can be used to present the distribution of item and person measures for a particular sample (persons) and instrument (items). Since item linking was used in the data collection, all students (regardless of grade and test form) can be compared, and all test items (regardless of test form) can be compared. Figure 3 shows the Wright map for the final set of 102 items and all 1856 students who completed the test. The map shows the distribution of person ability on the left side of the map and the distribution of item difficulty on the right hand side of the map. The lower student ability and easier item difficulty the closer to the bottom of the map a person or item is plotted. Higher ability students and difficult items are plotted closer to the top of the map. A student with an ability that equals an item's difficulty (e.g., a person in the same line with an item) has a 50/50 chance of solving that item correctly.

A detailed evaluation of the Wright Map reveals that the person ability measures are well distributed over the range of measures. The Wright Map shows, for example, that item MI11 (context ‘‘milk,’’ conception ‘‘forms and sources,’’ lowest complexity) is the easiest item and item WK43 (context ‘‘wind power station,’’ conception ‘‘conservation,’’ highest complexity) is the most difficult item. A review of person ability with regard to these items shows that numerous students have a person ability estimate lower than MI11's difficulty. For these students their chance of solving even the easiest item is below 0.5. Only a few students' ability is greater than WK43's difficulty. The distribution of items between these two items is slightly skewed towards the higher person abilities. However, since the ECA was developed

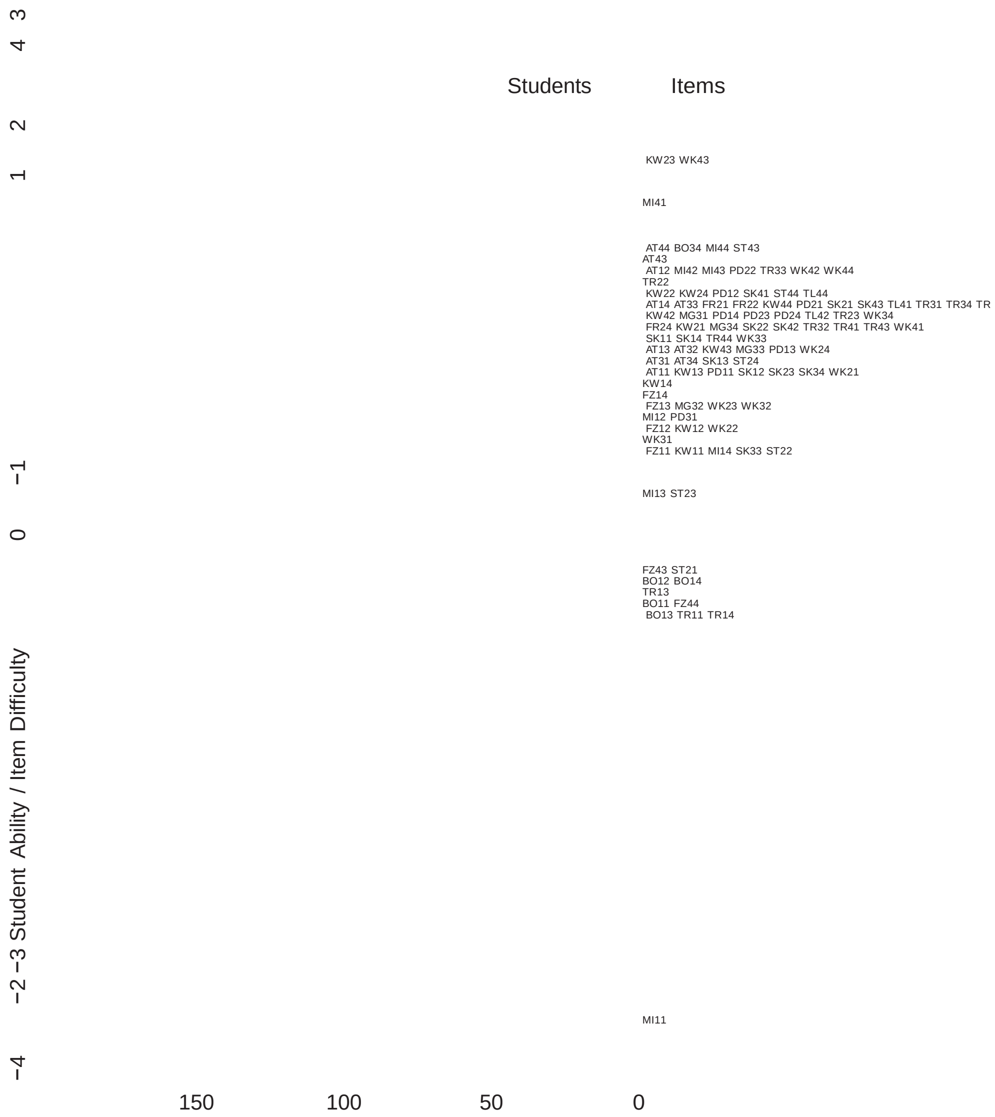

Figure 3. Wright map of the ECA with 102 items and N ¼ 1,856 students.

to identify students' progression in mastering an understanding of the energy concept, and the instrument was not designed to differentiate among students upon their entry into the learning progression, the distribution of items can be viewed as an intended feature of this instrument. For an instrument which aims to measure such a wide range of students, from 6th grade onward, having a larger number of harder items will provide more detail as to the progression of typical students as they learn the energy concept. However, the limited number of items that can differentiate between students with a particularly low ability measure might explain the limited overall reliability of the instrument. More items in this part of the scale would be needed to differentiate between students with a particularly low ability (e.g., when used to assess Grade 6 students' abilities).

Learning Progression Validity

To investigate the manner in which the latent trait defined by the ECA aligns with the hypothesized learning progression, item difficulty and person ability measures were investigated with respect to the hypothesized learning progression levels. We hypothesized that students develop an understanding of energy following a sequence of four individual conceptions of energy: energy as forms and sources, energy transfer and transformation, energy degradation and energy conservation. Moreover, we hypothesized that students will develop understanding of each conception by building a successively more complex knowledge base about the conception. If the ECA items define such a learning progression, item difficulty should depend on the conception of energy the respective item corresponds to. That is, items concerning energy forms and sources should in principle be among the easiest items, while items concerning energy conservation should be among the most difficult. Moreover, for each conception item difficulty should depend on item complexity. That is, for all items concerning the conception ‘‘energy forms and sources,’’ the least complex items should be the easiest, the most complex items should be the most difficult ones.

Progression of Difficulty by Conception. In order to investigate the progression of item difficulty by conception, items were classified by conception. The descriptive statistics in Table 1 show that the average item difficulty is the lowest for items regarding the conception ‘‘forms and sources,’’ whereas items regarding ‘‘energy conservation’’ have the highest average item difficulty. Items related to ‘‘transfer and transformation’’ and ‘‘degradation’’ range in the middle regarding item difficulty. The average item difficulty of ‘‘degradation’’ items is slightly lower than our hypothesized learning progression would predict. An analysis of variance (ANOVA) of item difficulty, grouping items by conception suggests that the effect of conception on item difficulty is indeed statistically significant, $F(3, 98)$ ¼ 12.58, $p < 0.001$, h^2 ¼ 0.28.

An additional analysis of the level of correlation between item difficulty estimates and conceptions was conducted to access whether there is a trend that items of higher (more advanced) conceptions exhibit higher item difficulty. For this analysis, unique labels were assigned to each of the four conceptions corresponding to the theoretically hypothesized hierarchy of the four conceptions [energy as forms and sources (1), energy transfer and transformation (2), energy degradation (3), and energy conservation (4)]. Analysis of items as a function of the four conceptions found a relationship between difficulty and conception level, $\mathbf{t}$ ¼ 0.39, $p < 0.001$. Comparison of the mean item difficulties using a Welch two sample t-test revealed that ‘‘energy forms and sources’’ items, on average, were easier than ‘‘energy transfer and transformation’’ items, $t(52.61)$ ¼ -3.79; $p < 0.001$, and that ‘‘energy conservation’’ items, on average, were more difficult than ‘‘energy degradation’’ items, $t(44.89)$ ¼ -2.35, $p < 0.05$. No significant differences were observed between mean item

Table 1
Item difficulty by conception

Conception	No. of Items

		Difficulty	
		Mean	SD
Forms	31	−0.53	0.70
Transformation	24	0.14	0.59
Dissipation	20	0.06	0.47
Conservation	27	0.44	0.61

Table 2
Item difficulty by conception and for each of the four conceptions by complexity

Conception	Complexity	No. of Items	Difficulty	
			Mean	SD
Forms	Facts	8	−0.83	0.72
	Mappings	7	−0.28	0.71
	Relations	8	−0.54	0.66
	Concept	8	−0.42	0.74
Transformation	Facts	6	0.03	0.64
	Mappings	7	0.18	0.65
	Relations	6	0.10	0.77
	Concept	5	0.25	0.26
Dissipation	Facts	5	−0.11	0.50
	Mappings	4	−0.14	0.35
	Relations	5	0.13	0.59
	Concept	6	0.29	0.41
Conservation	Facts	5	0.53	0.44
	Mappings	6	0.47	0.24
	Relations	8	0.43	0.78
	Concept	8	0.36	0.79

difficulty of ‘‘energy transfer and transformation’’ and ‘‘energy degradation’’ items, t(41.96) ¼ 0.46, p ¼ 0.65. The difference between ‘‘energy transfer and transformation’’ and ‘‘energy conservation’’ items was found to be significant at the level of a ¼ 0.10, t(48.66) ¼ −1.77, p ¼ 0.08.

Progression of Difficulty by Complexity. In addition to classifying items as a function of the four conceptions of energy, for each of the four conceptions of energy, items were also classified as a function of complexity. Table 2 shows the average item difficulties and standard deviations for each level of complexity for each of the four conceptions. We hypothesized that for a given conception, items of higher complexity would be more difficult than items of lower complexity. The descriptive statistics provided in Table 2, however, suggest that for none of the four conceptions is the average item difficulty a linear function of item complexity. For ‘‘forms and sources,’’ average difficulty of items of the complexities ‘‘relations’’ and ‘‘concept’’ is lower the average difficulty of items of complexity ‘‘mapping.’’ And for ‘‘conservation,’’ the average difficulty of items of complexities ‘‘mappings’’ and above is lower than of items of complexity ‘‘facts.’’

In order to investigate whether for each of the four conceptions item difficulty depends on item complexity, a two-way ANOVA with the factors ‘‘conception’’ and ‘‘complexity’’ was computed. As suggested from our earlier findings, a significant effect of conception on item difficulty was identified, F(3, 86) ¼ 11.43, p < 0.001. However, no statistically significant effect of item complexity on item difficulty was observed across all conceptions, F(3, 86) ¼ 0.93, p ¼ 0.43. Moreover, the interaction of ‘‘conception’’ and ‘‘complexity’’ was not found to have a statistically significant effect on item difficulty, F(9, 86) ¼ 0.39, p ¼ 0.94, suggesting that for none of the four conceptions item difficulty did depended on item complexity. In parallel with our exploration of analyzing item difficulty as a function of conception, we computed the correlation between item difficulty and item complexity for each of the four conceptions. For purposes of this analysis, each complexity level was assigned a unique number [facts (1), mappings (2), relations (3), concept (4)]. None of the four correlations was

found to be statistically significant, t ¼ 0.135, p ¼ 0.34 (forms and sources), t ¼ 0.037, p ¼ 0.84 (transfer and transformation), t ¼ 0.267, p ¼ 0.14 (degradation), and t ¼ 0.072, p ¼ 0.65 (conservation).

Progression of Ability by Grade. The evaluation of item difficulty by conception and complexity suggested that indeed items regarding higher conception levels of energy are more difficult than items concerning lower energy conception levels. With respect to validation of the learning progression this means that indeed more able students tend to correctly answer items of higher conceptions. However, that does not necessarily mean that more able students are only those students of higher grades. It might as well be that students' ability is equally distributed across grades. To investigate how students' ability measures might be related to student grade, a figure similar in some respects to a Wright map was created (cf. Figure 4). The upper half of this map shows the probability density functions for three normal distributions, one for each Grade of 6, 8, and 10. These functions were obtained by fitting a normal distribution to students' ability measures for each of the three grades. The lower half of the map shows the distribution of item difficulty as a box plot for each of the four conceptions. The figure suggests a trend toward higher person ability with higher grade level. An ANOVA with the grade as the independent variable and the person ability

Figure 4. Person ability by grade compared to items by conception.

measures as the dependent variable was used to evaluate these data. Results suggest a significant effect of grade on person ability, F(2, 1853) ¼ 161.29, p < 0.001, h^2 ¼ 0.15. Analysis of the correlation between grade and person ability measures moreover suggests that students of higher grades (in average) tend to solve items of higher conceptions at a higher success rate than students of lower grades, t ¼ 0.293, p < 0.001.

An analysis of Figure 4 reveals a number of interesting trends: Students of Grade 6 tend to typically solve items regarding energy forms and sources. The average Grade 6 student solved about one third of the energy forms and sources items, and also some of the least difficult items regarding energy transformation and degradation. Some of the more advanced Grade 6 students solved more difficult items of higher conceptions. Grade 8 students typically solve items regarding energy forms and sources of higher difficulty than those energy and sources items solved by most Grade 6 students. Grade 8 students also seem to be able to solve the less difficult items of energy transformation and degradation. However, these students typically remain mostly at the form and sources level. Items of energy conservation are solved by some of the Grade 8 students, but only the most able. From Grades 8 to 10 students again move toward more advanced levels of the continuum. Almost all students of Grade 10 master the energy forms and sources level, and above average students also master the transformation and degradation levels. However, items regarding the principle of energy conservation appear to be mastered by only the most able 10th grade students. Finally, it needs to be noted that the range of student ability widens from Grades 6 to 10.

Discussion

The purpose of this study was to work towards a learning progression for the energy concept. For this purpose we developed a new measurement instrument, the ECA, and used it to collect data upon students' progression in understanding the energy concept based on an initial learning progression. In the following, we will discuss the functioning of the instrument and how the findings on students' progression in understanding energy matched the initial learning progression.

Instrument Functioning

Analysis of item fit statistics revealed that the instrument fits the requirements of the Rasch model. The low number of misfitting items was likely due to detailed steps taken to author items, pilot testing and revision of items. Summary statistics investigated were well in line with values for instrument development of the type and sample utilized for this study. The WLE person separation reliability was lower than expected. This had in part to do with the number of items completed by respondents. Analysis of the Wright map suggested that the items, in general, were more difficult than the sample was able (in average), which was intended by design. The lack of items that could differentiate between students at the lower end of the trait and the fact that students of Grade 6 received a majority of items more difficult than their average ability (resulting in the items being less accurate), is a reason for the less than expected reliability. For further studies it is recommended to create test booklets with a larger number of items in the range of Grade 6 students' average ability. The instrument in its current state may not sufficiently differentiate students within classroom. The aim of this study was to develop a robust measurement for validation of an initial learning progression (not first and foremost to develop an instrument to differentiate between different students within tested classes). Validation of the learning progression with respect to students' ability measures only included investigation of students' average ability as a function of

students' grade. For this purpose, the reliability of the instrument may be considered sufficient. Thus, for the intended purpose our instrument provides useful and informative measures.

Learning Progression Validity

With respect to the validation of the hypothesized learning progression on energy, we found the difficulty of the ECA items to be higher when the items required that a higher level of conceptions of energy to be applied. This suggests that the hierarchy of conceptions indeed defines a progression in students understanding of the energy concept. However, whereas item difficulty was found to increase with higher level conceptions of energy, no statistical connection between the complexity of the knowledge required to solve an item and item difficulty could be identified. Regarding students' ability, we found their ability increased with school grade.

These results suggest that students may first develop an understanding of energy forms and sources, and only later understand transformation and degradation. Even later students develop an understanding of energy conservation. These results align with earlier research on students' growth in understanding the energy concept (Dawson-Tunik, 2006; Lee & Liu, 2010; Liu & McKeough, 2005). Interestingly, a particular variance of item difficulty within one conception and thus a particular overlap of item difficulty between items from adjacent conceptions were observed. The variance in item difficulty may be to due to the well-known effect of item context (e.g., Chi, Feltovich, & Glaser, 1981). That is, students being able to correctly answer an item on, for example, energy transformation in one context, while failing to identify the correct option in another context. This effect of item context might explain why students have often been found to have mastered an understanding of energy transformation quite early (e.g., Duit, 1981a), while other studies suggested students struggle with the concept of energy transformation (e.g., Dawson-Tunik, 2006).

Our findings do still align with the idea of the conceptions representing levels in mastering understanding of the energy concept. Namely in the sense, that those students that are able to correctly answer the majority of items related to one conception may be assumed to have mastered this particular conception. The particular overlap in item difficulty between adjacent levels suggests that the conceptions do not create distinct stages in mastering the energy concept in the sense that only if students have mastered one level, can they learn more advanced levels. Instead it seems students begin to develop an understanding of energy transformation, once they know about energy coming in different forms, still not possessing an in-depth understanding of energy forms and sources.

It is also worth to note that in our study students seem to develop an understanding of energy transfer and transformation in parallel with an understanding of energy degradation. One reason for this observation may lie in the way items utilized to assess students' understanding of dissipation were designed. These items required students to understand that in every transformation process some energy is rendered no longer useable, and what the mechanisms of this degradation of energy would be. The items did not include knowledge about or understanding of the idea of entropy. Thus our findings do not directly contradict earlier work which suggested that an in-depth understanding of energy degradation follows an understanding of energy conversation (e.g., Solomon, 1982). Instead our work adds to existing research suggesting that a qualitative understanding of energy degradation will be obtained at the same as an understanding of energy transformation, whereas an in-depth understanding of the energy degradation process may require a more elaborated understanding including the idea of energy conservation (e.g., Shultz & Coddington, 1981). Our results suggest an understanding

of energy conservation, however, seems to be hard to obtain for students until the end of middle school (Figure 4). This aligns with earlier research (e.g., Driver & Warrington, 1985; Duit, 1984).

Regarding complexity, our findings do not confirm that students—as hypothesized—develop understanding of the individual conceptions of energy by obtaining an increasingly complex knowledge base about these conceptions. Several reasons for these findings are possible. The findings may be due to the particular implementation of different complexity levels in items: all items were originally designed to detect conceptual understanding, hypothesized to corresponded to a particular complex knowledge base. In order to create items solvable for students with less complex knowledge, additional information was added to the original items. However, the information missing in a students' knowledge base might differ from student to student. Moreover, students would have to process the information provided during testing. This might have increased the cognitive load of students (for details see, e.g., Schnotz & Kürschner, 2007). The end result being that our items might not have been suitable for an assessment of different levels of complexity of students' knowledge base. Another possible reason for the missing effect of complexity on item difficulty is that it may be that multiple choice items are not the optimal technique of assessing the complexity of students' knowledge base. The work of Lee and Liu (2010) suggests that open-ended items may provide better measures of the complexity of students' knowledge. However, there also exists research in which multiple-choice items were successfully used to reliably and validly assess the complexity of students' knowledge base (e.g., Bernholt & Parchmann, 2011; Kauertz & Fischer, 2006). Thus we are confident, that multiple-choice items that assess the individual levels of complexity of students' knowledge can be created. And, that these items can be utilized to confirm reveal an effect of item complexity on item difficulty; supporting our original hypothesis.

The assumption underlying our learning progression and the ECA that students' develop an understanding of the individual conceptions of energy by obtaining an increasingly complex knowledge base about each conception may be wrong; in particular, as other approaches such as those of Dawson-Tunik as well as Lee and Liu (2010) suggest that students obtaining an increasingly complex knowledge base underlies students' progression in understanding the energy concept as a whole. If it is the case that students do not develop an understanding of the individual conceptions through developing an increasingly complex knowledge base, the question is: What is the mechanism underlying students' progression from one level of the learning progression (e.g., ‘‘energy forms and sources'') to the next (e.g., ‘‘energy transfer and transformation'')? Understanding this mechanism is of utmost importance for the informed design of curriculum materials and instructional components to foster students' progression through the learning progression. Given the particular variance on each level of conception we suggest that the description of how students develop an understanding of individual conceptions should build on the idea that mastering a particular level of understanding the energy concept (e.g., ‘‘energy forms and sources'') relates to students being able to describe scenarios in a greater variety of contexts (e.g., identify energy forms in more contexts). This conceptualization means that: First—students would progress in their understanding of energy by learning about new forms of energy (in addition to the ones they already know) and to use these forms in a variety of contexts. Second—students would learn about transformation processes between these new forms and the forms of energy students know about already. Third—students would learn about degradation within the transformation processes. Fourth—students would learn that the overall amount of energy remains constant for all of these transformation processes. This view aligns with the knowledge integration perspective suggested by Lee and Liu (2010). If students develop an understanding of energy in the described way, then we should be able to observe a smaller impact of item context on item difficulty for the more able students (i.e., students from higher grades). In this study we did not systematically differentiate between energy forms in contexts. Therefore, this hypothesis should be investigated in future studies.

With respect to the analysis of students' ability measures across grades, the results suggest a particular heterogeneity of students' understanding with respect to the energy concept, and a differential development of students. Approximately half of the students in Grade 6 have developed an initial understanding of energy transformations, but even in Grade 10 some students have not reached such a level of understanding. Only about one-third of the students in Grade 10 were found to have developed an understanding of energy conservation. This trend confirms previous research on students' understanding of the energy concept. Students exhibit difficulties in developing more elaborate conceptions of energy such as understanding energy dissipation or conservation (e.g., Driver & Warrington, 1985); and a remarkable number of students have not developed an understanding of energy by the end of Grade 10 (Duit, 1981b).

Conclusion

Our study's findings provide important information that can be used to further inform the detailing of an energy learning progression. We were able to confirm a general progression with respect to the levels described by four conceptions of energy (forms and sources, transfer and transformation, dissipation, conservation), but we could not confirm that these conceptions create distinct levels. Our findings in particular suggest that students develop an understanding of ‘‘energy degradation'' along with an understanding of ‘‘energy transfer and transformation.'' This suggests that while most researchers expect students to only obtain an understanding of degradation after having developed an understanding of conservation, a first qualitative understanding of degradation may be obtained earlier by students. As a consequence we suggest to rename the level ‘‘energy degradation'' in our learning progression to ‘‘energy dissipation.'' We moreover suggest to introduce a new level of understanding named ‘‘energy devaluation.'' This is the level of concept of energy understanding which follows ‘‘energy conservation.’’ The difference is that ‘‘energy dissipation’’ specifically relates to the view of students understanding that in each transformation or transfer process some of the energy is transformed into thermal energy and knowing the underlying mechanism (e.g., friction). ‘‘Energy devaluation’’ would relate to students understanding of the characteristics of the process of energy degradation itself (including understanding the idea of entropy). Despite the fact that our findings suggest that students may develop an understanding of transformation and degradation at the same time, we suggest to continue viewing these two levels as distinct and separate entities. Finally, in line with other research on students' understanding of the energy concept, our findings also suggest that only a smaller part of the students develops an understanding of ‘‘energy conservation’’—and only at the end of Grade 10.

Our findings may suggest that within an energy curriculum, initial teaching should focus on developing an understanding of energy with respect to forms and sources first. Then, the concept of transfer and transformation should be covered, before introducing energy dissipation and conservation. However, as detailed earlier, students seem to develop an understanding of energy transfer and transformation while still not having fully developed an understanding of energy forms and sources. Therefore, we suggest that it is not wise for all possible forms (and sources) of energy to be covered in the curriculum before the concept of

energy transfer and transformation is introduced. Instead we suggest to ensure that students obtain a thorough understanding (1) of what an energy form is and (2) that different energy forms exist—before moving on to the transformation of one energy form into the other (cf. Nordine et al., 2010). More research is needed with respect to how a revised curriculum would impact students learning and how the curriculum may be further optimized with respect to students developing an understanding of energy conservation and energy degradation.

We could not determine that students develop an understanding of individual conceptions by obtaining an increasingly complex knowledge base. Further research is necessary in this area. One area of further research would be to investigate to what extent an operationalization of each level of complexity into specific items (instead of varying complexity through the provision of additional information) will allow for a more precise assessment of the complexity of students' knowledge base regarding the individual conceptions of energy. Another area of further research would concern (1) the exploration of the effect of context and (2) the subsequent revision of the learning progression. However, while more research is needed to further detail an energy learning progression, some important implications for the teaching of energy and design of an energy curriculum can be identified from our findings.

In summary, our findings align with several researchers' findings about students' progression in understanding the energy concept (in particular with respect to the four conceptions of energy) who used different data methods and samples. We acknowledge that other approaches such as using newly developed Ordered Multiple Choice (OMC) items (Briggs, Alonzo, Schwab, & Wilson, 2006) and/or techniques such as concept mapping—maybe valuable tools for assessing the complexity of students' knowledge base.

We would like to thank the journal reviewers and editors for their detailed and helpful recommendations. We also thank Xiufeng Liu, Hendrik Härtig, and Ross Nehm for their helpful manuscript comments.

References

Adams, R. (Ed.). (2002). PISA 2000 technical report. Paris: OECD.

Aebli, H. (1980). Denken: Das Ordnen des Tuns (Vol. 1). Stuttgart: Klett-Cotta.

Alao, S., & Guthrie, J. T. (1999). Predicting conceptual understanding with cognitive and motivational variables. The Journal of Educational Research, 92, 243–254.

American Association for the Advancement of Science. (2001). Atlas of science literacy. New York: Oxford University.

American Association for the Advancement of Science. (2007). Getting Assessment Right. 2061 Today, 17(1), 2–7.

Berland, L. K., & McNeill, K. L. (2010). A learning progression for scientific argumentation: Understanding student work and designing supportive instructional contexts. Science Education, 94(5), 765–793. DOI: 10.1002/sce.20402

Bernholt, S., & Parchmann, I. (2011). Assessing the complexity of students' knowledge in chemistry. Chemistry Education Research and Practice, 12(2), 167. DOI: 10.1039/c1rp90021h

Black, P., & Solomon, J. (1983). Life world and science world: Pupils' ideas about energy. In G. Marx (Ed.), Entropy in the school. Proceedings of the 6th Danube Seminar on Physics Education (pp. 43–455). Budapest: Roland Eoetvoes Physical Society.

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

Boone, W. J., & Scantlebury, K. (2006). The role of Rasch analysis when conducting science education research utilizing multiple-choice tests. Science Education, 90(2), 253–269. DOI: 10.1002/sce.20106

Boyes, E., & Stanisstreet, M. (1990). Misunderstandings of ''law'' and ''conservation'': A study of pupils' meanings for these terms. School Science Review, 72(258), 51–57.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). How people learn: Brain, mind, experience and school. Washington, DC: National Academy Press.

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

Bruner, J. S. (1966). Toward a theory of instruction. London: Belknap Press of Harvard University Press.

Case, R. (1985). Intellectual development: Birth to adulthood. Orlando: Academic Press.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. Pittsburgh, PA: Learning Research and Development Center, University of Pittsburgh.

Commons, M. L. (2007). Introduction to the model of hierarchical complexity. Behavioral Development Bulletin, 13, 1–4.

Dawson-Tunik, T. L. (2006). Stage-like patterns in the development of conceptions of energy. In X. Liu & W. J. Boone (Eds.), Applications of Rasch measurement in science education (pp. 111–136). Maple Grove, USA: JAM Press.

Doménech, J. L., Gil-Pérez, D., Gras-Martí, A., Guisasola, J., Martínez-Torregrosa, J., Salinas, J., . . . Vilches, A. (2007). Teaching of energy issues: A debate proposal for a global reorientation. Science & Education, 16(1), 43–64. DOI: 10.1007/s11191-005-5036-3

Driver R., & Millar R. (Eds.). (1986). Energy matters: Proceedings of an invited conference: Teaching about energy within the secondary science curriculum. Leeds [England]: University of Leeds, Centre for Studies in Science and Mathematics Education.

Driver, R., Squires, D., Rushworth, P., & Wood-Robinson, V. (1994). Making sense of secondary science: Supporting materials for teachers. London: Routledge.

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

Duit, R. (1981a). Students' notions about the energy concept - before and after physics instruction. In W. Jung, H. Pfundt, & C. von Rhoeneck (Eds.), Proceedings of the international workshop on 'Problems Concerning Students' Representation of Physics and Chemistry Knowledge (pp. 268–319). Ludwigsburg: Paedagogische Hochschule.

Duit, R. (1981b). Understanding energy as a conserved quantity—remarks on the article by R.U. Sexl. International Journal of Science Education, 3(3), 291–301. DOI: 10.1080/0140528810030306

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

Duit, R. (2007). Energie: Ein zentraler Begriff der Naturwissenschaften und des naturwissenschaftlichen Unterrichts. Unterricht Physik, pp. 4–7.

Duit R. (1986). Der Energiebegriff im Physikunterricht. Kiel: IPN.

Duncan, R. G., & Hmelo-Silver, C. E. (2009). Learning progressions: Aligning curriculum, instruction, and assessment. Journal of Research in Science Teaching, 46(6), 606–609. DOI: 10.1002/tea.20316

Duschl, R. A., Schweingruber, H. A., & Shouse, A. W. (2007). Taking science to school: Learning and teaching science in grades K-8. Washington, DC: National Academies Press.

Erickson, G. L., & Tiberghien, A. (1985). Heat and temperature. In R. Driver, E. Guesne, & A. Tiberghien (Eds.), Children‘s ideas in science. Milton Keynes: Open University Press.

Feynman, R. P. (2011). The Feynman lectures on physics (New millennium ed.). New York, London: BasicBooks; Perseus Running [distributor].

Field, A. (2009). Discovering statistics using SPSS (3rd ed.). London: Sage.

Fischer, K. W. (1980). A theory of cognitive development: The control and construction of hierachies of skills. Psychological Review, 87(6), 477–531.

Gagne, R. M., & White, R. T. (1978). Memory structures and learning outcomes. Review of Educational Research, 48(2), 187–222.

Goldring, H., & Osborne, J. (1994). Students' difficulties with energy and related concepts. Physics Education, 29(1), 26–32.

Kauertz, A., & Fischer, H. E. (2006). Assessing students' level of knowledge and analysing the reasons for learning difficulties in physics by Rasch analysis. In X. Liu & W. J. Boone (Eds.), Applications of Rasch measurement in science education (pp. 212–246). Maple Grove, USA: JAM Press.

Kline, P. (2000). The handbook of psychological testing (2nd ed.). London, New York: Routledge.

Kolen, M. J., & Brennan, R. L. (2004). Test equating, scaling, and linking: Methods and practices (2nd ed.). New York: Springer.

Krajcik, J., Drago, K., Sutherland, L. A., & Merritt, J., (2012). The promise and value of learning progression research. In S. Bernholt, P. Nentwig, & K. Neumann, (Eds.), Making it tangible—Learning outcomes in science education. Münster: Waxmann.

Kultusministerium des Landes Nordrhein-Westfalen. (1993). Richtlinien und Lehrpläne. Physik. Gymnasium. Frechen: Ritterbach.

Lee H-.S. & Liu, O. L. (2010). Assessing learning progression of energy concepts across middle school grades: The knowledge integration perspective. Science Education, 94(4), 665–688. DOI: 10.1002/sce.20382

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

Liu, O. L., Lee, H.-S., Hofstetter, C., & Linn, M. (2008). Assessing knowledge integration in science: Construct, measures, and evidence. Educational Assessment, 13(1), 33–55. DOI: 10.1080/10627190801968224

Liu, X. (2010). Rasch measurement in science education. Charlotte, NC: Information Age Publishing.

Liu X., & Boone W. J. (Eds.). (2006). Applications of Rasch measurement in science education. Maple Grove, USA: JAM Press.

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. DOI: 10.1002/tea.20060

Liu, X., & Ruiz, M. E. (2008). Using data mining to predict K-12 students' performance on large-scale assessment items related to energy. Journal of Research in Science Teaching, 45(5), 554–573. DOI: 10.1002/tea.20232

Mohan, L., Chen, J., & Anderson, C. W. (2009). Developing a multi-year learning progression for carbon cycling in socio-ecological systems. Journal of Research in Science Teaching, 46(6), 675–698. DOI: 10.1002/tea.20314

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

National Research Council. (2012). A framework for K-12 science education: Practices, crosscutting concepts, and core ideas. Washington, DC: The National Academies Press.

Neumann, I., Neumann, K., & Nehm, R. (2011). Evaluating instrument quality in science education: Rasch-based analyses of a nature of science test. International Journal of Science Education, 10(33), 1–33. DOI: 10.1080/09500693.2010.511297

Neumann, K., Viering, T., & Fischer, H. E., (2010). Die Entwicklung physikalischer Kompetenz am Beispiel des Energiekonzepts.

Nordine, J., Krajcik, J., & Fortus, D. (2010). Transforming energy instruction in middle school to support integrated understanding and future learning. Science Education, 95(4), 670–690. DOI: 10.1002/sce.20423

Piaget, J. (1972). Intellectual evolution from adolescence to adulthood. Human Development, 15, 1–12.

Rasch, G. (1960). Probabilistic models for some intelligence and attainment tests. Copenhagen: Danish Institute for Educational Research.

Roberts, D. (2007). Scientific literacy/science literacy. In S. Abell & N. G. Lederman (Eds.), Handbook of research on science education (pp. 729–780). Mahwah, NJ: Lawrence Erlbaum Associates.

Schnotz, W., & Kürschner, C. (2007). A reconsideration of cognitive load theory. Educational Psychology Review, 19(4), 469–508. DOI: 10.1007/s10648-007-9053-4

Schwarz, C. V., Reiser, B. J., Davis, E. A., Kenyon, L., Achér, A., Fortus, D., . . . (2009). Developing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners. Journal of Research in Science Teaching, 46(6), 632–654. DOI: 10.1002/tea.20311

Sekretariatder Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK]. (2005). Bildungsstandards im Fach Physik für den Mittleren Schulabschluss. Beschluss vom 16.12.2004. München: Luchterhand.

Shultz, T. R., & Coddington, M. (1981). Development of the concepts of energy conservation and entropy. Journal of Experimental Child Psychology, 31(1), 131–153.

Smith, C., Wiser, M., Anderson, C. W., & Krajcik, J. (2006). Implications of research on children's learning for assessment: Matter and atomic molecular theory. Measurement, 14(1&2), 1–98.

Solomon, J. (1982). How children learn about energy or does the first law come first ? The School Science Review, 63(224), 415–422.

Solomon, J. (1983a). Learning about energy: How pupils think in two domains. International Journal of Science Education, 5(1), 49–59. DOI: 10.1080/0140528830050105

Solomon, J. (1983b). Messy, contradictory and obstinately persistent: A study of children's out-of-school ideas about energy. School Science Review, 65(231), 225–230.

Solomon, J. (1986). Energy for the citizen. In R. Driver & R. Millar (Eds.), Energy matters. Proceedings of an invited conference: Teaching about energy within the secondary science curriculum (pp. 25–31). Leeds [England]: University of Leeds, Centre for Studies in Science and Mathematics Education.

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. DOI: 10.1002/tea.20313

Stevens, S. Y., Delgado, C., & Krajcik, J. S. (2010). Developing a hypothetical multi-dimensional learning progression for the nature of matter. Journal of Research in Science Teaching, 47(6), 687–715. DOI: 10.1002/tea.20324

Trumper, R. (1993). Children's energy concepts: A cross-age study. International Journal of Science Education, 15(2), 139–148.

Viennot, L. (1979). Spontaneous reasoning in elementary dynamics. European Journal of Science Education, 1(2), 205–221.

Viering, T., Fischer, H. E., & Neumann, K. (2010). Die Entwicklung physikalischer Kompetenz in der Sekundarstufe I. In E. Klieme (Ed.), Kompetenzmodellierung. Zwischenbilanz des DFG-Schwerpunktprogramms und Perspektiven des Forschungsansatzes. 56. Beiheft der Zeitschrift für Pädagogik, Heft 2/2010 (pp. 92–103). Weinheim [u.a.]: Beltz.

Vosniadou S. (Ed.). (2008). International handbook of research on conceptual change. New York: Routledge.

Watts, D. M. (1983a). A study of schoolchildren's alternative frameworks of the concept of force. International Journal of Science Education, 5(2), 217–230. DOI: 10.1080/0140528830050209

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

Watts, M., & Gilbert, J. K. (1983). Enigmas in school science: Students‘ conceptions for scientifically associated words. Research in Science and Technological Education, 1(2), 161–171.

Wilson, M. R., & Bertenthal, M. W. (2006). Systems for state science assessment. Washington, DC: National Academies Press.

Wilson, M. (2009). Measuring progressions: Assessment structures underlying a learning progression. Journal of Research in Science Teaching, 46(6), 716–730. DOI:10.1002/tea.20318.

Wright, B. D., & Stone, M. H. (1979). Best Test Design. Chicago: MESA Press/University of Chicago.

Wu, M. L., Adams, R. J., & Wilson, M. R. (2007). ACER ConQuest version 2.0: generalised item response modelling software. Camberwell: ACER Press.