Introduction to the ASPECt Project

Introduction

In today's society, citizens are constantly confronted with a wide range of energy-related issues that have significant personal and societal implications: deciding between purchasing a hybrid car or a traditional car, setting their thermostats to a higher temperature in the summer, or unplugging electrical devices when not in use. To make well informed decisions regarding these issues, it is essential to have an understanding of what energy is and how it can be transformed, transferred, or dissipated. Ideas about energy also have an important place in the school science curriculum with connections to topics such as photosynthesis and respiration or weather and climate that are encountered in the life and earth sciences. In the *Next Generation Science Standards* (NGSS), it is considered both a disciplinary core idea and a crosscutting concept (NGSS Lead States, 2013).

Despite the central role that energy concepts play in both real-world and school settings, these concepts are highly abstract and often counterintuitive and can be particularly challenging to students at all levels. The research literature has documented a wide range of energy-related misconceptions and alternative ideas that students bring to the classroom, but without diagnostic assessments designed specifically to help teachers pinpoint the particular problems their students are having and why, it will be difficult to help those students make progress in their science learning.

A few research-based energy assessments are available; however, their usefulness in locating K-12 students on a learning progression for energy is limited. This is either because they were designed for only high school and university students (Hestenes, 1999; Ding, 2007; Singh & Rosengrant, 2003) or they target too narrow of a range of energy ideas (Wattanakasiwich, Taleab, Sharma, & Johnston, 2013; Lee & Liu, 2010).

To address the need for a more general energy assessment for K-12 students, we developed the Assessing Student's Progress on the Energy Concept (ASPECt) instrument. The ASPECt instrument is a set of three tests that are precisely aligned to important ideas about energy and can be used to monitor how students progress as they learn about energy from late elementary school through high school.

Overview of the Targeted Learning Goals

The ASPECt instrument is designed to measure students' understanding of important energy ideas. The learning goals to which the instrument is aligned are fully described in the document titled "Description of the Energy Learning Goals." In brief, the concept of energy is commonly separated into four categories: (1) Energy Forms and Transformations; (2) Energy Transfer; (3) Energy Dissipation and Degradation; and (4) Energy Conservation (e.g. Duit, 2014). In our study, we divided the Energy Forms and Transformations category into five forms of energy along with the idea of energy transformations itself, and we expanded the Energy Transfer category into six specific mechanisms of energy transfer (see Table 1).

Table 1: *Energy Ideas Targeted by the Assessment 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	

For each idea, we described a learning progression that, for most ideas, was made up of three levels (Herrmann-Abell & DeBoer, 2018). The basic level of the progression focuses on simple energy relationships and easily observable effects of energy processes, similar to the elementary expectations in NGSS. The next level, the intermediate level, focuses on more complex energy concepts and corresponds to the middle school NGSS expectations. The highest level, the advanced level, focuses on still more complex energy concepts, often requiring an atomic/molecular model to explain phenomena. The advanced level covers many of the high school NGSS expectations. For example, the basic level idea for conduction states that when warmer things are touching cooler ones, the warmer things get cooler and the cooler things get warmer until they all are the same temperature. The intermediate level expects students to know that conduction is the transfer of energy that occurs when a warmer object comes in contact with a cooler object without a transfer of matter. Finally, the advanced level expects students to know that energy is transferred by conduction by the random collisions of atoms and molecules that make up the objects.

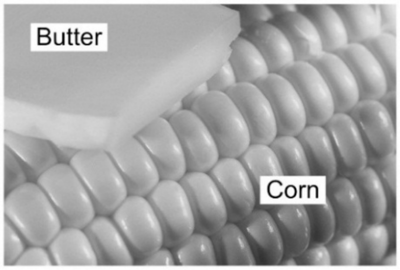
Summary of Item Development

Item development followed an iterative procedure that involved input and feedback from a number of sources. Detailed description of the process can be found elsewhere (DeBoer, Herrmann-Abell, & Gogos, 2007; DeBoer, Herrmann-Abell, Gogos, Michiels, Regan, & Wilson, 2008a; DeBoer, Lee, & Husic, 2008b; and DeBoer, Herrmann-Abell, Wertheim, & Roseman, 2009; Herrmann-Abell & DeBoer, 2014). Initially, multiple-choice items aligned to the different ideas and levels on the learning progression were drafted. Common student misconceptions (Driver, Squires, Rushworth, & Wood-Robinson, 1994) about energy were built into the assessment items as distractors (Sadler, 1998). Sample items aligned to the progression for conduction are shown in Figure 1.

The draft items were then pilot tested with students in grades 4 through 12 and reviewed by a panel of experts. During the pilot tests, students from across the country were asked to select what they thought was the correct response to the item and to answer a series of follow-up questions about the item (DeBoer, Herrmann-Abell, Gogos, Michiels, Regan, & Wilson, 2008a). The follow-up questions provided information about how well the item was performing, including information about what knowledge the students were using to answer the question and any difficulties they had in understanding the question. During the panel review, scientists, science education experts, and classroom teachers evaluated the items' scientific accuracy and alignment to the targeted learning goals. The panel members also flagged features that may cause problems for students such as comprehensibility issues, issues with the task contexts, or test-wiseness issues. The full description of the analysis criteria used during panel review can be found at: http://www.project2061.org/research/assessment/assessment_form.htm. The feedback from the pilot test students and the panel experts were then used to revise the items. The revised items were then field tested with a larger group of students. Rasch analysis (Bond & Fox, 2007) was used to evaluate how well the revised items were functioning and build the final item bank of over 300 items.

Basic level conduction item

A cook heats up some corn. Then she cuts a piece of cold butter and places it on top of the hot corn.



What will happen to the temperature of the corn and the butter as soon as she puts the butter on top of the corn?

A. Both the corn and the butter will get cooler.
B. Both the corn and the butter will get warmer.
C. The corn will get cooler, and the butter will get warmer.
D. The corn will stay the same temperature, but the butter will get warmer.

Intermediate level conduction item

The temperature of a plastic block is 60°F, and the temperature of a metal block is 40°F. A student puts the plastic block on top of the metal block. Will the blocks ever reach the same temperature? Why or why not?

A. Yes, but only for a little while because the metal block will continue to get warmer and the plastic block will continue to cool
B. Yes, because energy will be transferred from the plastic block to the metal block until they reach the same temperature
C. No, because the temperature difference is not large enough for energy to be transferred
D. No, because the blocks are made of different materials

Advanced level conduction item

A person pours a hot drink into a cup and then places a room temperature spoon in the cup. After a while, the person notices that the handle of the spoon has gotten hotter. What caused the handle to get hotter?

A. Heat molecules from the hot drink are absorbed by the spoon. These heat molecules travel to the handle of the spoon, making the handle hotter.
B. The molecules that make up the hot drink are rubbing against each other harder than the molecules that make up the spoon. The rubbing creates new energy that flows through the spoon to the handle, making the handle hotter.
C. The hot drink causes the molecules of the spoon to speed up. These faster moving molecules then move to the handle of the spoon, causing the handle to get hotter.
D. The hot drink causes the molecules of the spoon to speed up. When these faster moving molecules collide with slower moving molecules, energy is transferred to the slower moving molecules. These collisions continue to occur throughout the spoon until they reach the handle, making the handle hotter.

Figure 1: Sample items aligned to the progression for conduction

Summary of Instrument Development

The results of the field test, along with the clarification of the learning goals, were used to inform the selection of items for the instruments. Three instruments, each made up of 35 items, were drafted. The basic test includes mostly items that assess the basic level of each of the energy ideas. The intermediate test consists of primarily intermediate items and the advanced test consists of primarily advanced items. There were two items that appeared on all three tests, three items that appeared on the basic and intermediate tests, and three items that appeared on the intermediate and advanced tests. The instruments were then pilot tested to determine how well the items perform as sets. Rash analysis was used to evaluate the draft instruments (Hardcastle, Herrmann-Abell, & DeBoer, 2017). The results of the instrument pilot test were used to revise the instruments. On the final versions, five items appear on both the basic and intermediate versions and five items appear on both the intermediate and advanced versions. In addition, it was decided that item difficulties would be determined from the full data set so that all of the items in the bank would be on the same scale. Finally, the revised instruments were field tested and the comparability of computer-based and paper-based versions of the instruments was investigated. The comparability study results indicated that scores from the computer-based version and paper-based versions can be considered equivalent (Herrmann-Abell, Hardcastle, & DeBoer, 2018).

Summary of Support Material

In order to support users in interpreting the results of the instruments, support materials, including raw score to Rasch scale score tables, Wright maps, and option probability curves, were developed.

Raw score to Rasch scale score tables. Because we cannot expect users of our instruments to conduct their own Rasch analysis, we developed conversion tables that can be used to convert the raw scores to Rasch scale scores. These scaled scores are based on the assumption that students' response patterns follow a Guttman pattern, where if a student responds correctly to a particular item, they also respond correctly to the items that are easier. Common Rasch units of analysis are logits (log odds units) which range from approximately -3.0 to 3.0 for most respondents, with a mean score of zero. Because reporting scores as negative numbers can be confusing, we utilized a transformation to express all possible student performance with positive numbers ranging from 200 to 800.

Wright maps. We constructed Wright maps (Wilson, 2005) that help provide a more qualitative meaning to the scale scores. A Wright map can be considered a visual depiction of the learning progression for a concept (Wilson, 2009; Black, Wilson, & Yao, 2011). On the map, a vertical line is drawn to represent the scale. Students' performance level is commonly shown on the left-hand side, and the range of item difficulties are shown on the right-hand side. Less knowledgeable students and easier items are toward the bottom of the map, and more knowledgeable students and harder items are toward the top of the map. When a user finds a student's scale score on the map, they can determine where the student is on the progression, that is, what energy ideas have they mastered, what ideas are they developing, and what ideas have they yet to learn. When a student's performance level matches an item's difficulty, the student has a 50% chance of successfully responding to that item. Therefore, the student is more likely to respond correctly to items with a lower difficulty and less likely to respond correctly to items with a higher difficulty.

Option probability curves. Option probability curves plot the probability that students will select each answer choice as a function of their Rasch scale score. 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. Users can find a student's scale score on the x-axis and determine the probability the student will select each answer choice. This will provide information about possible misconceptions or alternative mental models the student may hold, which can inform the selection and sequencing of instructional activities.

Student Performance and Instrument Use

The data set used to determine the item difficulties was comprised of three separate sources; the item field test, the instrument pilot test, and the instrument field test. A total of 30,811 students from 45 states and Puerto Rico were included in the set (see Table 2). A sample of university students who were likely to have the knowledge being targeted by the items were included as a way of further validating the items. All of the students were studying science but not necessarily energy at the time of testing.

Table 2: *Demographic information for the anchoring data set*

	Item Field Test	Instrument Pilot Test	Instrument Field Test	Total
Year	Spring 2015	Winter 2016	Fall 2016	
Grade Band				
4th-5th	2967 (14%)	470 (11%)	848 (15%)	4285 (14%)
6th-8th	10390 (50%)	1651 (39%)	2425 (43%)	14466 (47%)
9th-12th	7414 (36%)	1895 (44%)	2408 (42%)	11717 (38%)
University/College	0 (0%)	244 (6%)	0 (0%)	244 (1%)
Gender				
Female	51%	53%	54%	52%
Male	49%	47%	46%	48%
Primary Language				
English	89%	88%	92%	89%
Not English	11%	12%	8%	11%

ANCOVA was used to perform a cross-sectional analysis of students' performance by grade 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. Student scale scores increased with increasing grade ($F(9, 29339) = 193$, $p < .001$) (see Table 4). Bonferroni post-hoc tests showed that the differences among grades were statistically significant for all pairs except between fourth and fifth grades, fourth and sixth grades, and eighth and ninth grades.

Table 4: *Estimated Marginal Student Means by Grade*

Grade band	Mean Student Score	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
4th	477	.80	475	478
5th	476	.70	474	477
6th	479	.54	478	480
7th	482	.49	481	483
8th	486	.47	485	487
9th	488	.60	487	489
10th	493	.58	491	494
11th	498	.59	496	499
12th	502	.88	500	503
University	509	2.25	504	513

To more accurately measure a student's location on the learning progression, it is best to use an instrument that is well matched to the student's current knowledge level. Therefore, we compared students' performance to the average item difficulties to provide guidance on which test is appropriate for which students. Table 5 summarizes the average item difficulty by instrument and Figure 2 compares the mean item difficulty of each instrument to the mean student scale score at each grade. The average difficulty of the items on the basic test was 474, which is below the average student score at each grade, even for elementary students. The average difficulty for the intermediate test, 498, is equal to the average score for the eleventh graders and below the average scores for twelfth grade and university students. The average difficulty for the advanced test, 521, is above the average student score at each grade. Based on this, we would suggest that users begin with using the basic test with elementary and middle school students and the intermediate test with high school students. After instruction on the energy concept, we would recommend retesting using the next level up to assess student's progress.

Table 5: *Mean Rasch Item Difficulty by Instrument*

Instrument	Min	Max	Mean	SD
Basic	416	557	474	32
Intermediate	438	552	498	29
Advanced	466	579	521	26

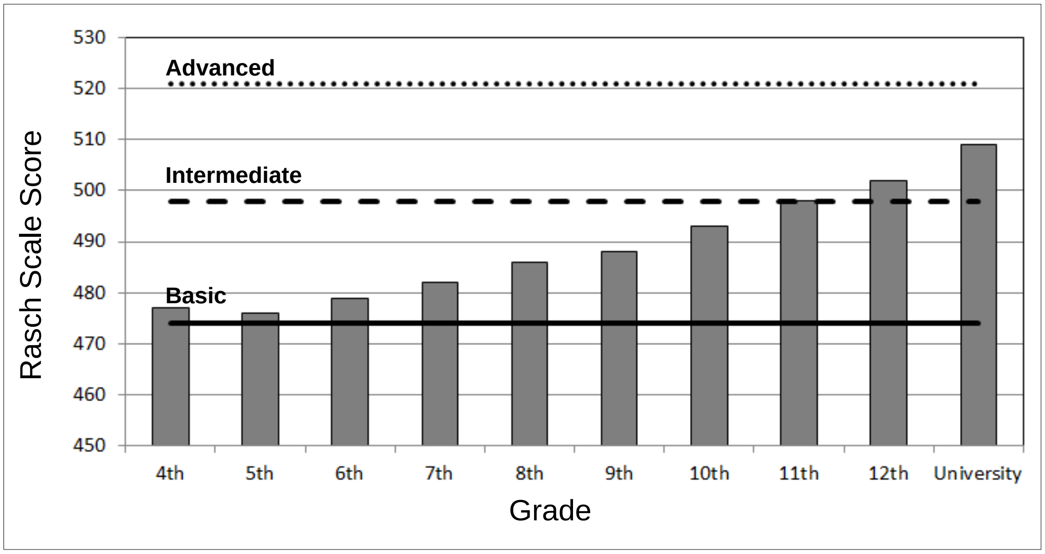


Figure 2: Mean Rasch scale score by grade compared to mean Rasch item difficulty by test

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 the views of the Institute or the U.S. Department of Education.

References

Black, P., Wilson, M., & Yao, S.Y. (2011). Road maps for learning: A guide to the navigation of learning progressions. *Measurement: Interdisciplinary Research and Perspectives, 9*(2–3), 71–123.

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

DeBoer, G. E., Herrmann-Abell, C. F., & Gogos, A. (2007). *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., Herrmann-Abell, C. F., Wertheim, J. A., & Roseman, J. E. (2009). *Assessment linked to middle school science learning goals: A report on field test results for four middle school science topics.* Paper presented at the National Association of Research in Science Teaching Annual Conference, Garden Grove, CA.

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.

Ding, L. (2007). *Designing an Energy Assessment to Evaluate Student Understanding of Energy Topics* (Doctoral dissertation). Retrieved from NCSU Libraries, https://repository.lib.ncsu.edu/handle/1840.16/4050

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

Duit, R. (2014). Teaching and learning the physics energy concept. In R. F. Chen, A. Eisenkraft, D. Fortus, J. Krajcik, K. Neumann, J. Nordine, & A. Scheff (Eds.), *Teaching and learning of energy in K-12 education* (pp. 67–85). New York: Springer.

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

Hardcastle, J., Herrmann-Abell, C. F., & DeBoer, G. E. (2017, April). *Validating an Assessment for Tracking Students' Growth in Understanding of Energy from Elementary School to High School*. Paper presented at the 2017 NARST Annual International Conference, San Antonio, TX.

Herrmann-Abell, C.F. & DeBoer, G.E. (2011). Using distractor-driven standards-based multiple-choice assessments and Rasch modeling to investigate hierarchies of chemistry misconceptions and detect structural problems with individual items. *Chemistry Education Research and Practice, 2,* 184-192.

Herrmann-Abell, C. F., & DeBoer, G. (2014). Developing and using distractor-driven multiple-choice assessments aligned to ideas about energy forms, transformation, transfer, and conservation. In R. F. Chen, A. Eisenkraft, D. Fortus, J. Krajcik, K. Neumann, J. Nordine, & A. Scheff (Eds.), *Teaching and learning of energy in K-12 education* (pp. 103–133). New York: Springer.

Herrmann-Abell, C. F., & DeBoer, G. E. (2018). Investigating a learning progression for energy ideas from upper elementary through high school. *Journal of Research in Science Teaching, 55*(1), 68-93.

Herrmann-Abell, C. F., Hardcastle, J., & DeBoer, G. E. (2018, March). *Comparability of Computer-Based and Paper-Based Science Assessments*. Paper presented at the 2018 NARST Annual International Conference, Atlanta, GA.

Hestenes, D. (1999). Development of an Energy Concept Inventory. Arizona State University: National Science Foundation.

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.NGSS Lead States, 2013).

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.

Singh, C., & Rosengrant, D. (2003). Multiple-choice test of energy and momentum concepts. *American Journal of Physics*, *71*(6), 607–617.

Wattanakasiwich, C., Taleab, P., Sharma, M., & Johnston, I. D. (2013). Development and implementation of a conceptual survey in thermodynamics. *International Journal of Innovation in Science and Mathematics Education, 21*(1), 29-53, 2013.

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

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

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.