Assessment of Thinking in Adult Learners

Darlene Eleanor Crone-Todd

Salem State College

Higher-order thinking is one of the defining characteristics of higher education, yet the rating of such behavior has been neither reliable nor valid. Recent research involving the use of categories has yielded relatively high inter-rater reliability, with nominal categories of behavior identified. The current study applies an ordinal-scale model of hierarchical complexity (MHC) to educational tasks. The initial results suggest that the MHC can be used to score questions with varying degrees of complexity and support, and that future research should concentrate on determining the reliability and validity of this scoring system for educational purposes.

In higher adult education, the attainment of higher-order thinking (HOT) skills is an expected, integral part of a liberal arts education (Halpern, 1998). However, the reliability and validity of HOT skills is not well established (Williams, 1999). The current state of the research suggests that there are many different ways to define and assess HOT skills, such as "reasoned argumentation" (Newman, 1991a, 1991b), comparison of similar elements (Carnine, 1991), application (Hohn, Gallagher, & Byrne, 1990; Semb & Spencer, 1976), or effective judgments (Paul & Heaslip, 1995) within a discipline. Despite the variety of such summary labels and definitions, so far little has been done to operationally define and rate the complexity of both the educational task, and the type of response provided by students.

Behavioral Research in Higher Order Thinking

Early categorical assessment

Within the last few decades, there has been movement on the part of behavior analysts to find more reliable methods of evaluating higher-order thinking. First, Semb and Spencer divided the behaviors into recall (point-to-point correspondence to media) versus more complex tasks. The more complex tasks included problem-solving in which students either had to identify a principle or concept in a novel example, or generate a novel example of a principle or concept. Johnson and Chase also broke down the lower-level and higher-level thinking behavior into elementary versus conceptual, respectively. The Elementary category was considered lower-level, and the conceptual category was higher-level that seemed to again require students to apply principles and concepts in a novel way through the use of examples. In both cases, the researchers were able to obtain high levels of inter-rater agreement. This is perhaps a testament to having two dichotomous, nominal, categories of behavior, in which the levels are clearly far apart: either the answer is already in the media presented, or it is not. Such a definition would make scoring much more reliable. However, finer discriminations between levels of higher-order thinking are needed to more precisely state goals and objectives involved in higher education.

Recent categorical assessment

In the past decade, there has been some attempt to now provide more precise definitions for the values between the two extreme categories. The first line involves attempting to modify Bloom's (1956) Taxonomy of Educational Objectives in the Cognitive Domain. The taxonomy has six general categories (see Table 1), as modified by Joseph Pear, myself, and several of our colleagues at the University of Manitoba (Crone-Todd, Pear & Read, 2000; Pear, Crone-Todd, Wirth, & Simister, 2001).

Table 1. Lower- and Higher-Level Categories in Higher Order Thinking

Lower-Level: Answers found in Instructional Media	Higher-Level: Going Beyond Instructional Media
Rote Manner Knowledge	Application
Comprehension	Analysis
	Synthesis
	Evaluation

This taxonomy is further delineated by two major categories: (a) based on material presented (either in lecture, text, or other media), and (b) based on going beyond the material presented. The first category consists of material that can be learned in a rote manner (Knowledge), or put in one's own words (Comprehension). The second has four sub-categories. The first is Application, in which an original example is provided, or a particular concept, principle, or procedure is identified in a new example. The second is Analysis, which involves comparison and contrast, including comparing an example to a definition. The third is Synthesis, which involves more complex application (i.e., combining several principles, concepts, or procedures) to solve a unique problem, or classifying several examples into classes. Finally, Evaluation involves all of the previous categories, in a manner consistent with developing one's own argument.

The research (Crone-Todd et al., 2000; Pear et al., 2001) shows that good inter-rater and inter-group reliability can be obtained with the first four Knowledge, Comprehension, Application, and Analysis.

However, thus far there have been too few exemplars of Synthesis and Evaluation to adequately determine Reliability for those categories. In addition, the research is divided in terms of whether the taxonomy is truly hierarchical or not. At best, one might suppose that the lower-level and perhaps the higher-level Application and Analysis categories may be ordinal measures. However, the lack of reliable measures for the Synthesis and Evaluation categories makes it difficult to assess the measurement scale involved.

Although beyond the scope of the present paper, it must be acknowledged that all of this complex human behavior involves

the use of verbal behavior (Skinner, 1957). As the level of complexity increases, the student moves through echoic and tact behaviors, and then to more and more complex intraverbal behavior, which involves learning concepts (i.e., stimulus classes) and equivalence relations, as well as relational frames (see Crone-Todd & Pear, 2002, Hayes, 1994, and Sidman, 1994 for a more discussion of these concepts).

One issue that is especially relevant to the current research on he modified taxonomy is whether in fact there are more difficult levels of each of the categories. For example, it might be relatively simple to demonstrate memorization or comprehension of a simple definition or fact. However, demonstrating memorization or comprehension of more complex definitions or facts would be more difficult.

Assessment of behavioral hierarchical complexity

Another line of research that may shed light on the complexity level is by Commons and his colleagues (e.g., Commons & Miller, 1998; Commons, Miller, Goodheart, & Danaher-Gilpin, 2005; Commons, Trudeau, et al., 1998) in the Model of Hierarchical Complexity (MHC). This yields a stage theory that has some initial similarity to Piaget's work, but goes beyond Formal Operations and divides task properties from performance properties. A property of tasks is that they can be described by orders of hierarchical complexity. Stages of performance are the corresponding numbers when a task of a given order has been successfully completed. Systematic stage is the first stage beyond Formal stage. Actions at that stage essentially organize two or more relationships among variables. Evidence of this stage is inferred from demonstration of an understanding of multivariate causation, for example. Metasystematic is the next stage, in which one can compare, contrast, and synthesize systems of formal relations. At the Paradigmatic stage, one synthesizes systems of systems (metasystems) into new fields based on new paradigms. A further synthesis across fields may result in an entirely new field emerging, or old ones being substantially revised. When this occurs, the person can be said to be operating at a cross-paradigmatic stage. Since the model appears to increase in complexity, it would be considered an ordinal measure.

The Relationship between Categories and Hierarchical Complexity

As one contemplates the stages in the complexity model, one can formulate some theories about how Bloom's and the hierarchical complexity model are related to one another. For example, the Systematic order task would require that one can demonstrate knowledge and comprehension of two relationships among variables and how they interact to form a system. To go onto the Metasystematic order, it seems reasonable to postulate that one must apply and analyze the properties of two or more systems, and how the nature of the variables within the system produce differences and similarities in the systems. It may also be the case that there is a synthesis needed to truly demonstrate understanding of an interaction effect. Arguing for why there is a main effect or interaction would surely indicate that this is the case at the systematic stage. The very wording (see above) of the Paradigmatic Order would imply synthesis, and if one wishes to establish and keep the new field alive, they must evaluate the new field, and argue for its existence. The Paradigmatic order necessarily involves all of the categories in Bloom's taxonomy, because one would have to analyze the metasystems involved, and then synthesize elements of those metasystems into a new one. According to the research, it is very rare to find people at the Paradigmatic or Cross-Paradigmatic stages.

One question that might arise is as follows: Why use both Bloom's taxonomy and the MHC if the phrases and definitions in the hierarchical complexity model are so similar? After all, perhaps the hierarchical complexity model itself would suffice to conduct research and to better define stages of complex behavior. However, research also suggests that at the various orders of complexity, individuals can only demonstrate understanding of a stage that is one or two above where they are currently scoring. If this is the case, then the use of the taxonomy may be a useful tool for describing the levels of support for engaging in increasingly more complex behavior.

There are seven levels of support identified (Commons & Miller, 1998; Commons, Danaher-Gilpin, Miller, & Goodheart, 2002) to help individuals move from one stage to the next one. They are identified in Table 2 below, with the associated Bloom's taxonomy categories.

Table 2. MCH level of support, action, and associated Bloom's taxonomy

MHC Level of Support	Action	Bloom's Taxonomy Category
Manipulation	Student performs with guided help	N/A
Transfer of Stimulus Control	Student performs with cue	Knowledge
Pervasive Imitation	Student can imitate	Knowledge
Direct	No support	Comprehension
Problem Finding	Discover task, give example	Application
Question Finding	Find and solve problems with stimuli	Synthesis/ Evaluation
Phoneme Finding	Discover new phenomenon without stimulus	Synthesis/ Evaluation

The purpose of this research, then, is to show how the MHC might be applied to questions that have already been rated reliably using the modified Bloom's taxonomy (i.e., Crone-Todd, et al., 2000). What follows is an application of this approach to demonstrate that despite lower-level assessments using a categorical system, that greater complexity may be required of students.

Scoring Questions Using MHC

Method

Two raters scored each of the following questions, which are taken verbatim from Martin & Pear (2003). All of the questions have previously been reliably assessed using the modified version of Bloom's taxonomy (Crone-Todd et al, 2000), and have been used extensively in a computer-aided personalized system of instruction (CAPSI, see Pear & Crone-Todd, 1999) in Canada and the United State of America.

The first rater assessed the questions, and then compared scoring with the second rater. The first rater was a novice, and the second rater was skilled in scoring using the MHC. Overall rates

of agreement were high (87.50%).

The questions were chosen on the basis of being previously assessed as being in the Knowledge or Comprehension categories. In all cases, the answers to the questions could be found by reading the textbook (Martin & Pear, 2003).

Definitions

The items assessed refer to the components of questions asked relating to the material presented in the text. The modified Bloom's taxonomy category is the level at which the item was previously scored. The MHC order is the stage at which the question, *prima facie*, would be scored (see above). The level of support provided by the text refers to whether or not the answer is contained in the text (one level of support), or not (none). Finally, the functional stage of behavioral complexity is arrived at by subtracting the level of support from the MHC order. This effectively results in a scored stage of behavioral complexity for the task.

Results & Discussion

The scored items (see Table 3) reveal that there are varying levels of complexity, despite only two previously assessed categories. What is suggested here is that what Bloom and his colleagues identified was perhaps summary labels that are related both to the level of complexity and to the levels of support that were responsible for moving students through the process of a particular stage related to the educational level at which they were, and could be, functioning. This also suggests that other behavioral researchers (e.g., Johnson & Chase, 1981; Semb & Spencer, 1976) also broke down complex tasks into categories in which they could specify the behaviors that must be performed. However, none of the previous approaches speak to the complexity of the task.

Table 3. Comparison of a modified Bloom's taxonomy and the MHC scoring system

Item from Martin & Pear (2003)	Modified Bloom's category	MHC Order	Level of Support Provided by Text	Functional Stage of Behavioral Complexity
What is PSI?	Comprehension	Abstract	Pervasive imitation	Concrete
Who was its founder?	Rote Knowledge	Concrete	Pervasive imitation	Concrete
State at least five characteristics of PSI	Comprehension	Systematic	Pervasive imitation	Formal
Briefly describe how PSI has made use of computer technology	Comprehension	Systematic	Pervasive imitation	Formal
Which is more effective for teaching undergraduates, the traditional lecturing approach or PSI?	Rote knowledge	Concrete	Pervasive imitation	Concrete
Justify your answer.	Evaluation	Metasystematic	Pervasive imitation	Formal
What is a positive reinforcer?	Rote Knowledge	Formal	Pervasive imitation	Abstract
In what way is positive reinforcement like gravity?	Comprehension	Metasystematic	Pervasive imitation	Systematic

One interesting aspect of using a scoring system that considers the hierarchical complexity of the task is that it may lead educators to consider the current skill complexity shown by students, and what the final goal would be for the educational level. For example, sometimes students struggle with course material, which may produce punishment or extinction, leading to less interest in the material (see Crone-Todd, Eyre, Hutchens, Jones, & Pear, in press). This problem might be alleviated by allowing for more support in developing the more complex skill, which could lead to more reinforcement and higher-order thinking.

The process involved in providing levels of support is consistent with both the behavioral and constructivist literature (see Pear & Crone-Todd, 2002 for a discussion of similarities between constructivism and behaviorism). The behavioral approach would call the initial support and cues, followed by removal of these supports, fading. As students engage in more and more independent actions (direct, problem finding, question finding, and phoneme finding), their behavior is being shaped and chained through the use of both extinction for lower-level performance, and reinforcement for closer approximations up through the levels of support. The constructivist approach would call this process the scaffolding that takes place to move the student through their zone of proximal development (Arlin, 1975, 1984; Fischer et al., 1984; Vygotsky, 1978).

An approach that considers the behavioral aspects involved in the order of hierarchical complexity should be useful to both stage theorists and to constructivists. An integrated approach should allow both educators and researchers to formulate individualized plans that are tailored to the student's current level of complex skill development, and would also allow for a more precise method to support movement to the next stage. Future research should center on further determining the reliability and validity of the MHC system in this context, which would include extending it to both instructor-led and computer-based environments. Such an approach should be encouraged to make educational practices more systematic, if not metasystematic.

References

Arlin, P. K. (1975). Cognitive development in adulthood: A fifth stage? *Developmental Psychology*, 11, 602-606.

Arlin, P. K. (1984). Adolescent and adult thought: A structural interpretation. In M. L. Commons, F. A. Richards, & C. Armon (Eds.), *Beyond formal operations: Vol. 1. Late adolescent and adult cognitive development.* (pp. 258-271). New York: Praeger.

Bloom, B. S. (Ed.) (1956). *Taxonomy of educational objectives, the classification of educational goals – Handbook I: Cognitive domain*. New York: McKay.

Carnine, D. (1991). Curricular interventions for teaching higher order thinking to all students: Introduction to the special series. *Journal of Learning Disabilities, 24*, 261-269.

Crone-Todd, D. E., Eyre, H. L., Hutchens, S. A., Jones. J. R., & Pear, J. J. (in press). The effect of changing grading criteria on student writing. *Behavior Analyst Today.*

Crone-Todd, D.E., & Pear, J.J. (2001). *Application of Bloom's Taxonomy to PSI.* Behavior Analyst Today, 3, 204-210.

Crone-Todd, D. E., Pear, J. J., & Read, C. N. (2000). Operational definitions for higher-order thinking objectives at the post-secondary level. *Academic Exchange Quarterly, 4*(3), 99-106.

Commons, M. L., & Miller, P. M. (1998). A quantitative behavior-analytic theory of development. *Mexican Journal of Experimental Analysis of Behavior. 24*(2), 153-180.

Commons, M. L., Miller, P. M., Goodheart, E. A., & Danaher-Gilpin, D. (2005). *Hierarchical complexity scoring system (HCSS): How to score anything.* Retrieved February 21, 2007, from http://dareassociation.org/Papers/Scoring%20Manual.htm

Commons, M. L., Trudeau, E. J., Stein, S. A., Richards, F. A., & Krause, S. R. (1998). The existence of developmental stages as shown by the hierarchical complexity of tasks. *Developmental Review, 8*(3), 237-278.

Fischer, K. W., Hand, H. H., & Russell, S. (1984). The development of abstractions in adolescence and adulthood. In M. L. Commons, F. A. Richards, & C. Armon (Eds.), *Beyond formal operations: Vol. 1. Late adolescent and adult cognitive development* (pp. 43-73). New York: Praeger

Halpern, D. F. (1998). Teaching critical thinking for transfer across domains. *American Psychologist, 53,* 449-455.

Hayes, S.C. (1994). Relational frame theory: A functional approach to verbal events. In S. C. Hayes, L. J. Hayes, M. Sato, & K. Ono (Eds.), *Behavior analysis of language and cognition: The fourth international institute on verbal relations* (pp. 11 – 30). Reno, NV: Context Press.

Hohn, R. L., Gallagher, T., & Byrne, M. (1990). Instructor-supplied notes and higher-order thinking. *Journal of Instructional Psychology, 17*, 71-74.

Johnson, K. R., & Chase, P. N. (1981). Behavior analysis in instructional design: a functional typology of verbal tasks. *The Behavior Analyst, 4*(2), 103-121.

Martin, G. L. & Pear J. J. (2003). *Behavior modification: What it is and how to do it.* (7th Ed.). Upper Saddle River, NJ: Prentice Hall.

Newman, F. M. (1991a). Promoting higher order thinking in social studies: Overview of a study of 16 high school departments. *Theory and Research in Social Education, 19*, 324-340.

Newman, F. M. (1991b). Classroom thoughtfulness and students' higher order thinking: Common indicators and diverse social studies courses. *Theory and Research in Social Education, 19*, 410-433.

Paul, R. W., & Heaslip, P. (1995). Critical thinking and intuitive nursing practice. *Journal of Advanced Nursing, 22*, 40-47.

Pear, J. J., & Crone-Todd, D. E. (1999). Personalized system of instruction in cyberspace. *Journal of Applied Behavior Analysis, 32*, 205-209.

Pear, J. J., & Crone-Todd, D. E. (2002). A social constructivist approach to computer-mediated instruction. *Computers & Education, 38*, 221-231.

Pear, J. J., Crone-Todd, D. E., Wirth, K., & Simister, H. (2001). Assessment of thinking levels in students' answers. *Academic Exchange Quarterly, 5*(4), 94-98.

Semb, G., & Spencer, R. (1976). Beyond the level of recall: An analysis of complex educational tasks in college and university instruction. In L. E. Fraley & E. A. Vargas (Eds.), *Behavior Research and Technology in College and University Instruction* (pp. 115-126). Gainesville, FL: Department of Psychology, University of Florida.

Sidman, M. (1994). *Equivalence relations and behavior: A research story*. Boston: Authors Cooperative.

Skinner, B. F. (1957). *Verbal behavior*. Acton, MA: Copley Publishing Group.

Vygotsky, L. S. (1978). *Mind and society: The development of higher mental processes*. Cambridge, MA: Harvard University Press.

Williams, R. L. (1999). Operational definitions and assessment of higher-order cognitive constructs. *Educational Psychology Review, 11,* 411-427.