Journal of Adult Development, Vol. 11, No. 2, April 2004 (© 2004)

Assessing Intellectual Development: Three Approaches, One Sequence

Theo Dawson[1]

In this paper, I compare three developmental assessment systems, employed to score a set of 152 interviews of engineering students: the Perry Scoring System (W. G. Perry, 1970), the Hierarchical Complexity Scoring System (T. L. Dawson, 2004, 1/31/03), and the Lexical Abstraction Assessment System (LAAS; T. L. Dawson & M. Wilson, in press). Overall, the Hierarchical Complexity Scoring System and Perry Scoring System agree with one another within the parameters of interrater agreement commonly reported for either one of the systems, and the Perry system and the LAAS agree with one another about as well as the LAAS and the Hierarchical Complexity Scoring System, upon which the LAAS is based.

KEY WORDS: adult development; conceptual development; hierarchical assessment; measurement.

INTRODUCTION

Until recently, there were only two types of cognitive-developmental assessment systems for texts. The first to be developed were *domain-based*[2] systems like those introduced by Perry (1970), for epistemological development, Kohlberg (1969), for moral development, Fowler (1981), for faith development, Armon (1984a), for the development of evaluative reasoning, and Kitchener and King (1981), for the development of reflective thinking. The second type of developmental assessment system to emerge was the *domain-general* scoring system. The Hierarchical Complexity Scoring System (Dawson, 2002b), which is based primarily on Commons' (Commons et al., 1995) and Rose and Fischer's assessment systems (Rose & Fischer, 1989), is the most developed and most thoroughly validated of these (Dawson, 1998, 2001, 2002a; Dawson, Commons, & Wilson, in review; Dawson & Gabrielian, in review; Dawson, Xie, & Wilson, in review). Only recently has a third developmental assessment alternative emerged. This is an objective computerized method for assessing the developmental level of texts called the Lexical Abstraction Assessment System (LAAS). Based on the construct, *hierarchical order of abstraction* (Dawson & Wilson, in review), the LAAS is the first accurate and reliable objective measure of the hierarchical complexity of texts. The domain-generality of the LAAS is presently under investigation.

Proponents of domain-specific assessments like Perry's frequently argue that each knowledge domain has a unique structure that can only be understood and measured in terms of the concepts and structures of that domain (Demetriou & Efklides, 1994; Kohlberg & Candee, 1984; Turiel, 1989). In this paper, I question this position by examining whether or not Perry's domain-based system for assessing epistemological development produces importantly different results from those of the Hierarchical Complexity Scoring System and the LASS. To do this, I compare the results obtained with each system on the same set of 152 interviews of students enrolled in an undergraduate engineering program. In addition, I show how generalized developmental assessment systems like the Hierarchical Complexity Scoring System and

[1]To whom correspondence should be addressed at Cognitive Science, Hampshire College, Amherst, Massachusetts 01002; e-mail: tdawson@hampshire.edu.

[2]These scoring systems are *domain-based*, in the sense that the range of conceptual content one can score with a given scheme is limited. The Perry scheme, for example, is employed exclusively to score epistemological content. Likewise, Kohlberg's scheme is used to score moral content.

1068-0667/04/0400-0071/0 © 2004 Plenum Publishing Corporation

the LAAS can aid in efforts to delineate sequences of conceptual development, by employing hierarchical complexity assessment along with conceptual content analysis to describe epistemological development.

Perry

Perry was one of the first researchers to suggest that observed differences in school performance might be due to developmental differences rather than differences in ability, intelligence, or personality. Using student self-reports of their experiences in college, Perry (1970) demonstrated that university students' conceptions of knowledge develop through nine developmental positions, from the absolutist position that knowledge is either "right" or "wrong" to the view that all knowledge is relative. Each position involves qualitative changes in the complexity of thinking about and conceptualizations of epistemology. Descriptions of the nine positions follow:

Position 1: The student views knowledge as either correct or incorrect. Knowledge is construed as an accumulation of facts collected through hard work and obedience.
Position 2: The student recognizes that there are conflicting opinions, but views some as correct and others as incorrect.
Position 3: Diversity and uncertainty are accepted, but only because the "answer" has not yet been found.
Position 4: The student comes to the conclusion that everyone is entitled to her own opinion, though right and wrong still prevail in the realm of authority.
Position 5: The student views all knowledge as contextual.
Position 6: The student comes to understand that it is necessary for him to commit to a position within a relativistic world.
Position 7: This commitment is made.
Position 8: The implications of commitment are explored as are notions of responsibility.
Position 9: The individual situates herself within an identity that incorporates multiple responsibilities, and views commitment as an ongoing process through which the self finds expression.

Several researchers have conducted investigations of epistemological reasoning using Perry's scheme (Benack, 1983; Cleave-Hogg, 1996; Clinchy, Lief, & Young, 1977; Kirk, 1986; Knefelkamp & Slepitza, 1976; Kurfiss, 1977; Widick, 1977). Higher Perry positions are related to educational level and have been shown to predict school performance (Kirk, 1986) and the development of empathy (Benack, 1983). Age differences in Perry position have also been reported. For example, Cleave-Hogg (1996) found in her study of 64 adult undergraduate students between the ages of 30 and 65, that 61 expressed epistemologies that were representative of the two highest levels of Perry's hierarchy. Traditionally aged undergraduate students, though their epistemological reasoning undergoes development during the college years, tend to perform at lower levels than older undergraduate students (Knefelkamp & Slepitza, 1976; Kurfiss, 1977; Perry, 1970).

Domain-Specific Versus Domain-General Scoring Systems

Most cognitive-developmental researchers agree that development across knowledge domains does not necessarily proceed at the same rate (Fischer & Bidell, 1998; Lourenco & Machado, 1996). However, there is still considerable disagreement about whether development across domains can be characterized in terms of a single, generalized process. Domain theorists argue that entirely different processes apply in different knowledge domains. Others, though they acknowledge that unique structures and processes are associated with particular domains, also argue that a single general developmental process applies across domains. Piaget called this process *reflective (or reflecting) abstraction*, through which the actions of one developmental level become the subject of the actions of the subsequent level. The product of reflective abstraction is *hierarchical integration*. In conceptual development, hierarchical integration is reflected in the concepts constructed at a new level by coordinating (or integrating) the conceptual elements of the prior level. These new concepts are said to be qualitatively different from the concepts of the previous level, in that they integrate earlier knowledge into a new form of knowledge. For example, notions of *play* and *learning* constructed at one level are integrated into a notion of *learning as play* at the next level. This new concept cannot be reduced to the original *play* and *learning* elements, because even their earlier meanings are transformed when they are integrated into the learning as play concept. This hierarchical integration is observable in performance in the form of *hierarchical complexity*. For example, the concept of learning as play is one order of hierarchical complexity more complex than the earlier concepts of play and learning from which it is constructed.

Hierarchical Complexity

The general developmental model employed here has been strongly influenced by Piaget's stage model, Fischer's (1980) skill theory, and Commons' General Stage Model (Commons, Richards, Ruf, Armstrong-Roche, & Bretzius, 1984; Commons, Trudeau, Stein, Richards, & Krause, 1998), which is also known as the Model of Hierarchical Complexity. The General Stage Model is a model of the hierarchical complexity of tasks. In this model, an action is considered to be at a given stage when it successfully completes a task of the order of hierarchical complexity associated with that stage. Hierarchical complexity refers to the number of nonrepeating recursions that coordinating actions must perform on a set of primary elements. Actions at the higher order of hierarchical complexity are: (a) defined in terms of the actions at the next lower order; (b) organize and transform the lower order actions; and (c) produce organizations of lower order actions that are new and not arbitrary and cannot be accomplished by the lower order actions alone. The General Stage Model does not define stages in terms of specific logical abilities like Piaget's INRC groups or class inclusion. Rather, it proposes that a given class inclusion task, for example, has a particular order of hierarchical complexity, depending on the nature of its elements and the kinds of classes formed.

The General Stage Model specifies 15 stages. The sequence is—(0) computory, (1) sensory and motor, (2) circular sensory–motor, (3) sensory–motor, (4) nominal, (5) sentential, (6) preoperational, (7) primary, (8) concrete, (9) abstract, (10) formal, (11) systematic, (12) metasystematic, (13) paradigmatic, and (14) cross-paradigmatic. The first three stages correspond to Piaget's sensorimotor stage, 3–5 correspond to his symbolic or preoperational stage, 6–8 correspond to his concrete operational stage, and 9–11 correspond to his formal operational stage. Stages 0 to 12 also correspond definitionally to the levels and tiers originally described by Fischer (1980). These are (0) single reflexive actions, (1) reflexive mappings, (2) reflexive systems, (3) single sensorimotor schemes, (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. The level and tier names from Fischer's skill theory are also used to denote orders of hierarchical complexity. Stages 13 and 14 are hypothetical postformal stages (Sonnert & Commons, 1994).

Not only are there definitional correspondences among analogous stages described by Commons, Fischer, and Noelting (and others not mentioned here), there is empirical evidence of correspondences between General Stage Model stages and at least three content-based systems, including Kitchener and King's (Dawson, 2002c; Kitchener, Lynch, Fischer, & Wood, 1993) stages of reflective judgment, Armon's Good Life stages (Dawson, 2002a), and Kohlberg's moral stages (Dawson & Gabrielian, in review; Dawson & Kay, in review).

As noted above, the Hierarchical Complexity Scoring System is based primarily on Commons' (Commons et al., 1995) and Rose & Fischer's (1989) assessment systems. This scoring system, like its predecessors, is designed to make it possible to assess the hierarchical complexity of a performance (its *complexity order*) without reference to *particular* conceptual content. Rather than making the claim that a person occupies a complexity order because he or she, for example, has elaborated a particular conception of justice, the Hierarchical Complexity Scoring System permits us to identify performances at a particular complexity order and then ask what the range of justice conceptions are at that complexity order. Thus, it avoids much of the circularity of many stage scoring systems, which define stages in terms of particular conceptual content (Brainerd, 1993).

It is possible to score the hierarchical complexity of text performances, because hierarchical complexity is reflected in two aspects of performance that can be abstracted from particular conceptual content. These are (a) hierarchical order of abstraction and (b) the logical organization of arguments. Hierarchical order of abstraction is observable in texts because new concepts are formed at each complexity order as the operations of the previous complexity order are "summarized" into single constructs. Halford (1999) suggests that this summarizing or "chunking" makes advanced forms of thought possible by reducing the number of elements that must be simultaneously coordinated, freeing up processing space and making it possible to produce an argument or conceptualization at a higher complexity order. Interestingly, at the preoperational, abstract, and metasystematic complexity orders, the new concepts not only coordinate or modify constructions from the previous complexity order, they are qualitatively distinct conceptual forms—representations, abstractions, and principles, respectively (Fischer, 1980). The appearance of each of these conceptual forms ushers in three repeating logical forms—definitional, linear, and multivariate. Because these three logical forms are repeated several times throughout the course of development, it is only by pairing a logical form with an hierarchical order of abstraction that a rater can make an accurate assessment of the complexity order of a performance. Other researchers have observed and described similar conceptual forms and repeating logical structures (Case, Okamoto, Henderson, & McKeough, 1993; Fischer & Bidell, 1998; Overton, Ward, Noveck, & Black, 1987; Piaget & Garcia, 1989).

Hierarchical complexity scoring can be conducted in any domain because hierarchical order of abstraction and logical structure guide scoring, rather than the identification of *particular* conceptual content as in conventional domain-based systems. The domain generality of hierarchical complexity scoring raises the question of whether the Hierarchical Complexity Scoring System assesses the same dimension of performance as conventional developmental assessment systems.

Four validation studies have shown that the Hierarchical Complexity Scoring System and one of its predecessors, the General Stage Scoring System (Commons et al., 1995) predominantly assess the same dimension of performance as more content-dependent stage-scoring systems. In the first of these, a think-aloud procedure was employed to compare the scoring behavior of five raters trained in the General Stage Scoring System with the scoring behavior of three raters trained in Kohlberg's Standard Issue Scoring System (Dawson, 2001). All raters scored the same 43 texts. A mean score for each text was calculated for each group of raters, resulting in two scores for each text, one based on the ratings of General Stage Scoring System raters and one based on the ratings of Standard Issue Scoring System raters. Despite the fact that the raters trained in the Standard Issue Scoring System justified their stage assignments on the basis of particular moral conceptions and interpersonal perspectives, whereas General Stage Scoring System raters justified their complexity order assignments in terms of logical structure, these mean scores were within one complexity order of one another 95% of the time ($r = .94$).

In a second study, the Hierarchical Complexity Scoring System, Armon's (1984b) Good Life Scoring System, and the Standard Issue Scoring System were employed to score three different interviews administered to 209 5- to 86- year-olds. Correlations of .90 and .92 were found between the results obtained with the Hierarchical Complexity Scoring System and the Standard Issue and Good Life Scoring Systems (Dawson, 2002a). Dawson argues that these correlations, combined with patterns in the acquisition of analogous good life stages, moral stages, and complexity orders provide evidence that the three scoring systems predominantly assess the same latent dimension: hierarchical complexity.

In a third study, Dawson et al. (Dawson, Xie, et al., 2003) conducted a multidimensional partial credit Rasch analysis of the relationship between scores obtained with the Standard Issue Scoring System and scores obtained with the Hierarchical Complexity Scoring System on 378 moral judgment interviews from respondents aged 5–86. They found a correlation of .92 between scores awarded with the two scoring systems, suggesting that to a large extent these two systems assess the same dimension of performance, though the Hierarchical Complexity Scoring System appeared to be somewhat "easier" than the Standard Issue Scoring System, particularly at the lower complexity orders. The Hierarchical Complexity Scoring System also revealed more stage-like patterns of performance than the Standard Issue Scoring System, including evidence of developmental spurts and plateaus.

In the fourth study (Dawson, 2002), the relationship between complexity orders and Kitchener and King's (1990) Reflective Judgment Stages was examined (Dawson, 2002c). In a sample of 209 interviews of adolescents and adults, the correlation between complexity order scores and Reflective Judgment scores was .84. Agreement between Reflective Judgment scores and complexity order scores was within one Reflective Judgment stage 90% of the time. This is higher than the reported median interrater agreement rate of 77% within one Reflective Judgment stage (Kitchener & King, 1990).

A fifth study examines the validity of the Hierarchical Complexity Scoring System as a stage scoring system. In this study Rasch scaling is employed to examine patterns of performance in a set of 747 moral judgment interviews of 5- to 86-year-olds scored with the system (Dawson, Commons, et al., in review). The authors examine these data for evidence supporting the specified developmental sequence as well as evidence for qualitative (as opposed to cumulative) change. They (a) identify six developmental stages in this age-range; (b) show that performances are all either consolidated at a single complexity order or a mixture of two adjacent complexity orders—a pattern that supports the specified order of acquisition; and (c) demonstrate that movement from stage to stage proceeds in a remarkably consistent series of spurts and plateaus across the six complexity orders, reflecting the tendency for individuals to spend less time in transition from one complexity order to another than in periods of consolidation. Domain-based scoring systems rarely reveal developmental patterns that are as consistent with the postulates of stage-developmental theory. The authors also show that two of the complexity orders—abstract systems and single principles—rarely occur before adulthood, and that patterns of performance on these two complexity orders are virtually identical to patterns of performance on the complexity orders found primarily in childhood and adolescence.

The LAAS

The LAAS is an objective, computer-based system for assessing the hierarchical complexity of texts. It is based primarily on Commons' (Commons et al., 1998) General Stage Model, Fischer's skill theory (Fischer, 1980), and the newly elaborated notion of hierarchical order of abstraction, described briefly above and in greater detail by Dawson and Wilson (in press). Unique manifestations of hierarchical order of abstraction are evident at every complexity order in the form of new conceptual understandings. Many of these new meanings are embodied in lexical items. For example, the abstract mappings conception of *honor* is constructed, in part, from abstract order conceptions of reputation, honesty, and fairness. The word *honor* rarely appears in texts before the abstract mappings order, and the words *reputation*, *honesty*, and *fairness* rarely appear in texts before the single abstractions order. Dawson and Wilson observed that each complexity order is probabilistically associated with a lexicon, and that lexical items from each lexicon are systematically distributed in text performances in ways that permit the computation of computerized scoring rules.

Another way in which hierarchical order of abstraction is evident in texts is in the mean word length of vocabulary. Because many morphemes stand for more abstract meanings than root words, an increase in mean word length should reflect increasing hierarchical order of abstraction. For example, the suffix *-ness* raises the hierarchical order of abstraction of a number of concepts such as *fair*, *kind*, *playful*, and *neat*. For this reason, mean word length is also taken into account in determining the hierarchical complexity of text performances with the LAAS. Taking into account both mean word length and the distribution of lexical items in vocabulary, the LASS awards scores that agree with human hierarchical complexity ratings within 1/3 of a complexity order about 83% of the time. A LAAS score is referred to as a Lexical Abstraction Position (LAP). Though the LAAS has not yet been specifically calibrated to score epistemological interviews, the scoring criteria are hypothesized to be general enough to function well in the epistemological domain.

Perry Positions, Complexity Orders, and LAPs

The following analyses examine the extent of convergence between scores assigned with the domain-based Perry scoring system and the domain-general LAAS and Hierarchical Complexity Scoring System. They further provide an analysis of some of the epistemological concepts associated with orders of hierarchical complexity, and compare the conceptual content associated with particular complexity orders with analogous Perry position definitions.

METHOD

The data analyzed here were collected for a study on the development of epistemological reasoning in engineering undergraduates. Data collection procedures are described in Marra, Palmer, and Litzinger (2000) and Wise, Lee, Litzinger, Marra, and Palmer, (2001). A total of 152 semistructured Perry interviews dealing with issues of truth, knowledge, and learning were collected from college freshmen ($n = 14$), sophomores ($n = 73$), and seniors ($n = 65$).

Perry Scoring

The original Perry scoring was done by an expert Perry rater who awarded a holistic score composed of three digits (333, 334, 344, 444, 455, etc.) to each interview. A score of 344, for example, means that a given interview is predominantly at position 4 with some position 3 reasoning, whereas a score of 334 means that a performance is predominantly at position 3 with some position 4 reasoning, and a score of 444 means that a performance is solidly at position 4.

Scoring with the Perry system involves the identification of particular epistemological arguments. For example, to be awarded a score at position 4, an individual must argue that knowledge is uncertain and therefore (1) everyone is entitled to their own opinion, (2) right and wrong prevail only in the realm of authority, (3) all methods of determining the truth are arbitrary, (4) many decisions about the truth are based on personal preference, or (5) we can never know the "truth." When these arguments or arguments of this type are present, they are considered to provide evidence of position 4 reasoning.[3] If all of the arguments in a given interview are of this kind the performance is scored at position 4. If position 5 arguments are also present, the performance is scored as mixed. The final score is a "holistic" judgment, based on an overall evaluation of the performance, rather than actual counts of arguments associated with different levels.

To make Perry scores more amenable to quantitative analysis, these scores were transformed as follows: 333 = 3.34, 344 = 3.67, 444 = 4, etc. This is a common practice in research involving the Perry scheme. The interviews in the present sample were awarded transformed Perry ratings ranging from 2.67 to 7 ($m = 3.72$, $sd = .73$). To facilitate comparisons across scoring systems Perry ratings are represented, in several of the tables and figures that follow, in terms of the dominant position only.

Unfortunately, reliability estimates for the interview instrument and scoring system employed in this study are unavailable. However, the scoring system applied to these data is similar to the system described by Mentowski, Moeser, and Strait (1983), which has undergone extensive interrater agreement evaluation. The authors report absolute interrater agreement rates (for independent assessments) from 43% to 54%, whereas rates of agreement on dominant Perry position are, on average, about 71%. They also report a wide range of Pearson correlations (.13–.70) between raters, depending upon the raters being compared, their level of training, the type of performance being rated, and the developmental range of the sample under consideration.

Hierarchical Complexity Scoring

Hierarchical Complexity Scoring involves identifying both the highest order of abstraction and most complex form of organization in text performances (Dawson, 2004). A protocol (scoring unit) is considered to be at a given complexity order if its elements embody the level of abstraction of that complexity order and the complexity of its logical structure meets the formal requirements of that complexity order. In these data, three complexity orders were identified: abstract mappings, abstract systems, and single principles. One score was awarded to each protocol. Ideally, a protocol should represent a complete argument on a given topic. Fragmentary arguments are usually treated as unscorable. However, because this results in loss of data we chose to score fragmentary protocols if adjacent protocols in a given text provided enough information to aid in their interpretation. This meant that the rater had access to the entire interview when scoring. This is a standard practice in this type of research (Armon, 1984b; Colby & Kohlberg, 1987b; Kitchener & King, 1990). Interrater agreement rates range from 80 to 97% within half a complexity order and from 98 to 100% within a full complexity order. This equals or exceeds interrater agreements commonly reported in this field (Armon, 1984b; Colby & Kohlberg, 1987a).

To conduct hierarchical complexity scoring, each interview was subdivided into a set of protocols by standard probe question. There were 15 possible standard probes. Each individual received from 5 to 10 of these probes. Though complete data are preferable, previous psychometric investigations of hierarchical complexity scoring of similar interviews have demonstrated that performance is remarkably stable across protocols (Dawson, 1998; Dawson, Commons et al., in review). This means that, for a given performance, the mean score calculated from a randomly selected subset of protocols is expected to be very similar to the mean score calculated from any other randomly selected set of protocols. Consequently, missing data, if they are missing at random, are likely to have little effect on mean scores.

Each of the protocols from each of the interviews was awarded a hierarchical complexity score. The scores for each case were averaged, so that each performance received a single mean score. The range of these scores in the data is from 4 = abstract mappings to 6 = single principles. Only two cases received scores above 5 = abstract systems. The mean is 4.5 ($SD = .34$).

LAAS Scoring

To calculate a LAP, we first determine a set of densities. Each density represents the percentage of lexical items in an individuals' vocabulary that are associated with a given complexity order. The density

[3]Mentkowski, Moeser, and Strait (1983) provide a much more detailed account of possible constructions for each Perry position.

of lexical items from each of eight *abstraction indices* (lists of lexical items, one for each complexity order from sensorimotor systems to single principles) is calculated. The mean length of words in the vocabulary of each respondent is also calculated. The result is nine bits of information for each respondent. A discriminant function, developed from a construction set of over 1,400 interviews, is then applied to the densities and mean word length, resulting in a LAP rating from 1 to 21. Ratings for the present sample ranged from 15 to 21. These ratings are interpreted as follows: 15 = *consolidated abstract mappings*, 16 = *first transition to abstract systems*, 17 = *second transition to abstract systems*, 18 = *consolidated abstract systems*, 19 = *first transition to single principles*, 20 = *second transition to single principles*, and 21 = *consolidated single principles.*

The hypothetical relationships among complexity orders, LAPs, and Perry positions are shown in Table I. The correspondences between Perry scores and complexity orders/LAPs were arrived at by "scoring" Perry position definitions with the Hierarchical Complexity Scoring System, as described in the second column of the table. Table II shows the descriptive statistics for LAPs, hierarchical complexity scores, and Perry position scores. In this table, LAPs and hierarchical complexity scores have been converted into Perry units. To ease interpretation, the following analyses employ these converted scores.

RESULTS

The Empirical Relationships Among Complexity Orders, LAPs, and Perry Positions

The correlations betweem LAPs and Perry positions, LAPs and complexity orders, and Perry positions and complexity orders are .57, .54, and .40, respectively. These moderate correlations reflect the narrow range of the sample and tell us little about how the levels are empirically associated across systems. Table III details these relationships for the entire sample and at the three levels that are well represented in the sample. As can be seen from this table, when complexity scores are at abstract mappings or are transitional from abstract mappings to abstract systems, complexity orders, LAPs, and Perry scores exhibit a high rate of agreement, comparable to reported rates of agreement between Hierarchical Complexity Scoring System raters or Perry raters. However, at the abstract systems complexity order, Perry ratings, to a great extent, do not agree as well with complexity orders and, to some extent, with LAPs. When Perry scores equal 5, OHCs and Perry ratings enjoy a high level of agreement, and LAPS are likely to agree with Perry ratings within one Perry level, but LAPs and Perry ratings never agree absolutely. When LAPs are abstract systems (Perry 5), Perry scores and LAPs show a high level of agreement, but when OHCs are abstract systems Perry ratings and OHCs do not show a high level of agreement.

Figure 1 shows that, at the abstract systems order (Perry 5), the Perry scheme tends to award scores that are about 3/4 of a Perry level lower than those awarded with the LAAS and about 1 complexity order lower than those awarded with the Hierarchical Complexity Scoring System. The difference between Perry scores, and LAPs or OHCs is greater at this level than at the abstract mappings (Perry 3) and abstract mappings/abstract systems orders (Perry 4).

Are hierarchical complexity scores inflated at this level or is the Perry scoring system underestimating reasoning ability in abstract systems performances? To address this question, I conducted a content analysis of the epistemological concepts associated with each complexity order and transition and compared these with the scoring criteria at analogous Perry positions. To conduct this analysis, I coded the conceptual content of interviews scored abstract mappings, abstract mappings/abstract systems, and abstract systems and identified the epistemological concepts that appeared for the first time at each of these levels. Six interviews were selected at random from each of the group of 37 interviews scored abstract mappings (Perry 3) with the Hierarchical Complexity Scoring System, the 88 interviews scored abstract mappings/abstract systems (Perry 4), and the 24 interviews scored abstract systems (Perry 5). Though there were too few interviews scored abstract systems/single principles or single principles (Perry 6/7 and 8) to conduct an equivalent analysis at this level, I include the arguments presented in the two performances scored abstract systems/single principles and the single performance scored single principles.

The results of the concept analysis are shown in Table IV. Clearly, there are numerous epistemological concepts associated with each complexity order and transition that are not explicitly referred to in Perry scoring criteria. Moreover, though the conceptualizations identified at each complexity order are similar in structure to the descriptions at analogous Perry positions, they are different in content. This is particularly true at the abstract systems order (Perry position 5). For example, in the interviews scored as abstract

Table I. Hypothesized Relationships Among Perry Levels, Complexity Orders, and LAPs

Perry position	Rationale for correspondence	Complexity order LAP[a]	Conceptual structure of complexity order	Logical structure of complexity order
Position 3 Diversity and uncertainty are accepted, but only because the "answer" has not yet been found.	Diversity and uncertainty are abstract mappings order concepts, in that they are more generalized abstractions of precursor abstract concepts of difference and doubt. The relative lack of conceptual differentiation at this level (the individual hasn't yet observed the full magnitude of uncertainty) allows the individual to maintain that truths are "out there" to be found.	*Abstract mappings* LAP = 15	*2nd order abstractions* These coordinate or modify abstractions. For example, the abstract mappings order concept of *personal truth* indicates that the individual differentiates between at least two categories of *truth* as a concept abstracted from concrete instances. Concepts like *universal truth*, *scientific truth*, and *belief vs. truth* are also not constructed before this complexity order.	*Linear* The most complex logical structure of this complexity order coordinates one aspect of two or more abstractions, as in "Because we are all raised differently, each person has their own personal truths, based on upbringing." Here, upbringing determines the kinds of truths we hold in adulthood. *Upbringing* and *personal truth* are both abstractions.
Position 4 Everyone is entitled to her own opinion, but right and wrong still prevail in the realm of authority (or religion).	The increasing differentiation of abstract mappings order concepts raises doubts about the possibility of universal truths. Contradictory arguments coexist because the linear reasoning style of the abstract mappings order prevents the coordination of multiple kinds of truth (religion, authority, scientific, personal).	*Abstract mappings/abstract systems transition* LAP = 16–17		
Position 5 All (most) knowledge is viewed as contextual, though there are gradations of truth and a few right/wrong exceptions.	The ability to represent and coordinate multiple truth domains helps the abstract systems reasoner to see that knowledge is contextual, even though some knowledge seems more certain than other knowledge.	*Abstract systems* LAP = 18	*3rd order abstractions* These coordinate elements of abstract systems. For example, the concept of *point of reference* can be employed at this order do differentiate between different methods of determining truth. Concepts like *gradations of truth*, *the pursuit of truth as an ongoing process*, and *selecting the appropriate method for determining truth* are also not constructed before this complexity order.	*Multivariate* The most complex logical structure of this complexity order coordinates multiple aspects of two or more abstractions. "Because some methods of determining truth, like the scientific method, produce more consistent results than others, in some cases where there is no absolute truth there are better and worse answers." Here the abstract systems order notion that there are better and worse methods for determining truth leads to the conclusion that even though knowledge is uncertain, some answers are better than others.

Positions 6–7 The student comes to understand that it is necessary for him to commit to a position within a relativistic world (6). This commitment is made (7).	The increasing differentiation of abstract systems concepts, along with the multivariate form of abstract systems thinking, make it possible for the individual to coordinate the necessity of making choices with methods for evaluating truth. However, the contradictions and correspondences between evaluative systems, while fleetingly recognized, are not yet extracted to define overarching principles for evaluating knowledge.	*Abstract systems/single principles transition* LAP = 19–20		
Positions 8+ The implications of commitment are explored as are notions of responsibility.	At the single principles order, it becomes possible to extract the principles that underlie systems by coordinating single aspects of these systems. This permits a kind of metacognitive analysis and synthesis of the processes employed at the abstract systems order.	*Single principles* LAP = 21 (single principles)	*First order principles* At the single principles complexity order, the new concepts are referred to as first order principles. These coordinate abstract systems. Concepts like *web of existing knowledge*, *interrelating truths to extract a single truth*, and *coordinating principle* are not constructed before the single principles complexity order.	*Definitional* The most complex logical structure of this complexity order identifies one aspect of a principle or axiom coordinating abstract systems, as in "Knowledge, viewed from a variety of perspectives can inform the structuring of truth, which is in eternal state of transformation." Here, the respondent defines a principle for structuring truth that involves the coordination of different systems of knowledge.

[a] Lexical abstraction position.

Table II. Descriptive Statistics for LAAS, Hierarchical Complexity (OHC), and Perry Position Scores in Perry Units ($N = 152$)

	Minimum	Maximum	Mean	Std. Deviation
LAAS in Perry units	3.00	8.00	4.06	0.76
OHC in Perry units	3.00	8.00	3.96	0.81
Perry	2.67	7.00	3.72	0.73

systems, a major emphasis is on *interpretation*. At this level, individuals can consider knowledge as a system. They also understand that different persons have different knowledge. They are therefore able to argue that two people with different knowledge systems can come up with quite different "truths." This argument is structurally similar to the Perry position 5 argument that knowledge is contextual, but it is not explicitly included in the Perry scoring criteria for this level.

Perhaps the best way to illustrate how the Perry scoring system, as a content-dependent scoring system, might lead to underestimation of an individual's epistemological reasoning level, is to examine an instance in which the Perry system awards a considerably lower score than the Hierarchical Complexity Scoring System. In one of the more extreme cases found in the present set of interviews, a Perry score of 334 was awarded to an interview scored as abstract systems (Perry position 5) with the Hierarchical Complexity Scoring System. In this interview the respondent explains that, "Knowledge is not necessarily truth, [it is] intelligence and wisdom combined. . . . It's a very powerful thing [for someone] to be able to . . . incorporate [their] experiences with [their] intelligence." Though the phrasing is unsophisticated, this respondent appears to mean that knowledge is a construction of the individual, and that good knowledge combines both wisdom and intelligence. The hierarchical complexity of this argument is systematic, because multiple second order abstractions (intelligence and wisdom) are interrelated to define good knowledge. Even when I examine this argument from the perspective of the Perry scheme, I am inclined to argue that it is at Position 5. I would argue that the notion that *knowledge is constructed* is unlikely to appear before an individual has observed that knowledge varies from person to person and situation to situation, which does not occur until Position 4. Moreover, the notion that, *given this variability, some knowledge is still better than other knowledge*, does not occur until Position 5. The idea that good knowledge is constructed on the basis of wisdom (experience) and intelligence is therefore unlikely to occur until Position 5. How then, was this interview awarded a score of 334? One possibility is that the Perry scheme does not explicitly include the idea that knowledge is constructed in any of the position definitions. Because this respondent does not explicitly state that knowledge is contextual (which is included in the Perry position 5 definition), the interview is scored much lower with the Perry scheme than with the Hierarchical Complexity Scoring System.

Table III. Percent Agreement Within 1 Perry Position (1/2 OHC) in Four Conditions

	OHC & Perry		LAP & Perry		OHC & LAP	
Range	When OHC in Perry units =	When Perry =	When LAP in Perry units =	When Perry =	When OHC in Perry units =	When LAP in Perry units =
Abstract mappings (Perry 3)	$n = 45$	$n = 63$	$n = 32$	$n = 63$	$n = 45$	$n = 32$
Absolute agreement	57.8%	41.3%	62.5%	31.7%	37.8%	53.1%
Agreement within 1 Perry position	95.6%	82.5%	96.9%	100%	91.1%	93.8%
Abstract mappings/abstract systems (Perry 4)	$n = 73$	$n = 70$	$n = 83$	$n = 70$	$n = 73$	$n = 83$
Absolute agreement	52.1%	54.3%	50.6%	80.0%	58.9%	51.8%
Agreement within 1 Perry position	100%	98.6%	100%	100%	100%	100%
Abstract systems (Perry 5)	$n = 31$	$n = 15$	$n = 35$	$n = 15$	$n = 31$	$n = 35$
Absolute agreement	19.4%	40.0%	40.0%	0%	38.7%	34.3%
Agreement within 1 Perry position	64.5%	93.3%	91.4%	93.3%	93.5%	88.6%
All	$n = 152$		$n = 152$		$n = 152$	
Absolute agreement	46.1%		50.0%		48.0%	
Agreement within 1 Perry position	90.8%		97.4%		96.1%	

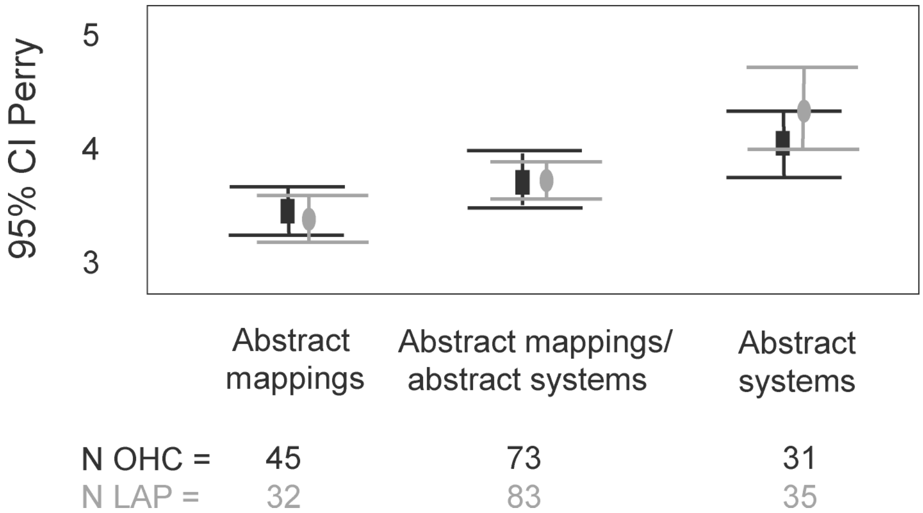

Fig. 1. The relationship between Perry scores and OHC and LAAS scores.

DISCUSSION

This analysis of the relationship between Perry's epistemological positions and orders of hierarchical complexity (as measured by the Hierarchical Complexity Scoring System and the LAAS) shows a strong correspondence between scores awarded with domain-specific and domain-general scoring systems while highlighting important differences. At Perry positions 3 and 4 (the abstract mappings complexity order and abstract mappings/abstract systems) the two types of systems predominantly agree. However, at Perry position 5, the rate of agreement drops considerably. This occurs, at least in part, because some of the arguments found for the first time at the abstract systems complexity order are not included in Perry position 5 scoring criteria even though they have the same structure as Perry position 5 conceptualizations.

I have argued elsewhere that domain-specific scoring systems tend to underestimate the performance level of those respondents who do not present arguments that are very similar to those exemplified in the scoring criteria (Dawson, 2003). Unfortunately, domain-based scoring systems have historically been developed on the basis of relatively small and homogeneous samples. The Perry scheme, for example, was originally based on the epistemological development of Harvard students during the middle of the last century. The particular epistemological concepts constructed by these students were likely to have been influenced by a number of factors, including social class, educational background, and historical context. The construction of a comprehensive domain-based scoring system for epistemological reasoning would require a large longitudinal sample that is representative of the population. This was not the case in the development of the Perry scheme.

Domain-general scoring systems like the Hierarchical Complexity Scoring System and the LAAS, by employing general scoring criteria, eliminate the problem of developing comprehensive domain-based scoring systems while providing valid and reliable estimates of a performance's position on the developmental continuum. Moreover, because hierarchical complexity and particular conceptual content can be analyzed independently, the researcher can legitimately address questions about the relationship between the developmental level of a performance and its conceptual content, that cannot be addressed when domain-based scoring systems are employed. For example, employing the Hierarchical Complexity Scoring System or the LAAS along with a separate conceptual content analysis, would permit one to ask (1) whether the epistemological reasoning of Harvard freshmen and U C Berkeley freshmen differs in hierarchical complexity, and (2) how their epistemological conceptions (within and across complexity orders) differ.

Table IV, though based on only 21 interviews, provides a rich range of epistemological conceptions. A thorough account of conceptual development in this domain would require a larger and more diverse

Table IV. Perry Position Criteria and the Concepts of Learning, Knowledge, Truth, and Truth & Knowledge Associated with Analogous Complexity Orders

Level	Perry	Learning	Knowledge	Truth	Truth & knowledge
Perry 3 *Abstract mappings* LAP = 15	• Diversity and uncertainty are legitimate but temporary. • What is not known can be determined • There are right ways to find the answers • Learning focus shifts away from memorization to process and method	• You learn from what you've experienced with family, friends, other environments; learn from life in general, not just in school • Happens over time • Involves taking in other people's opinions and "mixing" them with your own • Involves understanding • Involves hands-on experience • Pragmatic approach. Learn what you are told to learn, prepared for possibility of revision	• Information that can be taught, learned • Understanding • What you've learned from experience, interaction with others • More is better because provides more tools for seeking truth • Knowing a lot helps you form your own opinions • Should be able to apply it • Can be valid or invalid • Can be "good enough" for now • Logic, common sense, evidence, personal experience, experience of trusted other, quality of source, or "facts" can be used to evaluate information • When evaluating knowledge, listen to both sides and go with what suits you • Knowledge is the end result of learning • Knowledge is absorbed.	• Truth is fact, factual knowledge, proven knowledge • Something to figure out, but hard to find • Harder to find in some knowledge domains than others • Can be proven (if science-based) • Truth is found through problem-solving • Not speculation • There are some absolute scientific truths, scientific laws • What is accepted as truth may not be ultimate truth • Even scientific truths can change, be altered • Each person has their own personal truths, based on upbringing • People can fail to understand truth for psychological reasons (need to make self look good) • Universal truths are fixed, external • Universal truths are taken on faith (esp. if associated with religion)	• Knowledge should be the truth, but isn't always • Knowledge is needed to decide what is true • Even scientific knowledge isn't necessarily true • Truth is easier to verify in math and science than in humanities
Perry 4 *Formal/systematic* LAP = 16–17	• Uncertainty is extensive. There is an epistemological realm in which "everyone is entitled to his own opinion." • Right/wrong still prevails in Authority's realm or you just have to figure out what authority wants to get by, even though it may not be right. • There is no nonarbitrary method for evaluating truth. • We'll never know for sure. • Many decisions ultimately based on personal preference.	• Taking in knowledge from every experience in life • Involves "processing" knowledge • Grasping, fully understanding concepts • Knowledge is the path to truth (which doesn't actually exist).	• Something you've learned to do in an efficient way • What you see as the truth • Same knowledge source can be interpreted differently by different persons • Something you hope is true (act as if it is true) • Consistency can be used to evaluate knowledge • You can come to understand a perspective with which you do not agree • Multiple valid truths possible	• Can be multiple truths in some things	• Intertwined • Varies, depending on domain • The same knowledge can lead to different truths in some instances

Perry 5 *Systematic* LAP = 18	• All knowledge is contextual and relativistic. Right-wrong concerns contextualized. • There are a few right/wrong exceptions. • Some truths are better than others • Judgments and meaning making go on in both personal and professional contexts • Every choice requires taking a point-of-view • Sees self as participant in meaning-making.	• Involves interpretation • Existing knowledge can bias learning • You use your knowledge to learn	• Sought in the pursuit of Truth • In some cases there are no correct viewpoints, only those agreed upon by the majority • Involves interpretation, point of reference • In some cases where there is no absolute truth there are better answers than others • Science operates on theory not truth • No ultimate way to prove truth • Knowing how to find the answer	• Can be more absolute or concrete in some knowledge domains but not others • There are certain overall truths • Pursuit of truth is an ongoing process • Can't be entirely relative • Interpretation of knowledge can differ, resulting in different "truths" • May be cases in natural world that can be understood through science	• Experts in different fields can look at the same phenomenon and come to different conclusions because their knowledge leads them to observe differently
Perry 6,7 *Systematic/ metasystematic* LAP = 19–20	• Sees the necessity of making a personal commitment to a position, though knowledge is uncertain. • Good decisions consider many aspects of a given problem. • Commitment made in the face of legitimate uncertainty, and after experiencing doubt.	• Involves being exposed to different perspectives • Involves learning how to learn	• Subjective • Any learning you can accumulate through your senses • Is only as good as the expertise or the intelligence of the people before us	• Objective • No absolute truth • Would have to know everything to know truth • Human desire is to seek out truth	• Knowledge offers a way of approaching the truth
Perry 8+ *Metasystematic* LAP = 21	• The implications of commitment are explored as are notions of responsibility. • The individual situates herself within an identity that incorporates multiple responsibilities, and views commitment as an ongoing process through which the self finds expression.	• A thought process that is partially conscious and partially subconscious, a network integrating a new idea • Seeking principle that holds together a perspective.	• Knowledge defines an interacting network of experiences and factual information where every node in that network is weighted to one degree or another with respect to every other node • Neither "book smarts" nor pure abstract thinking • Facts and information are the prerequisites to knowledge • Refined through exposure to and reflection upon a variety of perspectives • To evaluate, go directly to the most comprehensive primary sources of evidence you can find • To evaluate, ask how new knowledge fits in with web of existing knowledge.	• Two separate truths can be interrelated to form a third truth • Reevaluated throughout our lives • Some mathematical truths may be absolute	• Knowledge, viewed from a variety of perspectives can inform structuring of truth, which is in eternal state of transformation.

sample, representing a wider age and educational range. By combining existing interviews with additional data from underrepresented age, educational, and cultural groups, we could begin the process of producing such an account. Only when we have constructed a more complete account of conceptual development in the epistemological domain can we provide adequately informative assessments of epistemological development or effective developmentally informed educational interventions.

The high level of agreement between the LAAS and the Perry scoring system suggests the possibility of large-scale assessments of epistemological development. To date, most assessments of epistemological development have been confined to relatively small samples, largely because of the prohibitive expense of scoring. Theoretically, as a computerized assessment system, the LAAS can be employed to assess the developmental progress of large numbers of students at frequent intervals, providing detailed information about developmental processes. The implications for research and education are numerous. First, frequently repeated assessments of developmental progress will permit researchers to trace spurts, dips, and plateaus in developmental progress. Second, such assessments will greatly aid in the description of sequences of conceptual development (using methods like those described here), allowing us to closely examine the various pathways individuals take from one complexity order to the next. Third, large-scale cross-sectional studies will make it possible to trace developmental trends in entire populations, such as student cohorts or company employees. Fourth, research suggests that one prerequisite for movement from one complexity order to the next may be the attainment of a "critical mass" of concepts at the current complexity order (Dawson, 1998; Dawson, Xie, et al., in review). Determining whether this is indeed the case and what constitutes this hypothetical "critical mass" in a given knowledge domain will require large-scale longitudinal developmental assessments. Finally, armed with knowledge of this kind, educators can employ the LAAS to conduct frequent developmental assessments to assist them in targeting the learning needs of individual students.

ACKNOWLEDGMENTS

Special thanks to John Wise for the use of his data, and to both John Wise and Bill Moore for helping me find my way through the Perry literature. The research reported in this paper was made possible in part by a grant from the Spencer Foundation. The data presented, the statements made, and the views expressed are solely the responsibility of the author.

REFERENCES

Armon, C. (1984a). Ideals of the good life and moral judgment: Ethical reasoning across the lifespan. In M. Commons, F. Richards, & C. Armon (Eds.), *Beyond formal operations, Vol. 1: Late adolescent and adult cognitive development* (pp. 357–381). New York: Praeger.

Armon, C. (1984b). *Ideals of the good life: Evaluative reasoning in children and adults*. Unpublished doctoral dissertation, Harvard, Boston.

Benack, S. (1983). Postformal epistemologies and the growth of empathy. In *Beyond formal operations: Late adolescent and adult cognitive development* (pp. 340–356). New York: Praeger.

Brainerd, C. J. (1993). Cognitive development is abrupt (but not stage-like). *Monographs of the Society for Research in Child Development*, *58,* 170–190.

Case, R., Okamoto, Y., Henderson, B., & McKeough, A. (1993). Individual variability and consistency in cognitive development: New evidence for the existence of central conceptual structures. In R. Case & W. Edelstein (Eds.), *The new structuralism in cognitive development: Theory and research on individual pathways* (pp. 71–100). Basel, Switzerland: S. Karger.

Cleave-Hogg. (1996). Learning perspectives of older students enrolled full-time in a reguler undergraduate program. In M. L. Commons, J. Demick, & C. Goldberg (Eds.), *Clinical approaches to adult development* (pp. 239–249). Norwood, NJ: Ablex.

Clinchy, B., Lief, J., & Young, P. (1977). Epistemological and moral development in girls from a traditional and a progressive high school. *Journal of Educational Psychology*, *69,* 337–343.

Colby, A., & Kohlberg, L. (1987a). *The measurement of moral judgment, Vol. 1: Theoretical foundations and research validation*. New York: Cambridge University Press.

Colby, A., & Kohlberg, L. (1987b). *The measurement of moral judgment, Vol. 2: Standard issue scoring manual*. New York: Cambridge University Press.

Commons, M. L., Richards, F. A., Ruf, F. J., Armstrong-Roche, M., & Bretzius, S. (1984). A general model of stage theory. In M. Commons, F. A. Richards, & C. Armon (Eds.), *Beyond formal operations* (pp. 120–140). New York: Praeger.

Commons, M. L., Straughn, J., Meaney, M., Johnstone, J., Weaver, J. H., Lichtenbaum, E., et al. (1995, November). *The general stage scoring system: How to score anything*. Paper presented at the Annual Meeting of the Association for Moral Education, Boston, MA.

Commons, M. L., Trudeau, E. J., Stein, S. A., Richards, S. A., & Krause, S. R. (1998). Hierarchical complexity of tasks shows the existence of developmental stages. *Developmental Review*, *18,* 237–278.

Dawson, T. L. (1998). *"A good education is . . . " A life-span investigation of developmental and conceptual features of evaluative reasoning about education*. Unpublished doctoral dissertation, University of California at Berkeley, Berkeley, CA.

Dawson, T. L. (2001). Layers of structure: A comparison of two approaches to developmental assessment. *Genetic Epistemologist*, *29,* 1–10.

Dawson, T. L. (2002a). A comparison of three developmental stage scoring systems. *Journal of Applied Measurement*, *3,* 146–189.

Dawson, T. L. (2002c, January). *Measuring intellectual development across the lifespan*. Paper presented at the Powerful learning & the Perry scheme: Exploring intellectual development's role in

knowing, learning, and reasoning, California State University, Fullerton, CA.
Dawson, T. L. (2004, 1/31/03). *The Hierarchical Complexity Scoring System*. Retrieved January, 2004, from http://gseacademic.harvard.edu/~hcs/
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., & Gabrielian, S. (2003). Developing conceptions of authority and contract across the life-span: Two perspectives. *Developmental Review*, *23,* 162–218.
Dawson, T. L., & Wilson, M. (in press). The LAAS: A computerized developmental scoring system for small- and large-scale assessments. *Educational Assessment*.
Dawson, T. L., Xie, Y., & Wilson, M. (2003). Domain-general and domain-specific developmental assessments: Do they measure the same thing? *Cognitive Development*, *18,* 61–78.
Demetriou, A., & Efklides, A. (1994). Structure, development, and dynamics of mind: A meta-Piagetian theory. In A. Demetriou & A. Efklides (Eds.), *Intelligence, mind, and reasoning: Structure and development. Advances in psychology* (pp. 75–109). Amsterdam, Netherlands: Elsevier Science.
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., & Bidell, T. R. (1998). Dynamic development of psychological structures in action and thought. In W. Damon & R. M. Lerner (Eds.), *Handbook of child psychology: theoretical models of human development* (5th ed., pp. 467–561). New York: Wiley.
Fowler, J. (1981). *Stages of faith: The psychology of human development and the quest for meaning*. San Francisco: Harper and Row.
Halford, G. S. (1999). The properties of representations used in higher cognitive processes: Developmental implications. In *Development of mental representation: Theories and applications* (pp. 147–168). Mahwah, NJ: Erlbaum.
Kirk, S. L. (1986). *The assessment of learning styles and cognitive development of adult students in a higher education setting: Implications for teaching, administration, and advisement*. Unpublished doctoral dissertation, Oklahoma State University.
Kitchener, K. S., & King, P. M. (1981). Reflective judgment: Concepts of justification and their relationship to age and education. *Journal of Applied Developmental Psychology*, *2,* 89–116.
Kitchener, K. S., & King, P. M. (1990). The reflective judgment model: Ten years of research. In M. L. Commons, C. Armon, L. Kohlberg, F. A. Richards, T. A. Grotzer, & J. D. Sinnott (Eds.), *Adult development* (Vol. 2, pp. 62–78). New York: Praeger.
Kitchener, K. S., Lynch, C. L., Fischer, K. W., & Wood, P. K. (1993). Developmental range of reflective judgment: The effect of contextual support and practice on developmental stage. *Developmental Psychology*, *29,* 893–906.
Knefelkamp, L. L., & Slepitza, R. (1976). A cognitive-developmental model of career development: An adaptation of the Perry scheme. *Counseling Psychologist*, *6,* 53–58.
Kohlberg, L. (1969). Stage and sequence: The cognitive-developmental approach to socialization. In D. Goslin (Ed.), *Handbook of socialization theory and research* (pp. 347–480). Chicago: Rand McNally.
Kohlberg, L., & Candee, D. (1984). Stage and sequence: The cognitive developmental approach to socialization. In *The psychology of moral development: The nature and validity of moral stages* (pp. 7–169). San Francisco: Jossey-Bass.
Kurfiss, J. (1977). Sequentiality and structure in a cognitive model of college student development. *Developmental Psychology*, *13,* 565–571.
Lourenco, O., & Machado, A. (1996). In defense of Piaget's theory: A reply to 10 common criticisms. *Psychological Review*, *103,* 143–164.
Marra, R. M., Palmer, B., & Litzinger, T. A. (2000). The effects of a first-year engineering design course on student intellectual development as measured by the Perry scheme. *Journal of Engineering Education*, *89*(1), 39–45.
Mentkowski, M., Moeser, M., & Strait, M. J. (1983). *Using the Perry scheme of intellectual and ethical development as a college outcomes measure: A process and criteria for judging student performance, Vol. 1*. Milwaukee, WI.: Alverno College, Office of Research and Evaluation.
Overton, W. F., Ward, S. L., Noveck, I. A., & Black, J. (1987). Form and content in the development of deductive reasoning. *Developmental Psychology*, *23*(1), 22–30.
Perry, W. G. (1970). *Forms of intellectual and ethical development in the college years*. New York: Holt, Rinehart, & Winston.
Piaget, J., & Garcia, R. (1989). *Psychogenesis and the history of science* (H. Feider, Trans.). New York: Columbia University Press.
Rose, S. P., & Fischer, K. W. (1989). *Constructing task sequences: A structured approach to skill theory* (Instructional manual). Cambridge, MA: Harvard.
Sonnert, G., & Commons, M. L. (1994). Society and the highest stages of moral development. *Politics and the Individual, 4*, 31–55.
Turiel, E. (1989). Domain-specific social judgments and domain ambiguities. *Merrill–Palmer Quarterly, 35*, 89–114.
Widick, C. (1977). The Perry Scheme: A foundation for developmental practice. *Counseling Psychologist, 6*, 35–38.
Wise, J. C., Lee, S., Litzinger, T. A., Marra, R. M., & Palmer, B. (2001, June). *Measuring cognitive change in engineering undergraduates: A longitudinal study*. Paper presented at the the American Society for Engineering Education Conference, Albuquerque, NM.