Elementary Education Online, 7(2), 310-322, 2008.
İlköğretim Online, 7(2), 310-322, 2008. [Online]: http://ilkogretim-online.org.tr

The Role of Abductive Reasoning in Cognitive-Based Assessment

Chong Ho Yu[1] Samuel DiGangi[1] Angel Jannasch-Pennell[1]

ABSTRACT. "Knowing what the students know" recently has become the guiding principle of cognitive-based assessment. In the past, the process of assessment was built around a form of quasi-deductive reasoning, in which test developers deduced items from certain premises, the blue print or the objective list. In this paper, the author(s) discuss how abductive reasoning contributes to cognitive-based assessment in three routes. First, knowing alternative explanations is essential in understanding different levels of conceptions and misconceptions in order to develop the constructs being measured. Second, converse reasoning or reverse engineering applied in an abductive fashion is employed to retrospectively build the student's mental model based on the end product. Third, analogical reasoning in the abductive reasoning mode is indispensable for cognitive modeling.
Key Words: Educational assessment, abductive reasoning, cognitive-based assessment

ÖZ. Geçmişte eğitsel değerlendirme süreci yarı-tümdengelim uslamlaması (quasi-deductive reasoning) yaklaşımına dayanmakta iken, günümüzde, özellikle eğitsel değerlendirmenin geleceğine ilişkin olarak, "öğrencilerin ne bildiklerini öğrenmek" olgusu bilişsel temelli değerlendirmeye yön vermektedir. Bu makalede, bilişsel temelli değerlendirmeye çıkarımsal uslamlama (abductive reasoning) yaklaşımının üç temel açıdan nasıl kaktı sağladığı tartışılacaktır. İlk olarak, ölçülen yapıları tam olarak anlamak için farklı düzeylerdeki doğru ve yanlış algılamaları ortaya koymada alternatif açıklamalar göz önüne alınacaktır. İkincisi, çıkarımsal uslamlama (abductive reasoning) yaklaşımına uygulanan tersine uslamlama açıklanacak ve üçüncüsü, çıkarımsal uslamlama (abductive reasoning) yaklaşımında analojik çıkarsama (analogical reasoning) tartışılacaktır.
Anahtar Sözcükler: Eğitsel değerlendirme, çıkarımsal uslamlama, bilişsel temelli değerlendirme

Since the National Research Council (NRC; 2004) released its report regarding the future of educational assessment, the phrase "knowing what the students know" has become the guiding principle of cognitive-based assessment. In the past, the process of assessment was built around a form of quasi-deductive reasoning, in which test developers "deduced" items from certain "premises," namely, the blueprint or the objective list. This kind of assessment is a "top-down" approach in the sense that it emphasizes mostly on what the test developers think students should know.

One of the common criticisms against exam-based assessment is that exam items are often irrelevant to practical applications in the real world. This problem is especially clear with exams that stress memorization of details and rote learning. Another common pitfall of exam-based assessment is that teachers are interested in using test scores as a screening tool, or a "showroom" of performance. On the student side, the objective of test-taking is gaining a good grade, rather than obtaining feedback for improvement. At the teacher end, the curriculum is tailored to conform to testing requirements. This phenomenon is often known as *teaching to the test*. In some extreme cases, teaching to the test can become *teaching the test*, where teachers learn the specific test items that assess various objectives, and teach those test items only (Ladner & Stone, 2007). This type of top-down approach, which draws teachers' attention to map objectives to exam items, has been questioned by many psychometrcians, such as Robert Mislevy.

[1] Ph.D.s, Applied Learning Technologies Institute, Arizona State University, USA

In the 1990s, Robert Mislevy recognized the importance of simulating real-life experience in assessment and providing diagnostic feedback to examinees. To accomplish this goal, Mislevy (1994) developed the Bayesian Inference Network (BIN) as a tool to capture the mental process of students during task performance. Long before the NRC report, Mislevy had foreseen the value of "knowing what the students know" in assessment by developing cognitive-based assessment. Unlike conventional test construction that is driven by deduction, cognitive-based assessment using BIN, according to Mislevy, tightly integrates abduction, deduction, and induction in test construction. Specifically, BIN builds around deductive reasoning to support subsequent inductive reasoning from realized data to probabilities of states. Yet, abductive reasoning is vital to the process in two aspects. First, abductive reasoning suggests the framework for inductive reasoning. Second, the BIN as a tool for reasoning deductively and inductively within the posited structure requires abduction to reason about the structure. Although Mislevy borrowed many ideas from philosophy of science, his focus was on developing assessment tools rather than elaborating the underlying philosophy of the methodology. To fill this conceptual vacuum, the authors discuss how abductive reasoning contributes to cognitive-based assessment in three ways. First, considering alternative explanations is essential in understanding different levels of conceptions and misconceptions for fully understanding the constructs being measured. Second, converse reasoning or reverse engineering applied in an abductive fashion is employed to retrospectively build the student's mental model based on the end product. Third, analogical reasoning in the abductive reasoning mode is indispensable for cognitive modeling.

In the following sections, the meaning of abduction will be briefly introduced. At first glance, this philosophical discussion will seem foreign to educational researchers. Nevertheless, readers will see its relevancy in the next section, in which applications of abductive reasoning will be illustrated with examples of Evidence-centered design (ECD), which is a form of cognitive-based assessment, and also a current project at Arizona State University.

What is abduction?

Abduction is a reasoning mode aiming to look for patterns in phenomena and suggest hypotheses (Peirce, 1934/1960). Logical reasoning is divided into formal types (symbolic logic) and informal types (critical thinking and plausible reasoning). Unlike deduction and induction, abduction is a type of critical thinking rather than symbolic logic. This process of inquiry can be analogous to exploratory data analysis. In exploratory data analysis, the researcher does not start with a pre-generated hypothesis and then collect data to test the hypothesis. Rather, after observing some surprising patterns, the researcher exploits them in as many ways as possible until a plausible "story" of the data emerges. The role of an exploratory researcher or an abductive inquirer can be compared to that of a detective. A detective does not collect just any information. Instead, he collects evidence and clues related to the central question of the case (Yu, 1994, 2006a). In short, abduction can be boiled down to two key ideas, namely, pattern-recognition and plausible reasoning.

How abduction is different from deduction and induction?

Abduction is vastly different from two other popular reasoning modes, namely deduction and induction. Aristotle is credited as the inventor of deduction (Trundle, 1994), though the pre–Socratics had used rudimentary deductive logics. Deduction is a form of analytic inference based upon logical and mathematical demonstrations. In other words, deduction involves drawing logical consequences from premises. An inference is endorsed as deductively valid when the truth of all premises guarantees the truth of conclusion. It is no wonder Quine (1982) stated that the mission of logic is the pursuit of truth, which is the endeavor to sort out the true statements from the false statements. Hoffmann (1997) further elaborated this point by explaining that the task of deductive logic is to define the validity of one truth as it leads to another truth. Since deduction presupposes the existence of truth and falsity, it confines the conclusion to a dichotomous answer (True/False). But in reality, matters outside the domains of mathematics and logic are not necessarily so clear-cut as "true" or "false."

In the 1600s, the English philosopher Francis Bacon (1620) defined the use of inductive reasoning as drawing conclusions from an exhaustive body of facts. According to Bacon, one should proceed regularly and gradually from one thesis to another based on the generalization of empirical input and particular instances by collecting all relevant data without any presuppositions, so that the generalization is not reached till the last available instance is examined. In other words, each thesis is thoroughly tested by observation and experimentation before the next step is taken. In effect, each confirmed thesis becomes a building block for a higher level concept, with the most generalized thesis representing the last stage of the inquiry. In brief, induction is an inference from observed facts to generalizations.

Goldman (2006) asserts that no scientist has ever been a strict Baconian. First, the scientist would go nowhere if Baconian induction was literally followed, for no one could inductively exhaust all facts. Second, the so-called presupposition-less approach inevitably presupposes that reasoning about nature begins with "input" data that are simply given to the mind in experience without interpretation. Third, it also presupposes the availability of objective relevance criteria. But if the mind is truly passive in reasoning, how do hypotheses arise?

Briefly speaking, deduction is a logical process for seeking certain knowledge, whereas induction is an inferential approach that requires a large amount of data. Often, we do not have the luxury of obtaining "true" knowledge and voluminous data, thus abduction, which relies on plausible reasoning, seems to be a more viable method for guiding our development of assessments. In the following section, three major routes of abductive reasoning will be discussed.

Three routes of abductive reasoning

Seeking alternative explanations

The function of abduction is to look for patterns in surprising phenomena and to suggest plausible hypotheses (Peirce, 1934/1960). The following example illustrates this function:

- The surprising phenomenon, B, is observed.
- But if A were true, B would be a matter of course.
- Hence there is a reason to suspect that A might be the explanation.

By the standard of deductive logic, the preceding reasoning is akin to the fallacy of affirming the consequent. Consider this example. It is logical to assert that "It rains; if it rains, the floor is wet; hence, the floor is wet." But any reasonable person can see the problem in making statements like: "The floor is wet; if it rains, the floor is wet; hence, it rained." Nevertheless, in Peirce's logical framework this abductive form of argument is entirely acceptable, especially when the research goal is to discover plausible explanations for further inquiry (de Regt, 1994). The importance of seeking plausible explanations after seeing the facts is echoed by the prominent mathematician George Polya (1954a) almost half a century later. According to Polya, all human knowledge outside mathematics and logic are built upon conjectures. Chances are, these conjectures are also based on other conjectures. In contrast to demonstrative reasoning in mathematics and logic that yields certainty, many other inferences belong to the realm of plausible reasoning, which is controversial and provisional, such as a lawyer's circumstantial evidence and a historian's documentary evidence. Indeed, a lawyer and a historian have to propose plausible explanations based on thin data *after the facts*.

In order to yield a set of plausible explanations, abduction is usually formulated in the following mode:

- The surprising phenomenon, X, is observed.
- Among hypotheses A, B, and C, A is capable of explaining X.
- Hence, there is reason to pursue A.

It is crucial to note that seeking a promising route for further investigation is *qualitative* in nature. Abductive reasoning does not necessarily start with quantitative data. For example, when Kepler developed his astronomical model, the basic preconceptions of Kepler were very general "hunches" about the nature of motion and forces, and also the basic idea that the Sun is the source of the forces driving the planetary system. Kepler's laws, which followed from these qualitative preconceptions, were later quantitatively confirmed by the variation of the orbital periods of the planets with the distance from the Sun (Myrstad, 2004).

Reverse engineering

Though abductive reasoning seems problematic from a deductive standpoint, it is a legitimate scientific methodology, which appears in what's known as "reverse engineering." Inquiry in evolutionary biology utilizes reverse engineering or converse reasoning because biologists trace back the causal history of evolution given the consequences (existing species and fossil records). A further point is that abductive reasoning sometimes aims at explaining theories. Darwinian Theory is said to be substantiated by abduction in the sense that a network of theories could be mutually supported. In other words, the credibility of one theory is enhanced by being explained by another theory. For instance, the theory of natural selection is partially inspired by Malthusian theory: Given the tension between exponential population growth and limited resources, some mechanism must exist to restore the equilibrium. These two theories appear to be compatible (Kleiner, 2003).

This type of reverse engineering is very similar to "regressive reasoning" or "working backwards" as proposed by Polya (1957). According to Polya, one can start from requirements, assumptions, and facts, and then draw consequences from them, and consequences from the consequences, until one reaches the starting point. Nonetheless, there is a subtle difference between the ideas of Peirce and Polya. Polya developed his notions in the context of problem solving, which involves a two-step process. To solve a problem, the first step, in Polya's view, is to conduct an *end-means* analysis by employing regressive reasoning, as opposed to the popular means-end analysis. When the end-means analysis is converted into action, the components of the execution plan have to be synthesized in a forward order. This execution plan is called "synthesis." Comparing Peirce's and Polya's notions is far beyond the scope of this article. Nonetheless, Peirce is regarded as the originator of converse reasoning for his ideas was developed around late 19th and early 20th centuries, hence the subsequent discussion will center on Peirce's abduction.

At first glance, abduction is an educated guess among existing hypotheses. But unifying conceptions is an important part of abduction, and it would be unfortunate if our understanding of abduction was limited to more mundane cases where hypotheses were simply assembled (Thagard & Shelley, 1997). Abduction does not occur in the context of a fixed language, since the formation of new hypotheses often goes hand in hand with the development of new theoretical terms such as "quark" and "gene." Indeed, Peirce (1934/1960) emphasized that abduction is the only logical operation that introduces new ideas.

For Peirce, progress in science depends on the observation of the right facts by minds furnished with appropriate ideas (Tursman, 1987). Definitely, the intuitive judgment made by an intellectual is different from that made by a high school student. Peirce cited several examples of remarkable correct guesses, however, not all success is simply luck. Instead, the opportunity was taken by the people who were prepared:

(a) Bacon's guess that heat was a mode of motion;

(b) Young's guess that the primary colors were violet, green and red;

(c) Dalton's guess that there were chemical atoms before the invention of microscope (Tursman, 1987).

Analogical thinking

The preceding creative ideas did not arise in a vacuum. This type of educated guessing is often based upon analogical thinking. Hempel (1966) suggested that discovery relies on creative imagination by citing Kekule's discovery of the hexagonal benzene ring: The chemist Kekule failed to devise a structural formula for the benzene molecule in spite of many trials. One evening he found the solution to the problem while watching the dance of fire in his fireplace. Gazing into the flames, he seemed to see atoms dancing in snakelike arrays and suddenly related this to the molecular structure of benzene. This is how the hexagonal ring was discovered. The dance of fire may serve as an analogy to the molecular structure that Kekule had contemplated.

Whether the discovery of the benzene ring is a result of a conscious process or an unconscious one is debatable. Nevertheless, von Baeyer Hans Christian (1989), whose great grandmother, Adolf von Baeyer, was the first research assistant of Kekule, told the story of exactly how the discovery occurred. Long before the well-known vision before the fireplace, Kekule was called by the court to be a witness in a murder trial. Kekule was asked about the suspect who had been caught selling stolen goods, including a gold ring consisting of two intertwined snakes biting their own tails. This image attached to Kekule's mind and he later devoted his lifelong efforts to studying the structure of chemical models. During the process of studying structural chemistry, Kekula found that there are numerous possible combinations of atoms, ranging from a ball-and-chain structure to a compact clump. This ill-defined and unstructured problem is a typical one for abduction. Kekule explored different combinations but to no avail. One night in the early 1850s, Kekule sat outside and stared at the empty streets. He had a vision similar to the later and more famous one inspired by the snake-like fire. By the time his daydream ended, he had developed the structure of methane: a carbon atom. Once the prototypical model was developed, Kekule and Adolf von Baeyer worked diligently to refine the model. The climax of the story occurred in front of fireplace, which led the discovery to consummation.

By citing the example of the benzene ring, Seifert, Meyer, Davidson, Patalano and Yaniv (1995) speculated that the final steps on the road to insight may be subconscious; what innovators consciously know is that the products of these processes often seem natural. However, why didn't others make scientific breakthroughs by observing a fireplace or seeing a gold ring? Obviously, the background knowledge accumulated by Kekule throughout his professional career played a more important role in the discovery of the hexagonal ring than a brief moment in front of a fireplace. Without the deep knowledge of chemistry and the creative use of analogy, it is unlikely that anyone could draw inspiration by the dance of fire. In Seifert et al.'s terms, consciously submerging oneself into relevant information is like going through an "incubation period" (1995).

Hence, analogical abduction, which is a conscious *effort*, plays an important role in proposing new concepts and theories that might introduce causal mechanisms. This is achieved by adopting the pragmatic strategy of conceiving unknown mechanisms in terms of what is known. In actuality, researchers engage in this type of activity all the time: It is so-called modeling. In psychological and educational research where we attempt to understand how the mind works, we call it cognitive or mental modeling.

It is important to point out that constructing a mental model is essentially finding the right analogy. However, the term "analogy" implies that one should not take the model as a *literal truth*. Interestingly enough, Polya (1954b) had also made a similar point about inference from analogy. As mentioned before, Polya realized the fact that very often inferences must be based on conjectures, which might not be factual at all. We could consider the analogy of two theorems, A and B, as the intention to discover a common ground H, from which both A and B would follow: A is implied by H and B is implied by H. But, we do not have H; we just hope there is such an H. According to Polya, we could reach a reasonable conclusion expressed by the following pattern:

- A is implied by H
- B is implied by H
- B is true
- A is more credible.

It is important to point out that following the inference by analogy: Even if B is true, at most, A is more credible, but not true. In a similar vein, the goal of abduction is to propose a promising route for in depth examination rather than describing the world in a realist sense. Mental modeling is subject to revision as the existing analogy faces insurmountable problems. For example, since the invention and popularization of computer technology, cognitive theories have compared human perception, thinking, and decision-making to input, processing and output. Interestingly enough, while cognitive psychologists used “cognitive architecture” in cognitive science through an analogy to computer, earlier scholars introduced the concept of “computer architecture” into computer science through an analogy to the architecture of building (Anderson, 2007). In other words, this line of cognitive research was inspired by an analogy of analogy. It is a typical example of how abductive reasoning, as an exploratory approach, links “irrelevant” things together by innovative analogy.

Anderson (1983, 1990, 1993, 1996, 2007) maintains that our intellectual processes, just like computer programming, follow the rule of logical branching (e.g. if-then-else); he also compares vectorized and recursive procedures in LISP, a programming language for research in artificial intelligence, with human cognition. However, Penrose (1989, 1994, 1997) argues against comparing human reasoning to formal algorithms. He objects to the optimistic view held by the artificial intelligence community that all thinking is computational and thus can be simulated in a computer. Penrose insists that humans know things to be true and build up insights that a computer is incapable of doing, and that correct computation by an automated process does not constitute understanding.

Nonetheless, this and other incorrect analogies should not hinder researchers from employing analogical abduction, because while it is true that human mental processes are not literally syntactical computing processes, some insights arising from this metaphor have indeed been helpful to psychological research. For example, based on this analogy, Anderson found that our mind solves problems by activating many neural nodes and links in the mental schemata (spread of activation), which debunks the myth that our brain works in a compartmentalized fashion (faculty psychology). It is important to repeat the point that the aim of analogical abduction is proposing rather than affirming a causal explanation, and thus researchers must bear the risk of incorrect analogy.

In the following section, how the preceding three routes of abductive reasoning can be applied to assessment will be illustrated with concrete examples.

Applications of abduction

<u>Alternate explanations in distracter analysis</u>

While many view the purpose of assessment as quantifying individuals’ abilities, an alternative perspective is the notion that assessment helps to identify inability. When teachers, parents, and students know where weaknesses lie, they can be more strategic in designing instruction that will affect learning. Assessment of student misunderstanding or misconceptions has gained recent popularity certain academic domains, for example Physics.

Physics and other sciences are an ideal opportunity for conceptual assessment for several reasons. First, a large body of cognitive psychology research exists that documents students’ models and methods of reasoning in the physical domain. This provides an empirical foundation for assessment design and score interpretation. Further, many physics teachers are frustrated with the inability of current physics assessments to assess student knowledge. Physics education is flooded with faulty ideas because it requires sophisticated abstract thinking, yet there is a persistent mismatch between the goals of physics instruction and student assessment. Typically, classroom practices and assessment drive students toward undesirable and short-term goals, such as getting the “model” answer. There is some consensus that a better assessment tool must inform the teacher about what students think, how they think, and why they don’t understand (Dufresne, Gerace, Mestre, & Leonard, 2000). Diagnosing misconceptions is challenging, because a student’s failure to understand a particular physics concept may be caused by one or many possible flawed ideas. A typical abductive methodology may approach this problem as follows:

- A surprising phenomenon occurs: Students cannot understand the routine motions of the heavenly bodies in our Solar system.
- Based on the information collected from qualitative methodology, in addition to our background knowledge, three plausible explanations, A, B, and C, are proposed.

- An item concerning the motions of the celestial bodies is written with three distracters corresponding to three different misconceptions. The correct understanding and the three misconceptions constitute a construct map for this topic.
- A probabilistic model of the construct map is constructed by collecting test data.

In brief, the construct map approach is similar to the ECD approach, except that we model the probability of different kinds of misconception as well as the probability of proper learning. Different approaches could be employed to categorize misconceptions. Interviewing teachers and students is the most straight-forward one, but it is time-consuming and labor-intensive. A more cost-effective approach is to administer exams with two-tier items (Treagust, 1995). In a test consisting of two-tier questions, not only do examinees choose an option from multiple choices, but also they are asked to write down a reason for the response. An interview using follow-up questions could be applied to classify different misconceptions.

The following is a simplified version of the construct map developed by Briggs, Alonzo, Schwab, and Wilson (2006), showing the full and partial understandings regarding the motions of the Earth, Moon and Sun. Level 4 represents correct and complete understanding whereas Level 1 to 3 denote incorrect or incomplete understanding.

Table 1. *Construct map for understanding the motion of the Solar system*

Level	Description
4	Student is able to coordinate apparent and actual motion of objects in the sky.
3	Students know that the earth orbits the Sun, the Moon orbits the earth, and the Earth rotates its axis. However, student has not put this knowledge together with an understanding of apparent motion to form explanations and may not recognize that the earth is both rotating and orbiting simultaneously. The common error is: It gets dark at night because the Earth goes around the Sun once a day.
2	Student recognizes that the sun appears to move across the sky everyday and the observable shape of the Moon changes every 28 days. However, student may misperceive that all motion in the sky is due to the earth spinning on its axis, the Sun travels around the Earth, it gets dark at night because the Sun goes around the Earth once a day.
1	Student does not recognize the systematic nature of the appearance of objects in the sky. Student may not recognize that the earth is spherical. Student may misperceive that it gets dark at night because something covers the Sun, the phrases of the moon are caused by clouds covering the Moon, and The Sun goes below the Earth at night.

In order to substantiate the claim that certain misunderstandings are attributed to certain causes, distracters are written to map different levels of the construct map, as shown in Table 2.

After the exam is given to the students, the data set is analyzed with a proper psychometric algorithm. The partial credit model (PCM) is suitable for this type of data because Level 1 to 4 is a continuum, not dichotomous. It is beyond the scope of this article to unpack the algorithm of PCM and thus a brief introduction will be given instead. Interested readers can consult Yu (2006b, 2007). Simply put, students who gave the correct answer (D) should be awarded with full credit (3 points) whereas students who gave incorrect answers but demonstrated partial understanding should receive partial credit (A → 2 points; B → 1 point).

But, students who have no clue at all should get no credit (C → 0 points). At the end, the computer program will yield a step function to indicate how difficult it is for students to move from Level 1 to Level 2, how hard it is to gain improvement from Level 2 to 3, and so on.

Table 2. *Sample item of the motions of the Solar System*

It is most likely colder at night because:		
A	The earth is at the furthest point in its orbit around the Sun.	Level 3
B	The Sun has traveled to the other side of the Earth.	Level 2
C	The Sun is below the earth and the moon does not emit as much heat as the Sun.	Level 1
D	The place where it is night on earth is rotated away from the Sun.	Level 4

Figure 1 shows a hypothetical probabilistic model of the above construct map. Given that the skill level, measured by the estimated person theta based upon the performance in the entire test, is 3, the probability of reaching Level 4 understanding is .7. If the skill level is -2, the probability of answering this item correctly is .2. At the same skill level (-2), the probability of committing Level 3 error is .8. It suggests that there is a strong causal link between Level 3 misconception and failure of understanding why the weather is cold is night.

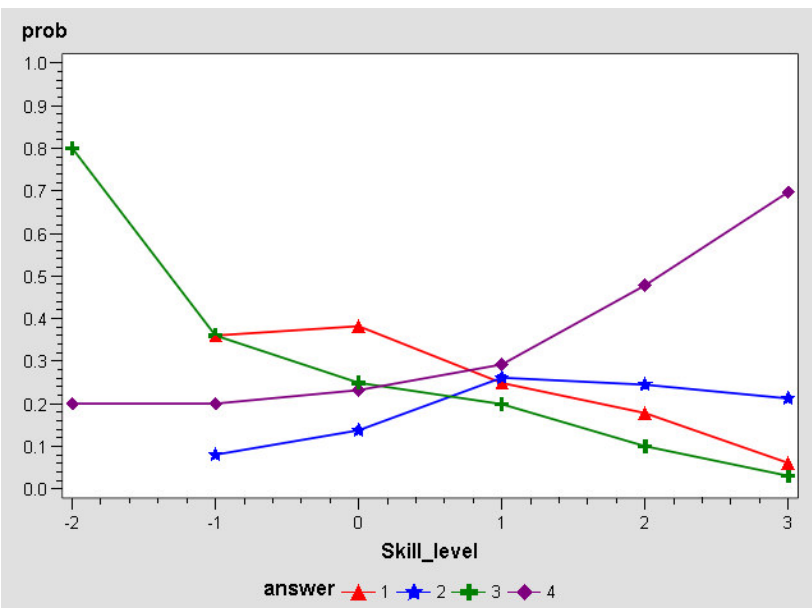

Figure 1. Hypothetical probabilistic model of construct map.

Reverse engineering in Evidence-Centered Design

An application of abduction can be found in the use of Evidence-Centered Design (ECD), which is a form of cognitive-based assessment (Behrens, Mislevy, Bauer, Williamson, & Levy, 2004; Mislevy, 1994; Williamson, Bauer, Steinberg, Mislevy, Behrens, & DeMark, S, 2004). In ECD, test development does not inherit a pre-determined blueprint of the test and make inferences about the test and the examinees based on test scores. Rather, the central questions are: what evidence can be provided by the work product of the examinee to substantiate the claim that he is capable of performing the job? How can the evidence be best measured?

Abductive reasoning is vital to ECD, as it suggests what constructs should be included or excluded, and how the constructs should be structured as a causal process. ECD is very different from conventional test development, in which constructs are pre-defined in the objectives listed in the syllabus. In the conventional assessment approach, test development is a top-down process in the sense that the item authors write items to "materialize" the abstract concepts. Table 3 shows one of the content standards of California Grade 4 Physical Sciences (California Department of Education, 2007):

Table 3. *Grade 4 Physical Sciences Standards*

4PS1.	Electricity and magnetism are related effects that have many useful applications in everyday life. As a basis for understanding this concept:
4PS1.a.	Students know how to design and build simple series and parallel circuits by using components such as wires, batteries, and bulbs.
4PS1.b.	Students know how to build a simple compass and use it to detect magnetic effects, including Earth's magnetic field.
4PS1.c.	Students know electric currents produce magnetic fields and know how to build a simple electromagnet.
4PS1.d.	Students know the role of electromagnets in the construction of electric motors, electric generators, and simple devices, such as doorbells and earphones.
4PS1.e.	Students know electrically charged objects attract or repel each other.
4PS1.f.	Students know that magnets have two poles (north and south) and that like poles repel each other while unlike poles attract each other.
4PS1.g.	Students know electrical energy can be converted to heat, light, and motion.

In the conventional approach, it is usually required that there are adequate items in a comprehensive test to cover every objective. But, the items are typically presented in *isolation* and very little attention is paid to the relationships between these objectives in the perspective of cognitive psychology. This conventional approach is problematic because no causal link is established between the capability of earning high scores in the exam and the conceptual comprehension of scientific applications. Even if examinees can identify the characteristics of each electromagnetic phenomenon, it does not necessarily mean that they can synthesize the related concepts and understand how a Maglev train works when they see one in Europe, Japan or China.

In ECD, test development does not start from a list of exam objectives; rather, the central question should be "what evidence can be provided by the work product of the examinee to substantiate the claim that he understands the constructs as an integrated entity?" The first stage of test construction is abductive in essence because converse reasoning is involved. In this stage, the end result (e.g. how a Maglev train works) is presented to a content expert. Then, the subject matter expert is asked to reason about the task process in a backward fashion: "What concept is required to successfully deliver the product? Which cognitive component is missing if the result is not desirable? What is the required concept just one step before the previous component?" The content expert may explain, "A Maglev, which stands for magnetically levitating train, is a form of mass transit system that suspends, guides and propels trains using electromagnetic forces." The test developer may probe the expert for more elaboration by asking, "How is electromagnetic force used in Maglev?" The expert may say, "One way to do it is to use the attractive magnetic force of a magnet beneath a rail to lift the train up. Another way is to use a repulsive force between two magnetic fields to push the train away from the rail. What it requires is the concepts of attractive and repulsive forces of magnets." The test developer can keep asking questions to establish the causal links between concepts, such as how maglev generates the attractive and repulsive forces. In other words, a causal map, which depicts how the mental processing of one concept leads to another, is conjectured in a qualitative manner, which is important to testing and assessment because educators should be concerned with the mental process of getting the answer and the constructs in each step rather than just the final answer.

Next, items are written to map the constructs of each task, test procedures are designed, and data are collected to verify this causal network. Instead of asking examinees to identify characteristics of each component in the content standard, test developers use scenario-based items to collect evidence of the examinee knowledge level. In this example, a maglev train is presented to the test taker and the examinee will be asked to explain how it works. This approach can avoid the pitfalls of so-called "teaching to the test" mentioned in the beginning of this article.

The test scores collected from these scenario-based items are used as evidence to substantiate the claims about the construct map, and thus this approach is named Evidence-Centered Design. As more and more examinees take the exam, the percentage of correct item responses and other numeric-based data can be utilized to estimate the probability of answering an item correctly conditioning on competence in another concept, which is manifested by performance on another item. For example, if 100,000 students demonstrate that they understand 4PS1.f., and 95,000 out of these 100,000 students can also comprehend 4PS1.e., then the model suggests that given competence in 4PS1.f., the probability of succeeding in 4PS1.e. is .95. If 85,000 out of those 95,000 students who understands 4PS1.e. can successfully explain how a maglev works, then we know that given the competency in 4PS1.e, the probability of delivering the final product is .89. It is important to point out that these probabilistic models are subject to revision as more and more data are accumulated in an inductive fashion.

Usually a traditional test simply gives a unitary score that indicates where a person may fall relative to a norm group, with no information regarding content knowledge. But based on the probabilistic models, ECD users can perform cognitive diagnosis. Specifically, unlike conventional assessment methodologies that can only classify students into "masters" and "non-masters" of the subject matter, instructors who use ECD can conjecture which concepts the failed examinee may miss given the consequence.

Computing analogy and cognitive-based assessment

As mentioned in the previous section, cognitive psychologists have been using computer programming as a metaphor to human mental structure and process. Anderson's ACT-R is one of the prominent cognitive models based on this analogy. It is important to emphasize that the authors of this article by no means endorse ACT-R as the best cognitive model, nonetheless, test developers can definitely benefit from the rich findings contributed by ACT-R researchers. According to ACT-R, there are two types of knowledge, namely, declarative and procedural. The former is conceptual knowledge while the latter is about "how-to." The connection between these two types of knowledge takes place through working memory, also known as short-term memory, which is compared to RAM in a computer system. A mental structure consists of nodes and links. Information stored in nodes and problem solving requires establishing links between nodes. An authentic problem solving skill could not be developed if weak links are built based on short term memorization. Unfortunately, most poorly written items are nothing more than a test of short term retention of the content. This mistake is like writing a paper on a word processor but shutting down the computer abruptly without saving the document. To ensure that the test taker has built a powerful problem solving skill that integrates declarative and procedural knowledge, *the instruction and the assessment* must be designed to enable the student to build strong links based on long-term memory. Take computing as a metaphor again. The student must store the information in his "hard disk drive." At first glance, this approach seems to be common sense. But, the insight is that learning and assessment can take place at the same time in the sense that not only can a carefully designed cognitive-based exam report test scores reflecting students' actual ability, but also it can enhance students' long term retention of the subject matter.

Conventional wisdom encourages instructors to teach the most relevant materials so that students can learn the content in the most efficient fashion. Interestingly enough, cognitive psychologists suggest that in contrast to conventional wisdom, redundant links connecting similar information in redundant nodes actually enhance our problem solving skills. By increasing the number of nodes and links in the long-term memory, experts are able to solve problems rapidly through pattern recognition (Gobet, 1996). The metaphor of a computer network is helpful to illustrate the redundancy approach of cognition. A point-to-point switching network is very vulnerable because a single failure might render the entire network inoperable. However, when there are redundant routes, packets of information can always be available and transportable.

Consider the following sample items in California Grade 6 Math Standards Test (California Department of Education, 2007):

Item 11. The vice president of sales took a client out to lunch. If the lunch was $44 and she gave a 20% tip, how much money did she spend on lunch?

A $8.80
B $35.20
C $52.80
D $53.80

Item 12. If 50% of a number is 20, what is 75% of the number?

A 8
B 15
C 30
D 45

Item 13. What is 60% of 30?

A 1.8
B 18
C 180
D 1800

Item 14. The original price of a new bicycle is $138.00. If the bicycle is marked down 15%, what is the new price?

A $20.70
B $117.30
C $123.00
D $153.00

Obviously, the preceding items are all about computing proportions. Item 11 and 14 are presented in specific contexts whereas Item 12 and 13 are presented in an abstract, decontextualized fashion. Typically, students complain about long exams because similar items pertaining to the same content recur over and over, but not only can this type of long exam improve reliability in terms of classical true score theory, it can also help students build redundant nodes and links in their cognitive structure by encoding information in both contextualized and decontextualized means (Anderson, Reder, & Simon, 2000).

It is worth repeating that although the computing analogy has led to fruitful results in research of cognitive psychology and its practical applications to cognitive-based assessment, it is by no means the only viable analogy. The very essence of abductive reasoning is to encourage inquirers keep an open mind to potentiality of other analogies.

Conclusion and Future Directions for Assessment Design

Abductive reasoning is no doubt beneficial to educators and test developers in developing cognitive-based assessment. Seeking alternate explanations to a surprising phenomenon leads item authors to contemplate the root causes of misconceptions whereas converse reasoning focuses on the final product and the cognitive processes that deliver the outcome. Although the objective of analogical reasoning is not concerned with the literal truth of human mental structure, cognitive psychologists find it useful in producing insightful findings and practical recommendations with regard to instruction and assessment.

However, the authors of this article are not asking readers to put aside deduction and induction. In actuality, both the ECD and construct map approaches are considered an integration of abduction, deduction, and induction. First, qualitative-oriented methodologies are used in an abductive fashion to capture the cognitive process of subject matter experts for correct understanding and that of students for faulty concepts. Next, a suitable psychometric algorithm is developed or chosen in a deductive mode. Last, a probabilistic model is inductively calibrated as more and more empirical data are accumulated.

While it is important to conduct abduction to understand how subject matter experts operate with correct constructs, it is equally important, if not more important, to understand how novices and struggling students are influenced by misconceptions. To address this issue, the Applied Learning Technologies Institute (ALT-I) at Arizona State University, USA is developing a construct-map for unveiling the causal relationship between cognitive flaws and failures in learning physics (Yu, DiGangi, Jannasch-Pennell, & Gorin, 2007). A multi-disciplinary team of researchers was assembled to contribute to the diagnostic assessment system design. Faculty with expertise in test design and cognitive theory from the Division of Psychology in Education provided a framework for educational diagnostic testing. Researchers from ALT-I, including programmers and assessment specialists, worked with existing educational software to design the necessary system components. Finally, faculty and graduate researchers from the Physics department provided content area expertise regarding student learning and misconception. The efficacy of cognitive-based assessment and abductive reasoning that guides the development process will be examined when more data are collected.

Acknowledgments

Special thanks to Dr. Joanna Gorin, Dr. Brad Arment, Dr. Samuel Green, Dr. Bernard Kobes, Dr. Ibrahim H. Diken, Mr. Charles Kaprolet, and Ms. Joy Gibson for their valuable input in this article.

References

Anderson, J. R. (1983). *Cognitive psychology and its implications*. San Francisco, CA: W. H. Freeman and Company.

Anderson, J. R. (1990). *The Adaptive Character of Thought*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Anderson, J. R. (1993). *Rules of the mind*. Hillsdale, N.J.: L. Erlbaum Associates.

Anderson, J. R. (1996). ACT: A simple theory of complex cognition. *American Psychologist, 51*, 355-365.

Anderson, J. R. (2007). *How can the human mind occur in the physical universe?* Oxford: Oxford University Press.

Anderson, J. R., Reder, L. M., & Simon, H. A. (2000). Applications and misapplications of cognitive psychology to mathematics instruction. *Texas Education Review, 1*, 29-49.

Bacon, F. (1620/1960). *The new organon, and related writings*. New York: Liberal Arts Press.

Behrens, J. T., Mislevy, R.J., Bauer, M., Williamson, D.M., & Levy, R. (2004). Introduction to evidence centered design and lessons learned from its application in a global E-Learning program. *International Journal of Measurement, 4*, 295-301.

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

California Department of Education. (2007). 2003 - 2006 CST Released Test Questions. Retrieved January 7, 2008 from http://www.cde.ca.gov/ta/tg/sr/css05rtq.asp

de Regt, H. (1994). *Representing the world by scientific theories: The case for scientific realism.* Tilburg, The Netherlands: Tilburg University Press.

Dufresne, W., Gerace, W., Mestre, J., & Leonard, W. (2000). *Accessing student knowledge with instructional technology*. Technical report PERG-2000#09-SEP. Boston, MA: University of Massachusetts.

Gobet, F. (1996). Discrimination nets, production systems and semantic networks: elements of a unified framework.

Proceedings of the 1996 International Conference on Learning sciences, Bloomington, IN, 398-403.
Goldman, S. L. (2006). *Science wars: What scientists know and how they know it* [CD-ROM]. Chantilly, VA: Teaching Company.
Hempel, C. G. (1966). *Philosophy of natural science.* Englewood Cliffs, NJ: Prentice–Hall.
Hoffmann, M. (1997). *Is there a logic of abduction?* Paper presented at the 6th congress of the International Association for Semiotic Studies, Guadalajara, Mexico.
Kleiner, S. (2003). Explanatory coherence and empirical adequacy: The problem of abduction, and the justification of evolutionary models. *Biology and Philosophy 18*, 513–527.
Ladner, M., & Stone, G. (2007, June). A test of credibility: NAEP versus TerraNova test score results in Arizona (Technical report 07-04). Phoenix, AZ: Goldwater Institute.
Mislevy, R. (1994). Evidence in educational assessment. *Psychometrika, 59,* 439–483.
Myrstad, J. A. (2004). The use of converse abduction in Kepler. *Foundations of Science, 9*, 321-338.
National Research Council. (2004). *Knowing what students know: The science and design of educational assessment.* Washington DC: National Academic Press.
Peirce, C. S. (1934/1960). *Collected papers of Charles Sanders Peirce.* Cambridge: Harvard University Press.
Penrose, R. (1989). *The emperor's new mind: Concerning computers, minds, and the laws of physics.* Oxford: Oxford University Press.
Penrose, R. (1994). *Shadows of the minds: A search for the missing science of consciousness*. Oxford: Oxford University Press.
Penrose, R. (1997). *The large, the small, and the human mind*. Oxford: Oxford University Press.
Polya, G. (1954a). *Mathematic and plausible reasoning -- Volume I: Induction and analogy in mathematics.* Princeton: Princeton University Press.
Polya, G. (1954b). *Mathematic and plausible reasoning -- Volume II Patterns of plausible inference.* Princeton: Princeton University Press.
Polya, G. (1957) *How to solve it: A new aspect of mathematical method*. Princeton: Princeton University Press.
Quine, W. V. (1982). *Methods of logic*. Cambridge, MA.: Harvard University Press.
Seifert, C. M., Meyer, D. E., Davidson, N., Patalano, A. L., & Yaniv, I. (1995). Demystification of cognitive insight: Opportunistic assimilation and the prepared-mind perspective. In R. J. Sternberg, & J. E. Davidson (Eds), *The nature of insight* (pp. 65-124). Cambridge, MA: MIT Press.
Thagard, P., & Shelley, C. (1997). Abductive reasoning: Logic, visual thinking, and coherence. In M. Chiara (Eds), *Logic and scientific methods* (pp. 413–427). Dordrecht: Kluwer.
Treagust, D. F. (1995). Diagnostic assessment of students' science knowledge. In S. M. Glynn & R. Duit. (Eds), *Learning science in the schools: research reforming practice* (pp. 327-346). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
Trundle, R. (1994). *Ancient Greek philosophy: Its development and relevance to our time*. Brookfield: Avebury.
Tursman, R. (1987). *Peirce's theory of scientific discovery: A system of logic conceived as semiotic.* Indianapolis, IN: Indiana University Press.
von Baeyer, H. C. (1989, Jan/.Feb). A dream come true. *Sciences, 29*(1), 6-8.
Williamson, D.M., Bauer, M., Steinberg, L.S., Mislevy, R.J., Behrens, J.T., & DeMark, S. (2004). *Design rationale for a complex performance assessment*. International Journal of Measurement, 4, 303-332.
Yu, C. H. (2006a). SAS programs for generating Winsteps control files and Web-based presentations. *Applied Psychological Measurement, 30*, 247-248.
Yu, C. H. (1994, April). *Induction? Deduction? Abduction? Is there a logic of EDA?* Paper presented at the Annual Meeting of American Educational Research Association, New Orleans, LA. (ERIC Document Reproduction Service No. ED 376 173)
Yu, C. H. (2006b). *Philosophical foundations of quantitative research methodology*. Lanham, MD: University Press of America.
Yu, C. H. (2007). Step functions in Winsteps. Retrieved January 8, 2008 from http://www.creative-wisdom.com/computer/sas/sf.htm
Yu, C. H., DiGangi, S., Jannasch-Pennell, A., & Gorin, J. (2007 October). *A multidisciplinary approach to cognitive-based assessment*. Paper presented at the Annual Meeting of EDUCAUSE. Seattle, WA.