Using Rasch Modeling and Option Probability Curves to Diagnose Students' Misconceptions

Cari F. Herrmann-Abell and George E. DeBoer

AAAS Project 2061

Paper presented at the
2016 AERA Annual Meeting
Washington, DC

April 8-12, 2016

Abstract

Understanding students' misconceptions and how they change is an essential part of supporting students in their science learning. This paper presents results from distractor-driven multiple-choice assessments that target students' misconceptions about energy. Over 20,000 elementary, middle and high school students from across the U.S. participated in the study. Rasch modeling was used to estimate item and student measures. Option probability curves were used to represent the distribution of correct answers and misconceptions across the range of student achievement levels. Analysis of the shapes of the curves and where they occur as a function of student achievement provides insights on changes in students' thinking as they learn more about a topic. These data can then be used to inform instruction.

Introduction

Students experience an array of natural phenomena through their life experiences, and before receiving instruction in science, they often construct their own ideas to explain these phenomena. While these ideas may not always be scientifically sound, they appear to the students to be correct because they sufficiently explain the student's world. For example, in a study by Chinn and Brewer (1998), students were presented with data that were incompatible with their misconceptions, but out of eight possible responses by students, only two of them involved a willingness to relinquish the misconceptions in favor of more scientific ideas. The persistence of these misconceptions is also evident in the fact that high school students often have similar misconceptions to middle school students (AAAS, 2015).

Researchers have documented the prevalence among many students of misconceptions and naive ideas about a variety of science topics (see for example Driver, Squires, Rushworth, & Wood-Robinson, 1994). The topic of energy is no exception, and because energy concepts are highly abstract and often counterintuitive, misconceptions about energy can be particularly persistent and difficult to challenge. For example, some students think that objects at rest have no energy because students associate energy only with obvious activity or movement (e.g., Brook & Driver, 1984). Students also think that living things have thermal energy but inanimate objects do not (e.g., Leggett, 2003) and, therefore, they may think that it is "coldness," not thermal energy, that is transferred between two inanimate objects (e.g., Newell & Ross, 1996).

Research has shown that students of teachers who have both knowledge of the science content and knowledge of misconceptions have higher learning gains than students of teachers who had only science content knowledge (Sadler, Sonnert, Coyle, Cook-Smith, & Miller, 2013). Therefore, it is imperative to identify tools that can help teachers become aware of common misconceptions and how those misconceptions change as students gain an understanding of the science. With this information in hand, teachers are better able to craft instruction that responds to where students are in their learning and that addresses their misconceptions more effectively.

Distractor-driven, multiple-choice assessment items in which the distractors are aligned to specific misconceptions provide one example of such a tool. This paper describes how student response data from these assessments can be analyzed to diagnose students' misconceptions and how Rasch modeling and option probability curves can be used to generate detailed visual representations of how students' thinking changes.

Methodology

The data presented here resulted from the field testing of assessment items aligned to energy concepts. The assessment items are being developed as part of a larger project to construct three vertically-equated instruments to measure students' understanding of energy from fourth to twelfth grade.

Learning Goals

Table 1 lists the energy concepts that were targeted by the field test items. These concepts were derived from several science standards documents including *Benchmarks for Science Literacy* (AAAS, 1993) and the *Next Generation Science Standards* (NGSS Lead States, 2013). To guide the development of the items and to articulate the progression of understanding being theorized for each concept, clarification statements were written to make explicit the boundaries around the targeted knowledge and to spell out the different expectations for students at each grade level.

Table 1
Energy Ideas Targeted by the Field Test Items

Ideas about the Forms of Energy	Ideas about Energy Transfer	Other Energy Ideas
Kinetic Energy	Conduction	Energy Conservation
Thermal Energy	Convection	Energy Dissipation & Degradation
Gravitational Potential Energy	Radiation	
Elastic Potential Energy	Transferring Energy by Forces	
Chemical Energy	Transferring Energy Electrically	
Energy Transformations	Transferring Energy by Sound	

Field Tests

A total of 372 distractor-driven, multiple-choice items were field tested with students from across the United States in May and June of 2015. Item construction followed rigorous item development procedures that included (1) the identification of documented misconceptions, which were then incorporated into distractors; (2) a careful evaluation of the items' alignment to the targeted ideas about energy; and (3) a close examination of the items for their overall psychometric effectiveness (DeBoer, Herrmann-Abell, & Gogos, 2007; DeBoer, Herrmann-Abell, Michiels, Regan, & Wilson, 2008; DeBoer, Lee, & Husic, 2008). The items were divided into 25 different test forms, ten for elementary students (Grades 4 and 5) and 15 for secondary students (Grades 6 through 12). There were 23 or 24 items on the field test forms for elementary school students and 31 or 32 items on the field test forms for the secondary school students. Linking items were used so that item characteristics could be compared across forms. On average, about 1,600 students responded to each item. The field tests were administered in both online and paper and pencil format. For each item, the students were asked to choose the one correct answer; students who chose more than one answer choice were marked incorrect.

Participants

A total of 21,061 students from 42 different states and Puerto Rico participated in the field test. This study included only the 20,870 students who responded to six or more items. Students with highly unexpected responses were also excluded as described below. Table 2 shows the demographic information for the 20,551 students included in this study. Elementary school students (grades 4 and 5) made up 14% of the sample. Middle school students (grades 6 through 8) were 50% of the sample, and high school students (grades 9 through 12) were 36% of the sample. Approximately half of the students were male and half were female. About 45% of the students identified themselves as white, 20% as Hispanic, 11% as African American, 5% as Asian, and 11% identified as being of two or more ethnicities. All of the students were studying science in school at the time of field testing but not necessarily physical science.

Table 2
Demographic Information for Field Test Participants

	Elementary	Middle	High	Total
Grades	4-5	6-8	9-12	4-12
Number of Students	2967 (14%)	10207 (50%)	7377 (36%)	20551
Gender				
Male	48%	49%	46%	48%
Female	50%	48%	55%	50%
Ethnicity				
White	38%	48%	44%	45%
Asian	7%	4%	7%	5%
Black	17%	11%	10%	11%
Hispanic	17%	19%	22%	20%
Two or more ethnicities	10%	10%	11%	11%
Primary language				
English	87%	88%	85%	87%
Other	11%	9%	13%	11%

Rasch Modeling

WINSTEPS (Linacre, 2016) was used to estimate Rasch student and item measures. In the dichotomous Rasch model, the probability that a student will respond to an item correctly is determined by the difference in the student's achievement level and the item's difficulty (Bond & Fox, 2007). The measures are expressed on the same interval scale, are measured in logits, and are mutually independent.

Option Probability Curves

Option probability curves plot the probability that students will select each answer choice as a function of their Rasch student measure. With traditional analysis of multiple-choice items, curves are often generated for correct and incorrect answers, and the results show two sigmoidal curves that cross. Because the focus is on whether or not the student selected the correct answer, all of the incorrect answer choices are lumped together. The curve corresponding to the correct answer typically increases monotonically while the curve for the set of distractors typically decreases monotonically with increasing student understanding (Haladyna, 1994). Past research has shown that the curves for distractor-driven items do not match the monotonic behavior of traditional items (Sadler, 1998; Herrmann-Abell & DeBoer, 2011; Wind & Gale, 2015). With distractor-driven multiple-choice items, therefore, it is important to look at the curves for each answer choice because the shape of the curves provides information about what types of students (in terms of their overall understanding) are more likely to select each answer choice, how persistent the misconception represented by the answer choice is, and how the popularity of the answer choice compares to other answer choices. Thus, analyzing the option probability curves for each answer choice provides additional information that is not available when the incorrect answers are analyzed in combination.

To generate the curves, we used a procedure similar to that described by Wind and Gale (2015). First, the Rasch student measures were rounded to the nearest .5 ranging from -3.0 to 3.0 logits. For each Rasch student measure value, the proportion of students selecting each answer choice was then calculated. A plot of this proportion as a function of Rasch student measure was produced for each answer choice. For some items, we determined the proportion of students selecting each answer choice by grade level to investigate whether the shape and position of the curves differed by grade level.

Findings

Rasch Fit Statistics

An initial Rasch analysis of the data revealed 10 misfitting items with outfit MNSQ values outside of the acceptable range of 0.7 to 1.3 (Bond & Fox, 2007). Starting with the item with the highest outfit MNSQ, we looked for student responses with a large Z-residual statistic indicating a highly unexpected response, perhaps due to students who frequently guessed or selected answer choices at random throughout the test. These students were flagged and their data were removed from the data set. Three hundred and nineteen misfitting students were identified and removed. The final fit analysis showed that all of the items were within the acceptable range for both infit and outfit indices. The final fit statistics, summarized in Table 3, suggest the final data set has a good fit to the Rasch model. The item standard errors were all low with a median of 0.06. The item separation index was 11.67, with a reliability of 0.99. The person separation index was 1.40, with a reliability of 0.66. The lower person separation index can be attributed to the fact that the students responded to a small percentage of the available items, about 7%. This means that there is less information available to estimate the student measures, which results in a lower reliability. In contrast, because there were so many students responding to each item, differences in difficulty level of the items is easier to determine, which can be seen in the very high item separation index and reliability estimate.

Table 3
Summary of Rasch Fit Statistics

	Item			Student		
	Min	Max	Median	Min	Max	Median
Standard error	0.02	0.11	0.06	0.37	1.93	0.40
Infit mean-square	0.84	1.27	0.99	0.44	2.17	0.99
Outfit mean-square	0.72	1.33	0.99	0.23	5.15	0.97
Point-measure correlation coefficients	0.00	0.53	0.34	-0.13	0.56	0.32
Separation index (reliability)	11.67 (0.99)			1.40 (0.66)		

Grade-to-Grade Differences

ANCOVA was used to perform a cross-sectional analysis of the students' performance by grade band controlling for gender, ethnicity, and whether or not English was their primary language. To control for differences in instructional focus across the country, we also controlled for the state students came from. The estimated marginal mean student performance was -0.54 for the elementary school students, -0.46 for the middle school students, and -0.17 for the high school students, $F(2,19789) = 395.54$, $p < .001$ (see Table 4). A Bonferroni post hoc test showed that high school students outperformed middle school students, and middle school students outperformed elementary school students. The negative means indicate that, overall, the items were relatively difficult for this sample of students regardless of their grade level.

Table 4
Estimated Marginal Student Means by Grade Band

Grade band	Mean Student Measure	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Elementary	-0.54	.014	-.67	-.51
Middle	-0.46	.008	-.47	-.44
High	-0.17	.009	-.18	-.15

Option Probability Curves

Kinetic Energy. The option probability curves shown in Figure 1 resulted from an item targeting the relationship between velocity and kinetic energy. Students were presented with a diagram of a ball rolling down a ramp and asked which of four line graphs represented the kinetic energy of the ball as it rolls down the ramp and its speed increases. The correct answer to this item requires that the students know that the kinetic energy of an object is proportional to the square of the velocity. The item was administered to students in middle and high school.

Figure 1(a) shows the curves for the middle school students and Figure 1(b) shows the curves for the high school students. Middle school students with a lower level understanding of energy (negative student measure) were most likely to select answer choice B. This answer choice included a graph that looked like the ramp in the diagram. This misconception of interpreting a graph as a picture of the event has been documented in the literature (e.g. Garcia Garcia & Cox, 2010). This answer choice was not as popular with the high school students. The high school students with a lower level of energy understanding were most likely to select answer choice A, which was the graph showing a linear relationship between kinetic energy and velocity. As students learn about energy, they may learn that the kinetic energy of an object increases with increasing velocity, but they may not be aware that the kinetic energy actually increases with the square of the velocity. The curve for this answer choice at the middle school level rises slightly across achievement levels, indicating that over time students develop the idea that velocity increases as kinetic energy increases. For both grade levels, the correct answer (C) is the most commonly selected answer by the students with higher understanding of energy (greater than 0.5 logits).

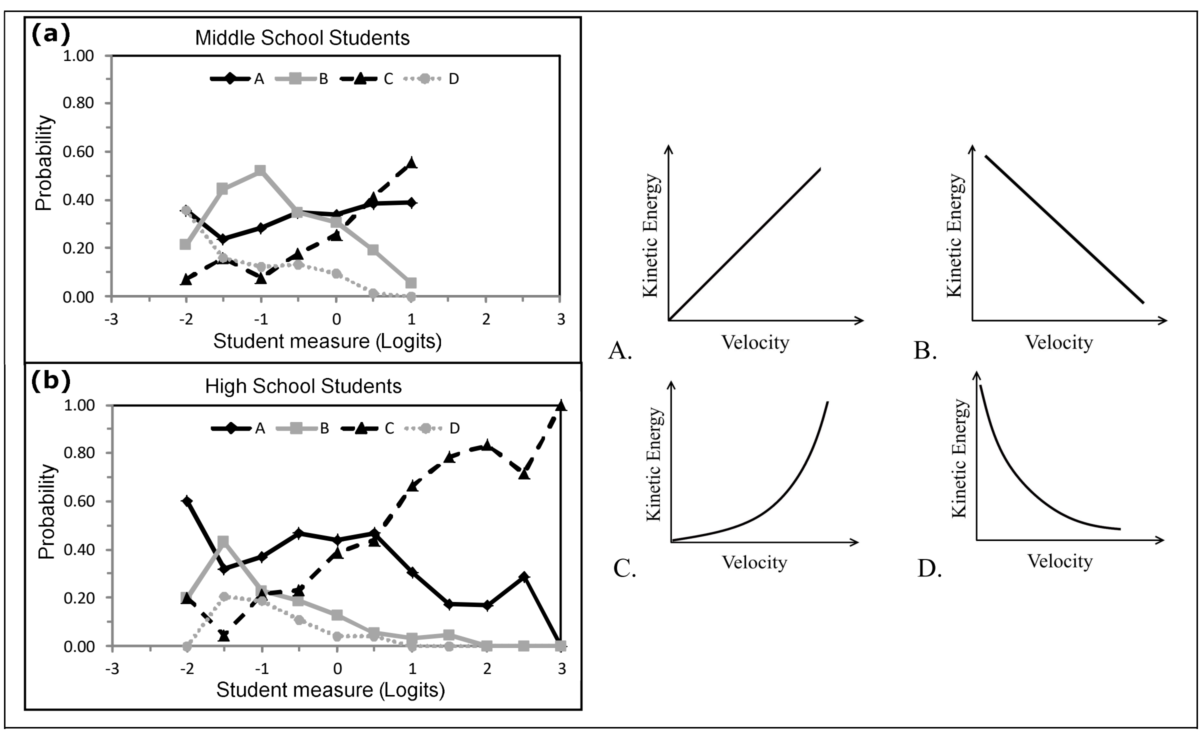
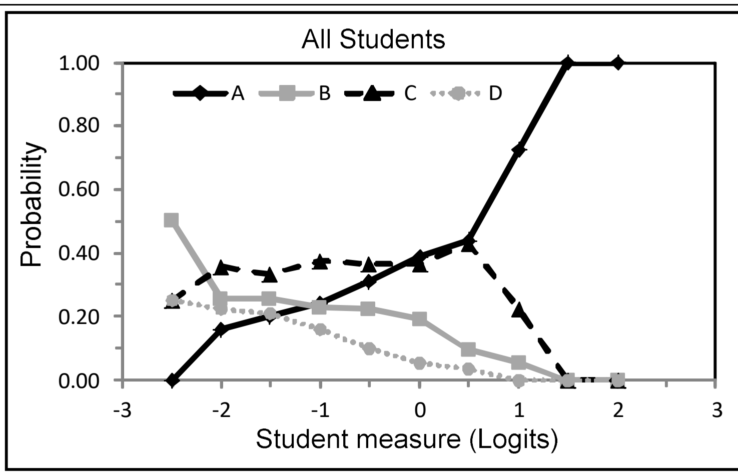


Figure 1: Option probability curves for a kinetic energy item and the graphs used in the answer choices to represent the kinetic energy of a ball rolling down a ramp.

Conduction. Figure 2 shows the option probability curves for an item aligned to the idea of transferring energy by conduction. The distractors in this item target the misconception that coldness is transferred between objects that are at different temperatures (Brook, Briggs, Bell, & Driver, 1984; Clough & Driver, 1985; Newell & Ross, 1996). Answer choice B says that coldness will be transferred from a cool counter top to a hot frying pan when the pan is placed on the counter. Answer choice C corresponds to the misconception that there is an exchange of both coldness and energy between the counter and frying pan. Answer choice D says that energy is transferred from the pan to the air around it but not from the pan to the counter.

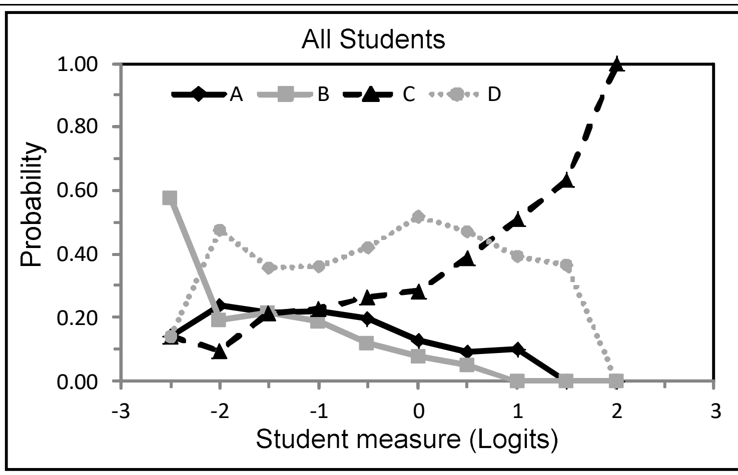
The curves show that students with the lowest level of understanding of energy think that only coldness is transferred. As students learn more about energy, they are almost equally likely to select any of the answer choices, which suggests they are unsure about what is transferred. In the mid-range, answer choices B and D decrease in probability, and students mainly choose between the correct answer (A) and answer choice C (an exchange of both energy and coldness). This is an indication that some students are transitioning from holding the misconception that only coldness is transferred to a middle ground where they think that both the misconception and the science idea are true. As students' understanding of energy improve, they begin to understand what happens during conduction and are likely to let go of the coldness misconception completely in favor of the correct science idea that energy is transferred.



A. Energy is transferred
B. Coldness is transferred
C. Energy and coldness are exchanged
D. Energy transferred from the pan to the air but not to the counter

Figure 2: Option probability curves for an item targeting the misconception that coldness is transferred between objects at different temperatures. Brief summaries of the answer choices are presented on the right.

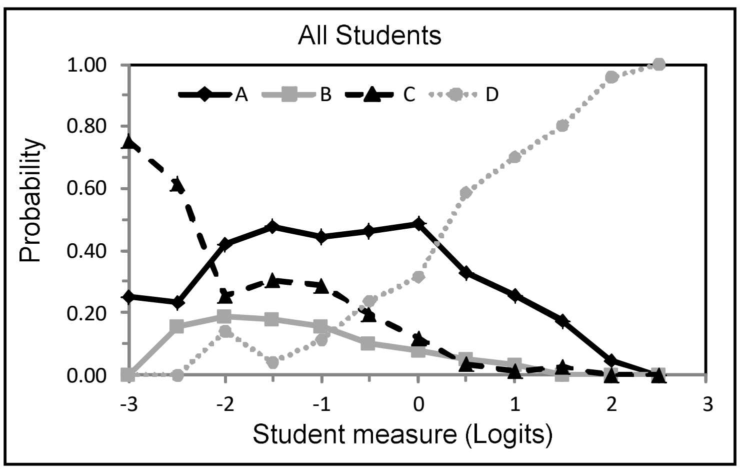
Convection. The curves for an item aligned to the convection idea is are shown in Figure 3. The item stem describes a room with a fireplace and a fan at one end. The question asks the students whether or not the air at the other end of the room will get warmer if the fan is used to blow the warmer air across the room. Answer choice A says that the air will not get warmer because fans are only used to cool a room. This answer choice is not popular with students at any achievement level. Answer choice B says that blowing the warmer air will not change the temperature of the air at the other end of the room. This misconception is only selected by students with a very low level of understanding of energy and quickly decreases as understanding of energy increases. The most popular distractor is answer choice D which says that the air will only get warmer if the air being blown is a lot warmer than the air at the other end of the room. The proportion of students selecting this answer choice is high for a wide range of achievement levels. It's possible that these students think that changes in temperature that result from the movement of slightly warmer air would not be significant, perhaps because they can usually not detect small changes in temperatures in their everyday experiences. As the student measure increases, answer choice C (the warmer air will make the air at the other end of the room warmer) becomes the most popular answer choice. These students have learned that the temperature of the other end of the room must increase even if the air blown across the room is slightly warmer.



A. Fans are only used to cool a room
B. Blowing warmer air will not change the temperature of the air across the room
C. Blowing warmer air will warm the air across the room
D. Only much warmer air will warm the air across the room

Figure 3: Option probability curves for an item about blowing warmer air across a room. Brief summaries of the answer choices are presented on the right.

Conservation of energy. One of the sub-ideas under the conservation of energy concept is the idea that all things have energy. Many studies have shown that students tend to associate energy with mainly human beings and other living things, not inanimate objects (e.g. Watts, 1983; Trumper, 1990; Liu & Tang, 2004). When students do associate energy with inanimate objects, they often only do so for objects that are moving (e.g. Brook & Driver, 1984; Kruger, 1990; Trumper, 1998). The item for which the option probability curves are shown in Figure 4 probed for these misconceptions. The context of the item includes two identical rocks. One is falling towards the ground and the other is sitting on the top of a cliff. Answer choice C corresponded to the misconception that neither rock has energy because only living things have energy. The probability of selecting this answer choice is high for students with the lowest level of energy understanding. The curve rapidly decreases with increasing energy understanding suggesting that students let go of this misconception as soon as they learn more about energy. Answer choice A corresponds to the misconception that only the falling rock has energy because it is moving. This answer choice is the most popular answer choice for students with a range of measures from -2 to 0 logits. Students with a measure greater than 0 logits were more likely to select the correct answer that both rocks have energy. Answer choice B, which said that only the rock on the cliff has energy, was not popular with students of any achievement level.



A. Only the falling rock has energy
B. Only the rock on the cliff has energy
C. Neither rock has energy
D. Both rocks have energy

Figure 4: Option probability curves for an item about the energy of two rocks, one falling and one sitting on a cliff. Brief summaries of the answer choices are shown on the right.

Significance

Using a combination of distractor-driven, multiple-choice items, Rasch modeling, and option probability curves provides valuable information about how students' knowledge and misconceptions change as they progress in their understanding of energy. This information can be used to raise teachers' awareness of misconceptions, which may help them better select and sequence appropriate instructional activities and respond to the needs of their students. For example, misconceptions that are appear as narrow peaks are not as persistent as misconceptions that spread across a wide range of overall student achievement levels. These short-lived misconceptions may not require as much class time to address as misconceptions that persist, allowing teachers to allocate instructional time more productively. Similarly, curriculum developers can leverage this information when designing and evaluating curriculum materials. For example, answer choices with curves showing a blip or hump toward the upper end of the achievement range could indicate a misunderstanding that students develop during instruction. This information would be helpful in evaluating and revising the instructional activities.

Conclusions

This paper describes how the use of Rasch modeling and option probability curves to analyze data from science assessment items can help diagnose students' misconceptions about energy and reveal how those misconceptions change as a function of students' overall energy understanding. A cross-sectional analysis of the student measures showed that the high school students have a better understanding of the energy concept than the elementary and middle school students. For most of the energy ideas, an analysis of the item measures validated the study's description of the energy concept and how it progresses from elementary to middle to high school. The shapes of the option probability curves provided insight into how students' thinking about energy changes as their Rasch measure increases. Taken together, these findings point to the potential richness and usefulness of the data that multiple-choice assessments can provide when coupled with appropriate analytical strategies.

Acknowledgements

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A120138 to the American Association for the Advancement of Science. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

References

American Association for the Advancement of Science. (2015) AAAS science assessment [website]. Retrieved from http://assessment.aaas.org.

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

Chinn, C. A. & Brewer, W. F. (1998). An empirical test of a taxonomy of response to anomalous data in science. *Journal of Research in Science Teaching, 35*(6), 623-654.

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

Brook, A., Briggs, H., Bell, B., & Driver, R. (1984). Aspects of secondary students' understanding of heat: Summary report. Leeds, UK: The University of Leeds, Centre for Studies in Science Education and Mathematics Education.

Brook, A., & Driver, R. (1984). Aspects of secondary students' understanding of energy: Full report. Leeds, UK: The University of Leeds, Centre for Studies in Science Education and Mathematics Education.

Clough, E. E. & Driver, R. (1985). Secondary students' conceptions of the conduction of heat: Bringing together scientific and personal views. *Physics Education, 20*(4), 176-182.

DeBoer, G. E., Herrmann-Abell, C. F., & Gogos, A. (2007, March-April). Assessment linked to science learning goals: Probing student thinking during item development. Paper presented at the National Association for Research in Science Teaching Annual Conference, New Orleans, LA.

DeBoer, G. E., Herrmann-Abell, C. F., Gogos, A., Michiels, A., Regan, T., & Wilson, P. (2008). Assessment linked to science learning goals: Probing student thinking through assessment. In J. Coffey, R. Douglas, & C. Stearns (Eds.), *Assessing student learning: Perspectives from research and practice* (pp. 231-252). Arlington, VA: NSTA Press.

DeBoer, G. E., Lee, H. S., & Husic, F. (2008). Assessing integrated understanding of science. In Y. Kali, M. C. Linn, & J. E. Roseman (Eds.), *Coherent science education: Implications for curriculum, instruction, and policy* (pp. 153-182). New York, NY: Columbia University Teachers College Press.

Driver, R., Squires, A., Rushworth, P., & Wood-Robinson, V. (1994). Making sense of secondary science: Research into children's ideas. New York, NY: Routledge.

Garcia Garcia, G. & Cox, R. (2010) "Graph-as-Picture" misconceptions in young students. In A.K. Goel, M. Jamnik, & N.H. Narayanan (Eds.), *Diagrammatic representation and inference* (pp. 310–312). Germany: Springer Berlin Heidelberg.

Haladyna, T. M. (1994). Developing and validating multiple-choice test items. Hillsdale, NJ: Lawrence Erlbaum Associates.

Herrmann-Abell, C. F., & DeBoer, G. E. (2011). Using distractor-driven standards-based multiple-choice items and Rasch modeling to investigate hierarchies of chemistry misconceptions and detect structural problems with individual items. *Chemistry Education Research and Practice, 12*, 184–192.

Leggett, M. (2003). Lessons that non-scientists can teach us about the concept of energy: a human-centered approach. *Physics Education, 38*(2), 130-134.

Linacre, J. M. (2016). *WINSTEPS Rasch measurement computer program*. Beaverton, Oregon: Winsteps.com.

Liu, X. & Tang, L. (2004). The progression of students' conceptions of energy: A cross-grade, cross-cultural study. *Canadian Journal of Science, Mathematics, and Technology, 4*(1), 43-57.

Kruger, C. (1990). Some primary teachers' ideas about energy. *Physics Education, 25*(2), 86-91.

Newell, A., & Ross, K. (1996). Children's conception of thermal conduction – Or the story of a woolen hat. *School Science Review, 78*(282), 33–38.

NGSS Lead States. (2013). *Next generation science standards: For states, by states*. Washington, DC: The National Academies Press.

Sadler, P.M. (1998). Psychometric models of student conceptions in science: Reconciling qualitative studies and distractor-driven assessment instruments. *Journal of Research in Science Teaching, 35*(3), 265-296.

Sadler, P. M., Sonnert, G., Coyle, H. P., Cook-Smith, N., & Miller, J. L. (2013). The influence of teachers' knowledge on student learning in middle school physical science classrooms. *American Educational Research Journal, 50*(5), 1020-1049.

Trumper, R. (1990). Being constructive: an alternative approach to the teaching of the energy concept. *International Journal of Science Education, 12*(4), 343-354.

Trumper, R. (1998). A longitudinal study of physics students' conceptions on energy in pre-service training for high school teachers. *Journal of Science Education and Technology, 7*(4), 311-318.

Wind, S. A. & Gale, J. D. (2015). Diagnostic opportunities using Rasch measurement in the context of a misconceptions-based physical science assessment. *Science Education, 99*, 721-741.

Watts, D. (1983). Some alternative views of energy. *Physics Education, 18*(5), 213-217.