

Chapter 5

Stage-Like Patterns in the Development of Conceptions of Energy

Theo L. Dawson-Tunik

Hampshire College

In this chapter, I examine developmental patterns in the acquisition of the energy concept in a group of 9th graders. The results suggest (1) that many ninth graders fail to achieve an adequate understanding of the energy concept; (2) that an understanding of this concept requires that students are able to construct linear arguments composed of abstract¹ concepts, a capability identified with the *abstract mappings* level in Fischer's (1980) developmental sequence of skill (complexity) levels; (3) that many of the conceptual elements of these arguments are formed at the complexity level preceding abstract mappings; and (4) that the difficulty of learning the energy concept is much greater for students who do not demonstrate the ability to construct abstract mappings.

The Development of Energy Conceptions

As has been pointed out elsewhere, curricula and instruction should be based on empirical evidence about the developmental pathways through which concepts develop (Driver, Rushworth, and Wood-Robinson, 1994; Liu and McKeough, 2005). Unfortunately, this is rarely the case, since few developmental sequences of this kind have been described. At present,

the typical teaching sequence for the energy concept has been determined primarily through expert consensus. This sequence is: (1) energy source, (2) work, (3) energy transfer, and (4) energy conservation (Liu and McKeough, 2005). It is expected that students will acquire the concept of energy conservation during the high school years (American Association for the Advancement of Science, 2001), and by the 8th or 9th grade, students are expected to understand the concept of energy transfer and the notion of energy as the ability to do work.

The results of the investigation described here suggest that many ninth graders achieve neither an understanding of energy as the ability to do work nor an understanding of energy transfer. Further, our results suggest that an understanding of these concepts requires that students are able to construct linear arguments composed of abstract² concepts, a capability that we identify with the *abstract mappings* level in Fischer's (1980) developmental sequence of complexity levels (discussed further below). Moreover, it appears that many of the conceptual elements of these arguments are formed at the complexity level preceding abstract mappings—single abstractions—and that these conceptual elements are built up from more concrete conceptions identified at earlier complexity levels. From the perspective of Fischer's (1980) skill theory, the fact that adequate understandings of energy transfer and of energy as the ability to do work require abstract mappings implies that many 9th grade students may be unable to understand these concepts because they are still building the requisite precursor conceptions of the single abstractions level—the level before abstract mappings. In addition, a growing body of research into the nature of developmental transitions provides evidence that movement from one developmental level to the subsequent level is much more difficult than learning new concepts at one's current developmental level (Dawson-Tunik, 2004a; Dawson-Tunik, Commons, Wilson, and Fischer, 2005; Liu and McKeough, 2005). This means that for some 9th graders, learning the energy concept may be a much more arduous task than for others, not because they are less able students, but because they are not developmentally prepared.

In this chapter, I explore this issue by employing the Rasch model to examine data from a developmental study of the acquisition of the energy concept. First, however, I provide short descriptions of the energy conceptions associated with three complexity levels: representational systems, single abstractions, and abstract mappings (defined in more detail

below), followed by some background on the theoretical and methodological frameworks that inform this work.

The developmental sequence

In an earlier paper, Stein and I (manuscript submitted for publication) described a developmental sequence of acquisition for the energy concept. These descriptions focus on the development of the concepts of energy transfer and transformation. They are based on students' responses to a series of questions about a bouncing ball, rolling balls, and balls shown in various positions relative to a spring. Three of these levels are described here.

Representational systems

At the representational systems level, children often provide elaborate observations of the movements of a bouncing ball, including the observation that a bouncing ball bounces lower and lower. Children performing at this level may employ the term energy, but their use of the term indicates that motion and energy are undifferentiated. To these children, there is a direct correspondence between energy and motion. Some children performing at this level use the term *gravity*—meaning something that pushes or pulls—as an explanation for the successively lower bounces of a bouncing ball, but they are unable to explain the mechanism.

Single abstractions

At the single abstractions level, the term energy means something “behind” motion—a cause for motion. Students performing at this level may employ the terms kinetic or potential energy, though their use of the term *potential energy* indicates that they do not think of it as a true form of energy, but rather as a potential for energy to happen. At this level, students may speak of energy transfer, explaining that the energy of a ball transfers to the floor during a bounce, much as a liquid flows from one location to another. They may also speak of gravity as a force that gradually takes away all of a bouncing ball's energy.

Abstract mappings

At the abstract mappings level, kinetic and potential energy are finally understood as different energy states. Students can describe transformations from one energy state to the other, sometimes referring to types of potential energy. Energy transformation is conceptualized as a change of energy from one state to another, as in a change of energy as motion to

energy as heat, sound, or potential energy. Energy transfer is described as a quantitative exchange, and gravity is classified as a constant force that provides objects with potential energy in some situations and kinetic energy in others. Some students define energy as the ability to do work, demonstrating in their explanations that they think of energy as a quantity rather than as a substance.

As can be seen from these descriptions, the conceptions of students operating below abstract mappings are inadequate for constructing an understanding of energy transfer or transformation. For example, an adequate understanding of energy transformation requires that potential and kinetic energy are both conceptualized as energy states. Students performing at single abstractions are able to conceptualize kinetic energy as the energy of motion, which arguably is an energy state, but potential energy is clearly not conceptualized as an energy state. Rather, it is viewed as the potential for energy to happen. State change, given the single abstraction conceptualizations of potential and kinetic energy would be a shift from having energy to having the potential to have energy.

The following section lays out the notion of hierarchical complexity as a central construct in the developmental model employed here.

Cognitive Development and Hierarchical Complexity

From a Piagetian (Piaget, 2000) perspective cognitive development is the activity of reflective abstraction, through which new knowledge is constructed from existing knowledge. Hierarchical integration is the product of reflective abstraction, and hierarchical complexity is the most direct observable manifestation of hierarchical integration. The concept of hierarchical complexity is at the core of most cognitive developmental theories, including those proposed by Piaget (Piaget and Inhelder, 1969), Fischer (1980), Case (1985), Commons (Commons, Danaher, Miller, and Dawson, 2000), Halford (1999), and Demetriou (Demetriou, Platsidou, Efklides and Metallidou, 1991). These theories are based on the idea that the growth of knowledge is not a simple cumulative process, but requires periodic reorganizations of thought structures. These reorganizations involve the creation of new concepts by integrating the operations/products of earlier thought processes in novel ways. The new concepts represent more than the sum of the operations/products of the previous level. For example, by around age 10, many Western children have constructed a notion of *learning as play*. While this notion requires previously con-

structed *learning* and *play* elements, it clearly requires more than these elements on their own. The new meaning is an argument about their relationship, and an entirely new concept. Moreover, both of the earlier conceptions are transformed by this construction. Not only is the concept of learning enriched by the idea that it can be fun, but the notion of play is enriched in that it is now consciously linked with a productive activity.

Properties of stages

Because developmental stages represent successive hierarchical integrations—meaning that each new stage is constructed from the actions of the previous stage—the sequence of development must be from one stage to the next with no skipping.³ It is conventionally held that the only way to provide evidence in support of invariant sequence is to conduct longitudinal research, showing empirically that, within persons, development proceeds sequentially (Armon, 1984; Armon and Dawson, 1997; Colby, 1981; Rest, 1975; Walker, 1982; Walker, Gustafson, and Hennig, 2001). Several longitudinal studies of this kind have been conducted, providing support for invariant sequence (Fischer and Bullock, 1981). Despite the fact that longitudinal data can make a compelling case for invariant sequence, there are patterns of performance in cross-sectional data that provide support for sequentiality. For example, sequentiality is supported by evidence that individuals always perform (within measurement error) at complexity levels that are adjacent to one another in the specified sequence (Dawson-Tunik, 2004a; Dawson-Tunik, Commons, Wilson, and Fischer, 2005).

Each of Piaget's original stages is defined by a set of formal properties that constitute a *structure d'ensemble*, or structure of the whole (Fischer and Bullock, 1981). This has sometimes been taken to mean that the entire knowledge system forms a single unified global structure (1996). In some (not all, see Lourenco and Machado (Demetriou and Efklides, 1994; Fischer and Bidell, 1998) interpretations of stage transitions based on Piaget's notion of *structure d'ensemble*, abrupt *global* reorganizations of the entire knowledge system characterize development, which is modeled as a staircase. However, because analogous structures—especially analogous structures in different knowledge domains—often do not develop in parallel, attempts to model development globally (in multiple domains) produce patterns that make development appear more or less smooth. In response to the lack of evidence for global step-like development, some have argued that development is better characterized as smooth and con-

tinuous. Flavell (1996), for example, suggests that progress through developmental stages is characterized by a gradual replacement of lower stage structures over time. Others, like Siegler (1994), have suggested that development is continuous, though not smooth. Still others argue that development within individual knowledge domains is characterized by periods of consolidation (plateaus) during which performance within a domain tends to be largely homogeneous (is predominantly at a single stage), and transitional periods (spurts) characterized by vacillation between the modal stage and its successor (Andrich and Styles, 1994; Case, 1991; Dawson-Tunik, 2004a; Dawson-Tunik, Commons, Wilson, and Fischer, 2005; Draney, 1996; Fischer and Rose, 1994; Fischer and Rose, 1999; Fischer and Silvern, 1985; Kitchener, Lynch, Fischer, and Wood, 1993; Liu and McKeough, 2005; Shultz, 2003; Thomas and Lohaus, 1993; Van Der Maas and Molenaar, 1992; Van Geert, 1998; Walker, Gustafson, and Hennig, 2001; Wilson, 1989). These authors have reported developmental patterns like the one idealized in Figure 1, in which each curve represents a developmental level, the horizontal axis represents time, and the vertical axis represents the proportion of reasoning expected at a given developmental level at a particular point in time. Note that there is little overlap between curves representing non-adjacent levels. This is a model of transition and consolidation in which there are periods within which performances tend to be at a single level followed by periods where performances represent a mixture of two adjacent levels. The developmental pattern revealed in Figure 1 also provides support for developmental discontinuities. Note that the likelihood that a person will perform at two adjacent levels is rarely as great as the likelihood that he or she will perform almost exclusively at a single level. Moreover, the period of time in which an individual is highly likely to perform at two adjacent levels is considerably shorter than the time during which an individual is highly likely to perform primarily at one level. This means that transitions are short and follow much longer periods of consolidation.

It has been argued that the results reported by these researchers are an artifact of the methods employed to assess developmental level. In particular, the use of whole level scoring, in which performances are assigned to a complexity level without distinguishing between periods of transition and consolidation, has been criticized. At present, there appear to be only two published reports in which transitional and consolidated performances have been modeled (Armon and Dawson, 2002; Dawson, 2002b). Interestingly, the analyses presented in these articles revealed developmental patterns

like the one idealized in Figure 2, which shows that transitional phases subsume the areas where level mixtures are expected, and are shorter and less prominent than consolidated periods. Interestingly, Figures 1 and 2 are not very different from one another in terms of the overall shape of development, despite the use of different scoring systems.

The results reported above suggest that developmental discontinuities between developmental levels are not an artifact of whole level scoring. I hypothesize that a Rasch model of students' performances on interviews about the energy concept will reveal similar discontinuities between developmental levels. The next section examines performances on a set of energy interviews of 9th grade physical science students. These were scored with the Lectical™ Assessment System⁴ (Dawson-Tunik, 2005), a do-

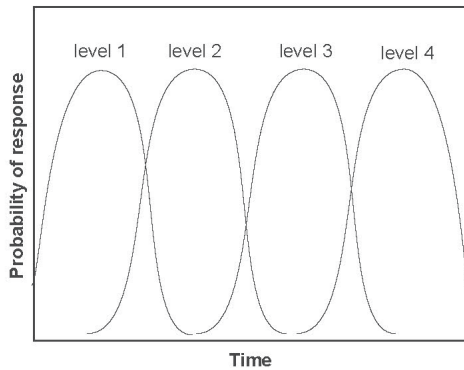


Figure 1. Distribution of developmental levels over time

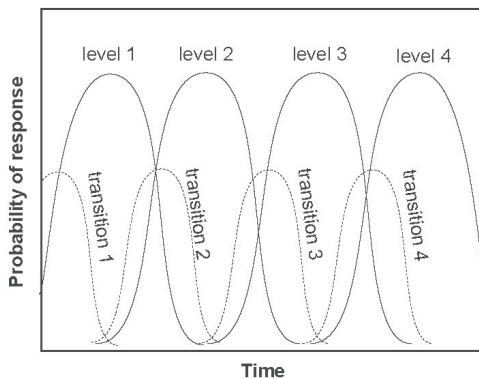


Figure 2. Distribution of developmental levels and transitions over time

main independent developmental assessment system based on Fischer's (1980) skill (complexity) levels and Commons' (Commons, Danaher, Miller, and Dawson, 2000) general stages.

Methods

Instrument

The interview instrument was the *Energy Teaser*, a worksheet that poses a series of questions about the energy of resting, rolling, and bouncing balls. Figure 3 depicts the bouncing ball problem. Students filled out a paper form of the teaser, then participated in a clinical interview that probed their responses. Probes were designed to expose the form and conceptual content of students' reasoning about the teaser questions.

Procedures

Because our primary intentions in conducting our larger project were to describe a sequence of conceptual development for the energy concept and to employ what we learned to inform the design of lesson plans and scoring rubrics for teachers (rather than to test a hypothesis), we were able to employ an informal research design. This was fortunate, as constraints imposed by participating schools made it impossible to implement formal sampling procedures.

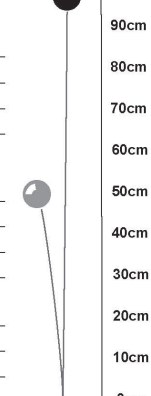
Figure 3: Bouncing ball problem from the *Energy Teaser*

The ball in the drawing on the right is dropped onto the floor from a height of 100 centimeters. It then bounces to a height of 50 centimeters.

(g) Explain your theory of what is happening to the energy of the ball as it is falling.

(h) Explain your theory of what happens to the energy of the ball at the moment when it hits the floor.

(i) Explain your theory of what happens to the energy of the ball after it hits the floor.



The diagram shows a vertical height scale on the right side, ranging from 0cm at the bottom to 100cm at the top, with major markings every 10cm. A ball is shown at the 100cm mark, representing its initial drop height. A second ball is shown at the 50cm mark, representing its maximum bounce height. A vertical line connects the two ball positions, and a horizontal line at the 0cm mark represents the floor. The balls are depicted as grey circles with a small white highlight to indicate a three-dimensional shape.

Figure 3. Bouncing ball problem from the *Energy Teaser*

A total of 6 ninth grade teachers, all of whom were participants in Hampshire Colleges' Collaboration for Excellence in Science Education (CESE) program, agreed to administer the Energy Teaser before and after teaching a unit on energy. The energy unit was approximately one week in duration and incorporated an activity in which students were required to observe and attempt to explain energy state changes in bouncing balls. See the CESE web site for a complete description of this activity (Dawson-Tunik, Wenk, and Paulman, 2005). Teachers further agreed to allow researchers to spend one day prior to instruction and one day following instruction interviewing student volunteers from each of their physical science classes. At the post-instruction interviews (Time 2), every effort was made to meet with the students who had participated in the pre-instruction interviews (Time 1). Extremely poor student attendance and the voluntary nature of participation made this difficult. Consequently, few students were interviewed at both time 1 and time 2.

Sample

This sample includes, 49 students who were interviewed at time 1 and 34 different students who were interviewed at time 2. The second interview of any student who was interviewed twice was eliminated from this analysis. Students ranged in age from 14 to 16. The two 16-year-olds were repeating a year.

Developmental Analysis

The Lectical™ Assessment System (LAS)

Several attempts have been made to develop a generalized developmental assessment system for human raters. Indeed, Piaget defined each of his developmental stages in generalized terms. Conservation, for example, is a general feature of concrete operations and can be observed on a wide range of tasks. Case (Case, Griffin, McKeough, and Okamoto, 1992), Fischer (Fischer and Bidell, 1998; Rose and Fischer, 1989), and their colleagues have employed generalized definitions to scale performances across domains, but have not disseminated generalized scoring systems. Grounded in Commons' General Stage Scoring System (Commons, Danaher, Miller, and Dawson, 2000) and Fischer's skill theory (1980), the LAS (Dawson-Tunik, 2004b), lays out explicit general criteria for determining the developmental level of performances in any domain of knowledge.

The LAS is based on a three-layer model of conceptual structure. In this model, the outer layer represents *conceptual content*, the middle layer

represents *domain-level structure*, and the inner layer represents *core structure*. The LAS targets core structures to determine the *complexity level* of a performance. These core structures are *hierarchical order of abstraction* and *logical structure*.

When assessing the complexity level of a text with the LAS, the analyst refers to two manifestations of hierarchical complexity: its conceptual structure, embodied in the *hierarchical order of abstraction*⁵ of the new concepts employed in its arguments, and the most complex *logical structure* of its arguments. These conceptual and logical structures are identical by definition and fundamentally interdependent. We make a distinction between the two types of structure for heuristic and pragmatic reasons. When scoring texts, hierarchical order of abstraction refers primarily to the structure of the elements of arguments, which often must be *inferred* from their meaning in context, whereas logical structure refers to the *explicit* way in which these elements are coordinated in a given text.

The level names from Fischer's (Fischer and Bidell, 1998) skill theory label the complexity levels: (0) reflexive actions, (1) reflexive mappings, (2) reflexive systems, (3) single sensorimotor actions, (4) sensorimotor mappings, (5) sensorimotor systems, (6) single representations, (7) representational mappings, (8) representational systems, (9) single abstractions, (10) abstract mappings, (11) abstract systems, and (12) single principles/axioms.

Here, I provide short definitions of each of the levels identified in the sample of interviews collected for this project, along with commonly reported modal ages of acquisition. See the LAS web site (Dawson-Tunik, 2005) for more information about these constructs and examples of performances from each level in several knowledge domains.

At the representational systems level, the new concepts are 3rd order representational sets. These coordinate elements of representational systems. For example, the concept of bounciness and the notion that gravity is something that pulls or pushes on objects can be employed in concert to explain the bouncing of a ball through a sequence of lower and lower bounces. "The ball bounces because it is made of bouncy stuff, but it can't come up so high again because gravity is pushing on it. [What is gravity?] I don't know. It just pulls on things so they don't fall up." Concepts like *energy as motion* and *gravity as something that pushes or pulls* are infrequently observed before this complexity level. The most complex logical structure of this complexity level is multivariate, coordi-

nating multiple aspects of two or more representations, as in, “If you drop the ball it will bounce lower and lower and make littler and littler noise... and keep on bouncing until it stops,” in which the child takes note of both the decreasing size of bounces and the decreasing sound made by the ball on successive bounces.

At the single abstractions order (abbreviated as SA in tables and figures), the new concepts are referred to as 1st order abstractions. These coordinate representational systems. For example, the concept of *kinetic energy* (energy of motion) coordinates many observations of moving objects, providing an explanation for their activity. It partially transcends the representational systems tendency to conflate energy with motion. Energy is now viewed as something behind motion—its cause. “[A hand is pressing down on a ball that is sitting on top of a spring, does the ball have energy?] It is kinetic energy going into potential energy. [Okay so what is the kinetic?] The hand, it’s pressing. [Okay and then when does it become potential?] When the spring is down, and then the spring pops up the ball to make it... Energy is, I guess,... movement (Case 10380).” A partial understanding of *energy transfer* emerges at single abstractions. “[What happens to the energy of the ball at the moment when it hits the ground?] I think the energy will move into the ground, the force of the ball hits the ground and... it would hit the ground and then I guess the shock from it would just go into the ground (Case 30008).” Here, energy is conceptualized as a substance that moves between two objects. The most complex logical structure of this complexity level identifies one aspect of a single abstraction, as in “Kinetic energy is the energy of motion” in which kinetic energy is defined in terms of motion.

At the abstract mappings order (abbreviated as AM in tables and figures), the new concepts are referred to as 2nd order abstractions. These coordinate or modify abstractions. For example, the abstract mappings order conception of *energy transformation* coordinates conceptions of kinetic and potential energy. “[When the ball is] speeding up, it gains more kinetic energy and loses potential, but it always keeps the same amount of energy, because of conservation of energy (Case 1641).”⁶ Concepts like *kinetic potential energy*, *heat energy*, *sound energy*, and *friction* are also rare in discussions of energy before the abstract mappings order. “[The reason that the ball is only going about half as high after the bounce is because some of the energy] has been transferred to the floor. [How is it transferred to the floor, do you know what the process is like?] It is when

the ...ball's energy [is] transferred into sound and heat" (Case 10634). The most complex logical structure of this complexity level coordinates one aspect of two or more abstractions, as in "[After it bounces] the ball ... will slow down. So, it will lose energy. Well, it's not really lost, it is still there, but it is being transferred into heat and sound" (Case 10530). Here, a conception of energy transfer coordinates types of energy—kinetic (movement), heat, and sound.

Scoring procedures

It is not possible to fully describe LAS scoring procedures in this short section, as it takes many hours of instruction and practice to become an accurate LAS analyst. To obtain a basic understanding of how LAS analysts assign a score, refer to the LAS web site (Dawson-Tunik, 2005). Scores are based on hierarchical order of abstraction and logical structure, and analysts must understand how these manifest in a given performance. Scoring is an iterative process; the analyst alternately examines each layer of structure until he or she converges on an interpretation of the core structure of the performance. For example, an analyst was asked to score the following interview segment:

[*Could you have a good life without having had a good education?*] Yeah, probably so, I would say. I wouldn't...it would be richer with education, but it wouldn't... [*Why do you think it would be richer with education?*] Well, you just, your mind would be open to a lot more things. (Dawson-Tunik, 2004a, case 0212)

The analyst's response illustrates how each layer of structure plays a role in the scoring process:

Well, this isn't a very sophisticated notion of the role of education in the good life. Especially because, at first, I thought that he was saying that you'd be richer, money-wise (laughter), with an education. That would make "richer" a [representational] notion, but I see that it's actually at least abstract, because it's related to this idea of open-mindedness. It seems there are two variables [richer life, open mind] that are in a logical relation to one another...as in, "If you get a good education, your mind will be more open, and therefore you will have a richer life." This would make it at least [abstract mappings], but could it be higher than that? Well, *richer life* could be higher than [single abstractions], and so could *open mind*, so I'm looking for evidence that they are...but the perspective here is of the individual person and his life, without bringing in anyone else's per-

spective, or a social perspective, so you can't say, really. [Abstract mappings]; I'll stick with that (Dawson-Tunik, 2004a).

In this example, the analyst appeals to all three levels of structure. The content level is referenced in her initial attempt to understand the argument, and again when she double checks her understanding at the end. The domain structure level is briefly included when she examines the social perspective of the respondent to see if there are grounds for considering the possibility that the statement is at a higher level than abstract mappings. The core structure is reflected in her analysis of the hierarchical order of abstraction and logical structure of the argument.

From this example, it is clear how *meaning* is central to the scoring process. Without a correct interpretation of the meaning of a statement, the analyst cannot even begin the process of scoring. In this case, knowing which sense of *richer* is intended by the respondent is essential to a correct interpretation of the hierarchical order of abstraction of the concept.

Reliability and Validity of the LAS

My colleagues and I have undertaken several studies of the reliability and validity of the LAS and its predecessors (Dawson, 2002a, 2003, 2004; Dawson and Gabrielian, 2003; Dawson, Xie, and Wilson, 2003; Dawson-Tunik, 2004a). We have examined inter-analyst agreement rates, compared scores obtained with the LAS with scores obtained with more conventional scoring systems, and examined the scale's functioning through statistical modeling. Inter-analyst agreement rates have been high, 80% to 97% within half of a complexity level (Dawson, 2004; Dawson and Gabrielian, 2003; Dawson-Tunik, 2004a).⁷ Correspondences between other developmental scoring systems and the LAS are also high, revealing agreement rates of 85% or greater within $\frac{1}{2}$ of a complexity level (Dawson, 2002a, 2004; Dawson, Xie, and Wilson, 2003). Employing Rasch scaling, which provides reliability estimates that are equivalent to Cronbach's alpha, we have consistently calculated reliabilities over .95 (Dawson, 2002a; Dawson, Xie, and Wilson, 2003; Dawson-Tunik, 2004a). Overall, our research shows that the LAS is a valid and reliable general measure of intellectual development from early childhood through adulthood.

LAS analysts are required to maintain an agreement rate of 85% within $\frac{1}{3}$ of a complexity level with a Certified Master LAS analyst. All of the interviews were initially scored by the second author. The first author scored a subset of 71 protocols. The agreement rate between the first author and the second author was 75% within $\frac{1}{5}$ of a complexity level,

92% within 2/5 of a complexity level, and 97% within 3/5 of a complexity level. There were no disagreements greater than 1 complexity level. Kendalls tau b was .87.

Interview scoring

Before scoring the responses to the energy interview, the interviews were divided into segments, or *protocols*, by probe question. There were 3 to 7 protocols per interview. Some respondents, due to time constraints, received only the rolling and bouncing ball questions, whereas others were also asked about balls in a variety of other situations, such as being held above a spring, sitting on a spring, or falling toward a spring. Each protocol was scored with the LAS, employing the 5 phase version of the scoring system. In this version, the analyst not only evaluates the general complexity level of a protocol, but also of 5 phases: transitional into the level, unelaborated, elaborated, highly elaborated, or transitional out of the level. The levels identified ranged from elaborated representational mappings to elaborated abstract mappings. Before conducting the Rasch analysis scores were collapsed into two phases per complexity level: transitional (transitional out and transitional in) and consolidated (unelaborated to highly elaborated).

Rasch analysis

A Rasch partial credit analysis was conducted with WINSTEPS (Linacre and Wright, 2000). Rasch models are designed specifically to examine hierarchies of person and item performance, displaying both person proficiency and item difficulty estimates along a single interval scale (logit scale) under a probabilistic function. In addition, they can be employed to test the extent to which items or scores conform to a theoretically specified hierarchical sequence. A central tenet of stage theory is that cognitive abilities develop in a specified sequence, making the statistical tests implemented in a Rasch analysis especially relevant to understanding stage data. The Rasch model permits researchers to address questions like, "Are all single abstractions items more difficult than all representational systems items and less difficult than all abstract mappings items?" Moreover, the detailed information about item functioning and individual performances provided by the software makes it possible to simultaneously examine group and individual effects. These properties make Rasch models uniquely suitable for the investigation of many developmental phenomena.

It is beyond the scope of this chapter to provide a comprehensive account of the Rasch model, though I do attempt to provide enough infor-

mation to allow readers who are unfamiliar with the model to follow the results of the analysis. For an introduction to the Rasch model, see Bond and Fox (2001), Rasch (1960), Smith and Smith (2004), or Wilson (2005).

When complexity level scores are in their raw ordinal form, it is possible to calculate a mean score for each case, to examine the proportions of respondents assigned to each complexity level or transitional phase, and to account for the range of complexity levels represented in a given performance. On the other hand, little can be said about the confidence we can place in these mean scores, the amount of difficulty associated with moving from complexity level to complexity level, or whether the difficulty of making transitions changes depending on where you are in the developmental sequence. Rasch analysis software makes it possible to address these questions by using a log-odds transformation (Wright and Masters, 1982) to convert ordinal data into distinct quantitative estimates of (a) item difficulty and (b) person performance, both of which are expressed in the same equal-interval units.

The formulation of the original dichotomous Rasch model can be expressed as:

$$\log_e(P_{ni1}/P_{ni0}) = B_n - D_i,$$

where P_{ni1} is the probability that person n encountering item i is observed in category 1, B_n is the “ability” measure of person n , and D_i is the “difficulty” measure of item i —the point where the highest and lowest categories of the item are equally probable. This model has been extended to specify the ploytomous partial credit model (Masters, 1982), which can be expressed as:

$$\text{Partial credit: } \log(P_{nij}/P_{ni(j-1)}) = B_n - D_i - F_{ij} = B_n - D_{ij},$$

where P_{nij} is the probability that person n encountering item i is observed in category j , B_n is the “ability” measure of person n , D_i is the “difficulty” measure of item i —the point where the highest and lowest categories of the item are equally probable, D_{ij} is the difficulty measure of item i , category j , and F_{ij} is the “calibration” measure of category j relative to category $j-1$, the point where categories $j-1$ and j are equally probable relative to the measure of the item.

The product of a Rasch analysis is an equal-interval scale, along which both item difficulty and respondent performance estimates are arranged. Each unit on the scale is referred to as a logit, each of which represents an

identical increase in difficulty. The range of difficulty represented in the items of a scale determines the number of logits.

All Rasch analysis software packages provide error terms for all item and person estimates, establishing the confidence one can place in them. Items and performances with more missing data are associated with larger error terms than those with less missing data, and performances that are predominantly at a single level are associated with larger error terms than performances that include a mixture of levels.⁸ Estimates for cases with missing data are biased toward the mean.

The fit statistics included in the following analysis are called infits. Infit statistics are used to assess whether a given performance (or item) is consistent with other performances (or items). They are based on the difference between observed and expected performances. Infits near 1 are desirable. Z-scores are calculated to assess the significance of both positive and negative divergences of infit statistics from 1. Z-scores between -2.0 and +2.0 are considered acceptable.

The analysis presented below was conducted with the computer program, WINSTEPS (Linacre and Wright, 2000).⁹

Results

Item analysis

Because my hypothesis is concerned only with the item estimates, only item fit statistics are reported here. The overall item separation reliability of the item estimates is .88. Infit statistics for all item category estimates fit the model with z-scores between -1.3 and .4. Table 1 displays the item fit statistics.

Figure 4 shows the item and person map from the partial credit analysis. The logit scale in the center of the figure spans 16 logits. Person performance estimates are shown on the left of the figure and item estimates (category estimates) are shown on the right. Category names denote the name of the interview probe and the developmental level represented by the category. Confidence intervals for each item category are included in parentheses. Note that transitional item estimates (SA/AM) cluster with the item estimates at the level into which they are transitioning (AM) and are statistically significantly distinct from estimates at the lower level (SA), in that there is no overlap between the confidence intervals of items in these two clusters. Moreover, the category probability curves for the

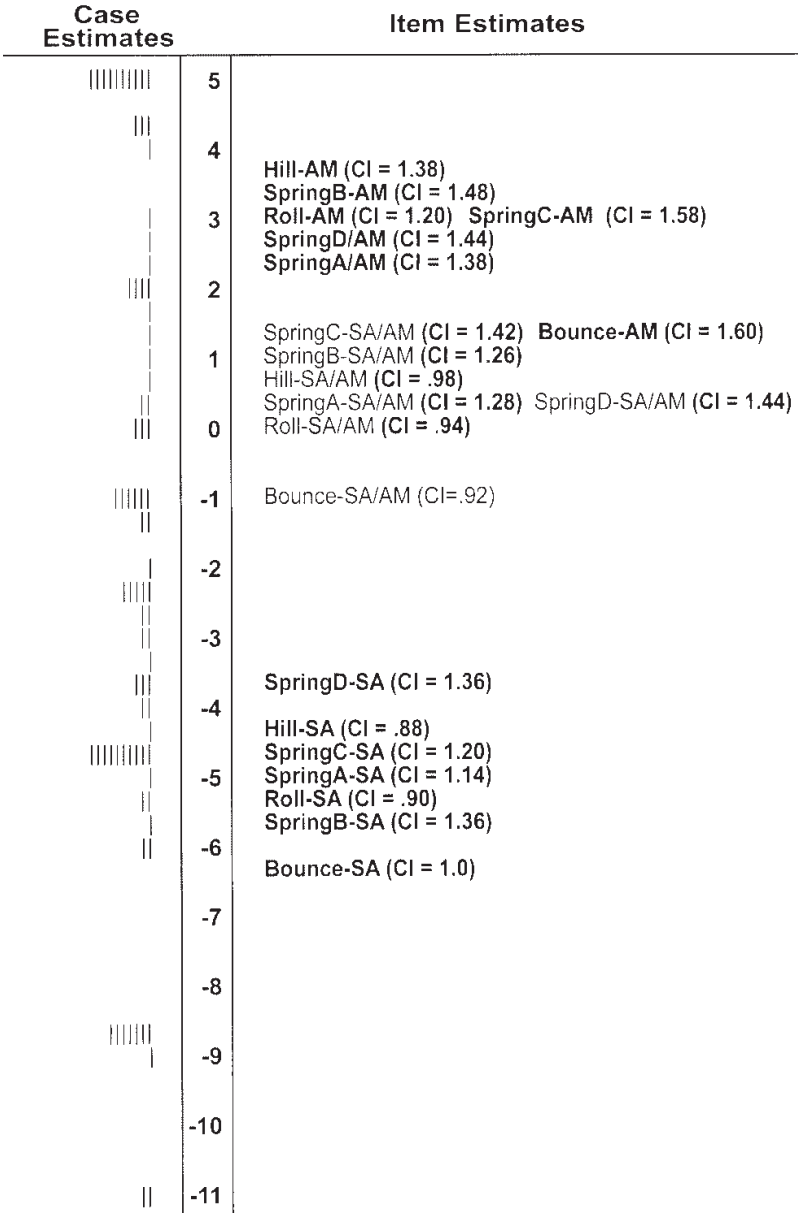


Figure 4. Item and case map ($N = 83$)



Figure 5. Response probabilities for bounce item

Table 1
Item fit statistics

Entry Number	Raw Score	Count	Measure	Error	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD
5 roll	207	66	.23	.261	.04	.21	.35	1.1
2 spring2	122	37	.56	.341	.09	.31	.18	.4
4 spring4	88	26	.86	.35	.93	-.31	.09	.3
3 spring3	103	32	.83	.341	.05	.2	.79	-.5
1 spring1	127	38	.24	.321	.01	.0	.79	-.6
7 bounce	226	67	-1.05	.26	.98	-.1	.92	-.3
6 hill	181	61	-1.67	.26	.77	-1.4	.75	-1.3

items reproduce the pattern shown in Figure 2. As can be seen in Figure 5, the probability curve for the SA/AM transitional category for the bounce item subsumes the small area in which SA and AM curves overlap. Category probability curves for all of the items are very similar.

Discussion

The pattern of item estimates in Figures 4 and 5 replicates patterns reported in two previous studies in which both transitional and consolidated performances were modeled (Armon and Dawson, 2002; Dawson, 2002b). They provide additional support for the existence of the developmental discontinuities expected by proponents of a stage theory of devel-

opment. More importantly, from the point of view of science education, these patterns highlight the difficulty of moving from one complexity level to another. Clearly, a great deal of consolidation must take place at single abstractions before students are ready to construct abstract mappings. In this paper I examined only a few of the conceptual building blocks coordinated at abstract mappings. These were specific to the energy topic and to the particular questions asked in the interviews. Development from one level to another in any domain requires a wide variety of skills and concepts—far more than were identified in this study. It is in this variety that we may find at least part of the explanation for the difficulty associated with moving from one complexity level to another.

Most of the students in the population we sampled were functioning at single abstractions in the energy domain, yet their teachers and their textbooks assumed they were capable of constructing abstract mappings. The needs of these students were not being met by this curriculum. Only students who demonstrated highly elaborated single abstractions or some abstract mappings prior to instruction were able, following instruction, to describe energy transformations or to understand energy as the ability to do work. About half of the students performing at single abstractions demonstrated some learning, but this primarily took the form of a greater elaboration of single abstractions—the addition of new concepts at the single abstractions level. Unfortunately, many of these new concepts were incorrect or shallow (as compared to more fully formed single abstractions constructions¹⁰). This is not surprising given that very little classroom instruction was devoted to conceptual learning.

Describing the pathways through which important concepts typically develop is necessary if we wish to design curricula that meet the needs of all students. A fuller understanding of the prerequisites for constructing important physics concepts is possible, but will require extensive research into the development of these concepts over a wide range of the developmental continuum, along with a coordinated effort to integrate results from studies conducted by different groups of researchers. Educators, researchers, curriculum developers, and funding agencies will have to work closely with one another to carry out the iterative cycles of research, implementation, and assessment required to construct an adequate understanding of how we can best support any student's learning. The tools are available. The LAS makes it possible to (1) describe developmental sequences reliably and accurately and (2) evaluate the developmental progress of individual students. Modeling tools like Rasch analysis provide insight into

developmental processes and yield important information about the learning needs of students performing within different developmental levels. Computers provide the means for evaluating students' developmental progress and distributing customized curricula to every classroom teacher. It is time to take a less piecemeal approach to science education. The time has come to concentrate our efforts on the task of gaining a comprehensive understanding of conceptual development in the sciences, so that we can design curricula and assessments that will optimize learning for every student.

Footnotes

^{1,2} Here, I employ the word *abstract* to describe a level of abstraction that generally appears for the first time sometime between the ages of 8 and 10. See Footnote 5 for a more detailed explanation.

³ The criterion of no regressions is often added to the invariant sequence requirement. However, some dynamic systems models of developmental processes predict regressions at stage transitions. These have been identified longitudinally .

⁴ The Lectical™ Assessment web site contains a wealth of information on scoring and methodology, as well as numerous descriptions of developmental sequences in a variety of knowledge domains.

⁵ The word *abstraction* as used in the term *hierarchical order of abstraction* refers to the way in which conceptions increase in generality over the course of development. The concepts that occur for the first time at the single abstractions complexity order are abstract in a more particular sense; the new conceptions of this complexity order are defined in terms of qualities that are increasingly detached from the concrete.

⁶ Note that this respondents' understanding of conservation of energy is inaccurate because she fails to conceptualize the system in which energy is conserved. Instead, she focuses exclusively on the ball and its movement, claiming that the energy of the ball itself must remain constant.

⁷ As of January 2004, certified LAS analysts must maintain an agreement rate of 85% within 1/3 of a complexity level with a Certified Master Analyst (Dawson-Tunik, 2005).

⁸ In a sense, individuals that are scored predominantly at a single level provide less information than individuals scored at a mixture of levels. Consequently, consolidated performances are associated with larger error

terms in the same way that performances with fewer data points are associated with larger error terms.

⁹ In keeping with the original formulation of the Rasch model, Winsteps treats person parameters as fixed effects. It has been argued that this limitation of the model restricts the generalizability of the results of Rasch analyses (Bartholomew and Knott, 1999; Mislevy and Wilson, 1996), though the specific implications for research of the present kind are not entirely clear. Moreover, several researchers employ Winsteps and other software that treats person parameters as fixed effects to explore developmental constructs similar to those examined here (Bond and Fox, 2001; Dawson, 2002b; Müller, Sokol, and Overton, 1999). In any case, concerns about generalizability are minimized in the present project by the large size and heterogeneity of the sample (Canadian Christian families, boys from New England private schools, working class mid-western families, California elementary school students, and a convenience sample from all over the United States).

¹⁰ Fully “correct” constructions are not possible at this level, since an “complete” understanding of the energy concept requires more elaborated concepts than are possible at this level as well as the ability to coordinate these concepts in systems of abstractions. This becomes possible at the abstract systems level.

References

- American Association for the Advancement of Science, National Science Teachers Association. (2001). *Atlas of science literacy*. Washington, DC: Author.
- Andrich, D., and Styles, I. (1994). Psychometric evidence of intellectual growth spurts in early adolescence. *Journal of Early Adolescence*, 14, 328-344.
- Armon, C. (1984). Ideals of the good life and moral judgment: Ethical reasoning across the lifespan. In M. Commons, F. Richards, and C. Armon (Eds.), *Beyond formal operations, Volume 1: Late adolescent and adult cognitive development* (pp. 357-381). New York: Praeger.
- Armon, C., and Dawson, T. L. (1997). Developmental trajectories in moral reasoning across the lifespan. *Journal of Moral Education*, 26, 433-453.
- Armon, C., and Dawson, T. L. (2002). The good life: A longitudinal study of adult value reasoning. In J. Demick and C. Andreoletti (Eds.), *Handbook of Adult Development* (pp. 271-300). New York: Plenum Press.
- Bartholomew, D. J., and Knott, M. (1999). *Latent variable models and factor analysis*. London: Oxford University Press.

- Bond, T. G., and Fox, C. M. (2001). *Applying the Rasch model: Fundamental measurement for the human sciences*. Mahwah, NJ: Lawrence Erlbaum.
- Case, R. (1985). *Intellectual development: Birth to adulthood*. New York: Academic Press.
- Case, R. (1991). *The mind's staircase: Exploring the conceptual underpinnings of children's thought and knowledge*. Hillsdale, NJ: Lawrence Erlbaum.
- Case, R., Griffin, S., McKeough, A., and Okamoto, Y. (1992). Parallels in the development of children's social, numerical, and spatial thought. In R. Case (Ed.), *The mind's staircase* (pp. 269-284). Mahwah, NJ: Lawrence Erlbaum.
- Colby, A. (1981). *Invariant sequence and internal consistency in moral judgment*. Harvard University, Cambridge, MA. Graduate School of Education.
- Commons, M. L., Danaher, D., Miller, P. M., and Dawson, T. L. (2000, June). *The Hierarchical Complexity Scoring System: How to score anything*. Paper presented at the Annual meeting of the Society for Research in Adult Development, New York.
- Dawson, T. L. (2002a). A comparison of three developmental stage scoring systems. *Journal of Applied Measurement*, 3, 146-189.
- Dawson, T. L. (2002b). New tools, new insights: Kohlberg's moral reasoning stages revisited. *International Journal of Behavioral Development*, 26, 154-166.
- Dawson, T. L. (2003). A stage is a stage is a stage: A direct comparison of two scoring systems. *Journal of Genetic Psychology*, 164, 335-364.
- Dawson, T. L. (2004). Assessing intellectual development: Three approaches, one sequence. *Journal of Adult Development*, 11, 71-85.
- Dawson, T. L., and Gabrielian, S. (2003). Developing conceptions of authority and contract across the life-span: Two perspectives. *Developmental Review*, 23, 162-218.
- Dawson, T. L., Xie, Y., and Wilson, M. (2003). Domain-general and domain-specific developmental assessments: Do they measure the same thing? *Cognitive Development*, 18, 61-78.
- Dawson-Tunik, T. L. (2004). "A good education is..." The development of evaluative thought across the life-span. *Genetic, Social, and General Psychology Monographs*, 130(1), 4-112.
- Dawson-Tunik, T. L. (2005, 06/1/06). The Lectical™ Assessment System. Retrieved December, 2005, from <http://www.lectica.info>
- Dawson-Tunik, T. L., Commons, M. L., Wilson, M., and Fischer, K. W. (2005). The shape of development. *The International Journal of Cognitive Development*, 2(2), 163-196.

- Dawson-Tunik, T. L., and Stein, Z. (manuscript submitted for publication). "It has bounciness inside!" Developing conceptions of energy.
- Dawson-Tunik, T. L., Wenk, L., and Paulman, V. (2005, 11/01/04). The Collaboration for Excellence in Science Education. Retrieved June, 2005, from <http://cese.hmpshire.edu>
- Demetriou, A., and Efklides, A. (Eds.). (1994). *Intelligence, mind, and reasoning: Structure and development*. Amsterdam, Netherlands: North-Holland/Elsevier Science Publishers.
- Demetriou, A., Platsidou, M., Efklides, A., and Metallidou, Y. (1991). The development of quantitative-relational abilities from childhood to adolescence: Structure, scaling, and individual differences. *Learning and Instruction*, 1, 19-43.
- Draney, K. L. (1996). *The polytomous Saltus model: A mixture model approach to the diagnosis of developmental differences*. Unpublished doctoral dissertation, University of California at Berkeley, Berkeley, CA.
- Driver, R. A. S., Rushworth, P., and Wood-Robinson, V. (1994). *Making sense of secondary science: Research into children's ideas*. New York: Routledge.
- Fischer, K. W. (1980). A theory of cognitive development: The control and construction of hierarchies of skills. *Psychological Review*, 87, 477-531.
- Fischer, K. W., and Bidell, T. R. (1998). Dynamic development of psychological structures in action and thought. In W. Damon and R. M. Lerner (Eds.), *Handbook of child psychology: Theoretical models of human development* (5th ed., pp. 467-561). New York: Wiley and Sons.
- Fischer, K. W., and Bullock, D. H. (1981). Patterns of data: Sequence, synchrony, and constraint in cognitive development. In K. W. Fischer (Ed.), *Cognitive development* (pp. 1-22). San Francisco: Jossey-Bass.
- Fischer, K. W., and Rose, S. P. (1994). Dynamic development of coordination of components in brain and behavior: A framework for theory and research. In G. Dawson and K. W. Fischer (Eds.), *Human behavior and the developing brain* (pp. 3-66). New York: Guilford Press.
- Fischer, K. W., and Rose, S. P. (1999). Rulers, clocks, and non-linear dynamics: Measurement and method in developmental research. In G. Savelsbergh, H. van der Maas, and P. van Geert (Eds.), *Nonlinear developmental processes* (pp. 197-212). Amsterdam: Royal Netherlands Academy of Arts and Sciences.
- Fischer, K. W., and Silvern, L. (1985). Stages and individual differences in cognitive development. *Annual Review of Psychology*, 36, 613-648.
- Flavell, J. H. (1971). Stage-related properties of cognitive development. *Cognitive Psychology*, 2, 421-453.

- Flavell, J. H. (1996) Piaget's legacy. *Psychological Science*, 7(4), 200-203.
- Halford, G. S. (1999). The development of intelligence includes the capacity to process relations of greater complexity. In M. Anderson (Ed.), *The development of intelligence* (pp. 193-213). Hove, UK: Psychology Press.
- Kitchener, K. S., Lynch, C. L., Fischer, K. W., and Wood, P. K. (1993). Developmental range of reflective judgment: The effect of contextual support and practice on developmental stage. *Developmental Psychology*, 29, 893-906.
- Linacre, J. M., and Wright, B. D. (2000). WINSTEPS [Computer Program]. Chicago: MESA Press.
- Liu, X., and 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.
- Lourenco, O., and Machado, A. (1996). In defense of Piaget's theory: A reply to 10 common criticisms. *Psychological Review*, 103, 143-164.
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149-174.
- Mislevy, R. J., and Wilson, M. (1996). Marginal maximum likelihood estimation for a psychometric model of discontinuous development. *Psychometrika*, 61, 41-47.
- Müller, U., Sokol, B., and Overton, W. F. (1999). Developmental sequences in class reasoning and propositional reasoning. *Journal of Experimental Child Psychology*, 74, 69-106.
- Piaget, J. (2000). *Studies in reflecting abstraction* (R. L. Campbell, Trans.). Hove, UK: Psychology Press.
- Piaget, J., and Inhelder, B. (1969). *The psychology of the child*. New York: Basic Books.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen, Denmark: Danish Institute for Educational Research. (Expanded edition, 1980. Chicago: University of Chicago Press.)
- Rest, J. R. (1975). Longitudinal study of the Defining Issues Test of moral judgment: A strategy for analyzing developmental change. *Developmental Psychology*, 11, 738-748.
- Rose, S. P., and Fischer, K. W. (1989). *Constructing task sequences: A structured approach to skill theory* (Instructional Manual). Cambridge, MA: Harvard University.
- Shultz, T. R. (2003). *Computational developmental psychology*. Cambridge, MA: MIT Press.
- Siegler, R. S. (1994). Cognitive variability: A key to understanding cognitive development. *Current Directions in Psychological Science*, 3, 1-5.

- Smith, E. V., and Smith, R. M. (Eds.). (2004). *Introduction to Rasch measurement*. Maple Grove, MN: JAM Press.
- Thomas, H., and Lohaus, A. (1993). Modeling growth and individual differences in spatial tasks. *Monographs of the Society for Research in Child Development*, 58(9), 1-169.
- van der Maas, H. L., and Molenaar, P. C. (1992). Stagewise cognitive development: An application of catastrophe theory. *Psychological Review*, 99, 395-417.
- van Geert, P. (1998). A dynamic systems model of basic developmental mechanisms: Piaget, Vygotsky, and beyond. *Psychological Review*, 105, 634-677.
- Walker, L. J. (1982). The sequentiality of Kohlberg's stages of moral development. *Child Development*, 53, 1330-1336.
- Walker, L. J., Gustafson, P., and Hennig, K. H. (2001). The consolidation/transition model in moral reasoning development. *Developmental Psychology*, 37, 187-197.
- Wilson, M. (1989). Saltus: A psychometric model of discontinuity in cognitive development. *Psychological Bulletin*, 105, 276-289.
- Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Mahwah, NJ: Lawrence Erlbaum.
- Wright, B. D., and Masters, G. N. (1982). *Rating scale analysis*. Chicago: MESA Press.