“A good education is...”
A life-span investigation of developmental and conceptual features of evaluative reasoning about education

by

Theo Linda Dawson
B. A. Antioch University LA, 1993
M. A. University of California at Berkeley, 1995

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy
in
Educational Psychology–Human Development

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, BERKELEY

Committee in Charge:

Professor Elliot Turiel (Chair)
Professor Paul Ammon
Professor Mark Wilson
Professor Marty Covington

"A good education is..."
A life-span investigation of developmental and conceptual features of evaluative reasoning about education

by

Theo Linda Dawson

Abstract

"A good education is..."

A life-span investigation of developmental and conceptual features of evaluative reasoning about education

by

Theo Linda Dawson

Doctor of Philosophy in Educational Psychology, Human Development

University of California, Berkeley

Professor Elliot Turiel, Chair

A total of 153 semi-structured, probed, clinical interviews, designed to elicit evaluative reasoning about education, were administered to a convenience sample of predominantly Caucasian, middle class respondents, ranging in age from 5 to 86 years. Demographic information was also collected, including age, educational attainment, and sex. Ninety of the interviews were administered by Armon for her longitudinal, life-span study of evaluative reasoning about the good (Armon, 1984).

While attitude and belief research into education is predominantly evaluative and lacks an ontological perspective, developmental research focuses on the ontogenesis of educational conceptions, neglecting the evaluative aspect. The first goal of this project was to bring together these two components, the evaluative and the developmental, to examine how evaluative reasoning about education changes as the hierarchical complexity (stage) of conceptions advances with development. The

second goal was to clarify the relationship between the content of evaluative pedagogical conceptions and their hierarchical complexity by independently determining their (1) hierarchical complexity and (2) propositional content. The attempt to differentiate and then reintegrate the conceptual and structural features of reasoning revealed that though these aspects of performance are part of a dialectical whole, they are not identical in what they have to teach us about development. To illustrate this point, the development of conceptualizations along two metaphorical strands, *education as play,* and *education as preparation for the good life*, are described and discussed.

The third goal of this project was to examine the nature of stages as assessed with the General Stage Scoring System (GSSS). A Rasch analysis of GSSS scores supported postulates of both invariant sequence and *structure d'ensemble*.

The final goal was to add to the body of evidence regarding the relationship of stage of performance with age, education, and sex. A curvilinear relationship, weakening with advancing age, was found between stage and age, while the relationship between educational attainment and stage was linear, with an average advance of one stage for every six years of formal education. No sex differences were found after educational attainment was taken into account.

_____ _____

Chair Date

Table of Contents

List of Tables

List of Figures

Acknowledgments

I would like to thank Cheryl Armon for her inspiring introduction to developmental psychology and the research enterprise, for ongoing feedback about this work, and for her generous contribution of data. Special thanks also to Michael Commons, who has given generously of his time and expertise over the last several years, and to Larry Walker, whose encouragement gave me confidence. I also want to extend many thanks to my chair, Elliot Turiel, who made sure I knew what I was doing; to Trevor Bond for introducing me to Rasch analysis; to Mark Wilson for helping me understand the value of measurement in qualitative research and for believing in this work; to Paul Holland for teaching me that you can never look at your data enough; and to William P. Fisher for dialectics, metaphor, and metrology. Thanks also to Paul Holland, Paul Ammon, Marty Covington, Karen Draney, and Willis Overton for their critical comments and support.

Chapter 1

What *is* a good education?

Part of the purpose of an education is to advance society. It's kind of easy to say "to help society," but what's the point to the individual or to the world? If you're advancing society it shouldn't really matter to anyone except for the collection of individuals themselves. [It's an advancement] if a greater proportion of individual lives are enhanced by the advancement.

—College student, female, 19

Philosophical discussions of education are invariably value-laden. According to Peters (1964, as quoted in Frankena, 1973),

> ...to call something 'educational' is to intimate that the processes and activities...contribute to or involve something that is worthwhile... being worthwhile is part of what is meant by calling it an 'education' (p. 20).

Educational themes and concepts are fundamentally tied to evaluations of the appropriate ends, means, and content of educational enterprises. This is apparent in the pedagogical literature, which is rich in discourse that explores the merits of the various themes and concepts of educational theory (Cooney, Cross & Trunk, 1993; Doyle, 1973; Frankena, 1973; Labaree, 1997; Pratte, 1971). In a sense, the central question of pedagogy is, "What is a good education?"

Over the last 50–60 years, numerous researchers have examined beliefs and attitudes about education in both adults and children, but to date, research into the

development of educational conceptions has been sparse. Most of the research into developing educational conceptions has its origins in early work by Perry (1970), and/or is influenced by Kohlberg and his colleagues' application of cognitive developmental theory to the study of reasoning outside the logicomathematical domain (Ammon, Hutcheson & Black, 1985; Basseches, 1984; Black, 1989; Kegan, 1994; King & Kitchener, 1994; Loevinger, 1987). A growing, independent strand of research into conceptions of learning has recently developed in a field called phenomenography (Giorgi, 1986; Marton, 1994a; Sälgö, 1979; van Rossom, Deijkers & Hamer, 1985). Though phenomenographers are hesitant to make strong developmental claims about the different ways of thinking identified in their work, their findings are remarkably similar to those of developmental researchers, as shown in Chapter 2.

Although, while the attitude and belief literature frequently focuses on evaluations of educational practices and aims, the developmental and phenomenographic literatures are not directly concerned with evaluative reasoning. The first goal of this project is to bring together these two components, the evaluative and the developmental, to determine how evaluative reasoning about education changes as the hierarchical complexity of conceptions advances with development.

The second goal is to clarify the relationship between the content of evaluative pedagogical conceptions and their hierarchical complexity by engaging in a deliberately iterative, relational research process. This research process is similar, in spirit, to the qualitative analytical process developed by Marton (1981), but

because the present project employs quantitative as well as qualitative methods, it more closely adheres to the relational methodology described by Overton (Overton, in press).

A relational approach to the research enterprise is deliberately dialectical (and consequently sometimes conceptualized as circular [Fisher, 1990 #4420]). One way in which this dialectic is played out in the present project begins with the attempt to independently determine (1) the hierarchical complexity and (2) the content, of conceptions. Though the hierarchical complexity and content of conceptions do not exist separately, it is possible to distinguish distinct features for each. Once these are highlighted, the interrelations of content and hierarchical complexity are illuminated. A second way in which the relational dialectic manifests in this research is in the alternating employment of interpretive and quantitative methods, each of which provides checks on the other and leads to otherwise unavailable insights into the development of evaluative reasoning about education.

The third goal for this project is to examine the nature of stages. By employing variants of Rasch's models for measurement, I (1) demonstrate that Commons' General Stage Scoring System (Commons, Straughn, Meaney, Johnstone, Weaver, Lichtenbaum, Sonnert, & Rodriquez, J., 1995a) produces results that are compatible with Piagetian stage theory, including evidence of invariant sequence and *structure d'ensemble,* and (2) that the propositional content of performances appears to develop in an invariant sequence across stages, but does not exhibit characteristics

compatible with the notion of *structure d'ensemble.*

The fourth goal of this project is to add to the body of evidence regarding the relationship of stage of performance and age, education, and sex.

Chapter organization

The literature review that follows in Chapters 2 and 3 explores the theoretical issues and empirical evidence central to this study. Chapter 2 is an overview of research on educational concepts, including research conducted from information processing, social cognitive, cognitive developmental, and phenomenographic perspectives. This review is followed, in Chapter 3, by a discussion of cognitive-developmental theory and research in other areas of evaluative thought, with particular attention to cognitive-developmental research on moral reasoning and evaluative reasoning about the good. As mentioned above, one of the aims of this research is to come to a greater understanding of the development of meaning by distinguishing between structure and content. Consequently, in Chapter 3, I examine the controversies surrounding the structure/content distinction, and attempt to provide a rationale for the view that it is not necessary to choose from among contextualist (in which content is privileged), universalist (in which structure is privileged), or holistic (in which structure and content cannot be examined independently) perspectives. I hope, instead, to show that research can be conducted from a fourth, discursive frame of reference, which incorporates and makes use of elements of all three original perspectives.

The methods used in this research are discussed in chapters 4 and 5. Chapter 4

includes background information about (1) the participants, (2) the interview, (3) stage scoring, (4) concept coding, and (5) data management. The General Model of Hierarchical Complexity (GMHC) (Commons, Trudeau, Stein, Richards, & Krause, in press). and its associated scoring system (the General Stage Scoring System or GSSS), without which the attempt to examine developmental stage and conceptual content with some degree of independence would have been impossible, is described in this chapter.

Chapter 5 presents two of the analytical tools used in this project. The first section explores the problem of examining the relationship between the propositional content of performances and their level of hierarchical complexity, and introduces a means for facilitating this end, a computer program developed for this project, called *Sequencer.* The second section presents a rationale for the application of Rasch models in the analysis of data collected for this project.

Chapters 6 and 7 present the results of the several analyses conducted to fulfill the goals of this project. The first step in the discursive research process attempted here involves examining patterns of stage performance in 153 Good Education interviews (Chapter 6). This analysis is conducted without reference to the particular content of performances. Not only does this analysis address the third goal of this project—to examine the nature of developmental stages—it lays the ground for achieving its other three goals: (1) to bring together evaluative and developmental research, (2) to examine the relationship between content and hierarchical complexity, and (4) to explore relationships between the hierarchical

complexity of performances and age, educational attainment, and sex. The fourth goal is directly addressed at the end of chapter 6.

Next, in Chapter 7, the results of the content analysis and the stage analysis are brought together to address goals 1 and 2, and to further flesh out the nature of developmental stages (goal 3). To accomplish these ends, descriptions of evaluative reasoning about education at each GMHC stage are constructed by reintegrating the structural features of each stage with the propositional content of the interviews. Because the enormous amount of data collected for this project made it impossible to conduct an in-depth analysis of the development of reasoning on every conceptual theme identified in the data, only two threads of meaning (metaphorical strands) are analyzed. These are *education as play* and *education as preparation for the good life.* Employing a deliberately iterative strategy, I conduct both quantitative and qualitative tests of the degree to which propositions—originally selected for inclusion on strictly interpretive grounds—contribute to the delineation of a single thread of meaning.

Chapter 8 brings together the results and explores directions for future research.

Chapter 2

Research on educational conceptions

You need more of a community surrounding instead of a classroom—where you have your own desk, and where you can work with other people to learn stuff. I think that would help.

Student, male, 10

This chapter serves two, related purposes. It is intended, (1) to provide a background for the present investigation of evaluative reasoning about education by presenting a review of existing research on educational values, beliefs, attitudes, preferences, and goals, and (2) to establish the importance of evaluative reasoning about education to learning and achievement.

Investigations into conceptions of education, including conceptions of learning, teaching, epistemology, and education in general, have been conducted from a variety of theoretical perspectives. Because each of these perspectives has fostered research relevant to the present project, this review of the literature is wide-ranging. Included in the following review are studies that are predominantly guided by (1) social cognitive theory, (2) information processing theory, (3) phenomenography, and (4) developmental theory. The social cognitive and information processing studies, for the most part, do not take a developmental perspective, though the phenomenographic and the developmental studies do. Consequently, in the following review, studies that are predominantly guided by social cognitive or information processing theory are grouped together, as are those guided by phenomenological or developmental theory.

There is another notable difference between the two groups of studies. Those in the first group are most often concerned with identifying the content of various educational values, beliefs, attitudes, etc., then demonstrating their relationship to learning, voting, teaching, or parenting behaviors and outcomes. Participants in these studies are frequently required to make evaluations of educational practices, institutions, or goals. Studies in the second group are more often concerned with describing forms of educational thought. Only a few of these studies attempt to relate educational thought to other behaviors, or to elicit evaluations of educational practices, aims, or institutions.

Note that the use of terms like concept, conception, attitude, element, preference, belief, and goal, which usually have relatively specific meanings in any given body of literature, are used rather loosely in this chapter. Later in this document, some of these terms will be assigned more particular meanings, but this review, which is focused primarily on commonalities across perspectives rather than differences, does not demand the precise use of terminology.

Social learning and information processing studies

Attitudes

One of the first researchers to investigate educational attitudes was Peterson (1933). He assessed college lecturers' attitudes toward the role of education in several areas, including mental training, the transmission of culture, preparing students for life, the relative contributions of training and talent to teacher excellence, and the origin and nature of moral values. He identified two primary pedagogical

orientations, which he called *traditional* and *progressive.* These orientations were conceived as two poles of a continuum. Since Peterson reported his results, several researchers have identified progressive and traditional dimensions of teachers' beliefs, attitudes, or conceptions (e.g., Cook, Leeds & Callis, 1952; Harding, 1944; Koch, Kentler, Dysart & Streit, 1934; Rose & Medway, 1981; Ryans, 1953; Sontag & Pedhazur, 1972).

Closely related research into educational attitudes was conducted by Kerlinger (1956; 1958; 1966), who also identified *progressive* and *traditional* orientations to education in a sample of professors and non-professors. However, Kerlinger found that these orientations, rather than representing a continuum, were orthogonal. Respondents whose attitudes were characteristic of the progressive orientation valued problem-solving, a focus on children's interests and needs, and warm social relations. They were also inclined to see education as an instrument for social change. Respondents who fit the traditionalist orientation were more likely to stress teacher authority, particular subject matter, moral training, and discipline. In a later study of 344 education graduate students and teachers, Kerlinger (1967) identified seven *second order* factors, each of which loaded onto either the progressive or traditional orientation. Factors associated with the traditionalist orientation were *criticism of the schools, learning as storing knowledge,* and *educational conservatism*, while factors associated with the progressive orientation were *experimentalism, reconstructionism, life adjustment,* and *romantic naturalism.*

Sontag and Pedhazur (1972) reported evidence of traditional and progressive

orientations in his ratings study of the educational attitudes of 356 graduate students. Seven primary thematic factors emerged from the analysis of their responses. These were *education as career training, education as religious or moral training, corporal punishment and education, individualized instruction, subject matter and knowledge, the need for expansion of education*, and *school as a social force.* These factors formed two clusters, corresponding to Kerlinger's traditional and progressive orientations. Participants whose responses were categorized as traditional were more likely to emphasize *corporal punishment, religious training,* and *vocational training,* while participants who were categorized as progressive were more likely to emphasize *individualized instruction, educational expansion,* and *school as a force for social change.*

In other early attitude research, clear traditional versus progressive orientations were not found. When Oliver and Butcher (Oliver, 1962; Oliver & Butcher, 1968) examined teachers' educational values, they found three bipolar factors, *naturalism-idealism, radicalism-conservatism,* and *tendermindedness-toughmindedness.* These did not fit the traditional versus progressive typology. In another study, Wehling and Charters (1969) assessed the educational attitudes and beliefs of 966 teachers. They found that teachers' attitudes clustered into 8 largely interdependent dimensions: *subject-matter emphasis, personal adjustment ideology, student autonomy versus teacher direction, emotional disengagement, consideration of student viewpoint, classroom order, student challenge,* and *integrative learning.* The traditional versus progressive bipolarity was not suggested

by the pattern of interdependence of these 8 dimensions. Finally, Bunting (1984) identified 4 overlapping dimensions of teachers' educational beliefs. These were *affective educational values, cognitive educational values, directive teaching behavior,* and *relevancy in subject matter.* She suggests that individual teachers may simultaneously hold beliefs that, under the traditional versus progressive bipolar assumption, would appear contradictory. For example, teachers who are highly concerned with the emotional development of students may also support values associated with a traditional orientation, like content mastery and student compliance with authority. On the other hand, some teachers who are committed to progressive values may also express concern for the affective development of students.

A shortcoming of attitude research in this area is that researchers do not appear to have a strong theoretical basis for the categorization of sets of beliefs into orientations. They appear to be predominantly guided by the factor analysis of their results, and offer no compelling rationale for the associations among elements, other than the fact that they are found together in enough performances to correlate with one another. Though the contradictory evidence found in this body of research demonstrates that the division of educational attitudes into progressive and traditional orientations (whether conceived as bipolar or orthogonal) does not accurately portray the educational attitudes of individuals, the basic themes of these orientations appear frequently in other educational research. The extent to which individual children's needs are taken into account, the nature of the teacher-

student relationship, the content of instruction, the nature of learning, and the ultimate aims of education appear repeatedly as major themes, both in the literature and in the interviews conducted for this project.

In the analyses of the metaphorical strands, education as play and education as preparation for the good life, in Chapter 7, similar themes figure prominently: ways of thinking about the student teacher relationship, the nature of learning, and the ultimate aims of education. Moreover, several of the conceptualizations associated with the traditional orientation (teacher-directed learning, vocational training, education as career training) appear explicitly in interviews scored at lower stages than interviews containing conceptualizations associated with the progressive orientation (supporting children's interests, warm social relations, education as an instrument for social change, student autonomy, educational expansion, and relevancy of subject matter).

Motivation

The self-reported educational motivations of students have also been the subject of research (e.g., Ames & Ames, 1984; Courtney, 1991; Dweck, 1991; Houle, 1961; Nicholls, Patashnick & Nolen, 1985; Smiley & Dweck, 1994). In his landmark work on motivation, Houle (1961, reported in Cross,1978), identified three sets of motivations for adult students' learning. The first set focuses on learning for the sake of learning; the second on the possibility of social interactions in learning activities, and the third on external goals or objectives. Courtney (1991) reports that individuals with a strong inclination for learning see education as a positive force,

equate it with happiness, and view it as a preparation for solving problems.

Students' educational motivations and goals have been the subject of several studies by Nicholls and Thorkildsen (Nicholls, 1992; Nicholls & Thorkildsen, 1988; Nicholls & Thorkildsen, 1989; Nicholls & Thorkildsen, 1995; Thorkildsen, 1988). In one of these studies (Thorkildsen, 1988), high achieving adolescents and preadolescents were administered a learning scale as well as measures of purpose, motivation, and beliefs about education. Students who (1) valued learning for its own sake, (2) were satisfied with school learning, and (3) believed that success emerges from learning, were likely to assert that school should promote the development of better citizens and unlikely to assert that school should help individuals to attain status. Students who (4) believed that schools should help promote individual wealth and success and (5) supported a competitive orientation to school, were less oriented toward learning and less satisfied with their school experience.

In a second study, Nicholls, Patashnick, and Nolen (Nicholls et al., 1985) examined the relationship between educational values and educational goals in a group of 315 ninth graders and 272 twelfth graders. One of their findings was that students who expressed academic alienation were more likely to view education as a way to gain wealth or status, and that those students who valued education as a way to gain wealth or status were unlikely to value learning for its own sake. In a later report, Nicholls (1992) describes 4 motivational orientations and 5 beliefs that emerged from children's responses to 2 sentence stems: (1) I feel really pleased with

math when...; and (2) students do well in math if... The orientations that emerged were categorized as (a) *effort,* (b) *insight and collaboration,* (c) *ego orientation,* and (d) *work avoidance.* The beliefs were categorized as (a) success requires *interest and effort;* (b) success requires *collaboration and attempts to understand;* (c) success requires *competitiveness;* (d) success requires *conformity;* and (e) success reflects *extrinsic factors.*

The motivation research paradigm suffers from the same problem as the attitude research paradigm, in that work in this area is largely atheoretical. In the Thorkildsen research, for example, there is no overarching theoretical perspective that explains either why elements emerge as components of particular factors, or why factors generated from purpose, motivation, and belief instruments should intercorrelate. Why, for instance, should the concept of learning for the sake of learning and the concern that education should promote the development of better citizens be related? What do two such disparate notions have in common? I submit that at least part of the answer is that these two ideas are constructed at the same order of hierarchical complexity. Evidence for this assertion follows in the review of the developmental literature as well as the results of this project.

Despite shortcomings in the motivation research, the themes identified in this work appear frequently in other research (as will be shown below), including the interviews for the present project. In the analysis of the metaphorical strand, education as play, in Chapter 7, several themes are present: interactive learning, the relationship between student interest and learning, and the idea that learning is

intrinsically valuable. Similarly, in the analysis of the strand, education as preparation for the good life, the following notions are present: education is a positive force, leads to happiness, leads to success, promotes the development of better citizens, and helps individuals attain status.

As claimed above, some of the relationships found between the conceptualizations identified in motivation research can be explained developmentally. For example, the finding, in Nicholls, Patashnick, and Nolen (1985), that students who value education as a way to gain wealth or status are unlikely to value learning for its own sake could be partially explained by the finding, reported in Chapter 7, that the explicit assertion that education is a way to gain wealth or status appears at a lower stage than the idea of learning for its own sake.

The developmental component of attitude and motivation research

The body of research in which educational conceptions are investigated through the assessment of attitudes, beliefs, opinions, and preferences is large and diverse, as can be seen in the sample of studies represented above. As will be seen in Chapter 7, many of these attitudes, beliefs, motivations, etc., are found, often in a variety of forms, in the interviews conducted for this project.

This literature provides conflicting evidence about age differences in educational concepts. Some researchers report agreement between students' and adults preferences and conceptions (Cooper & McIntyre, 1993; Elam, 1978). Other studies reveal few age differences. For example, Lowe and Brock's (1994) finding that graduate students and their professors agreed on 20 of 26 qualities of effective

teaching suggests that there are few differences between the educational values of graduate students and their professors. The elements of good teaching rated more highly by teachers in this study, in particular the tendency of teachers to have more global and far-reaching goals than students, hint at possible developmental effects, but the elements of good teaching that were rated more highly by students, such as more feedback from instructors, indicate more contextual effects. Context effects are also suggested by students' preference for more student control over their education, reported by Elam (1978).

Some of the evidence demonstrates that educational conceptions are malleable, though resistant, to change (Bird, Anderson & Swidler, 1993; Bramald, Hardman & Leat, 1995; Powell, 1992; Wubbels, 1993; Zeichner & Liston, 1987). Unfortunately, this body of research is not concerned with what the direction of change might be when it occurs. Most developmental theories hold that the direction of developmental change is from less differentiation and integration to more differentiation and integration. In the attitudes and motivation literature reviewed here, the occasional classification of educational conceptions into less differentiated authority- and content-based concepts versus more differentiated student-centered and cognition-oriented concepts, hints at the possibility of developmental differences. For example, Kerlinger's (1956) early identification of *traditional* and *progressive* types and Biggs (1984) finding that some students have more differentiated learning concepts than others both suggest this possibility. However, the research reviewed here does not take a developmental perspective

and thus cannot provide clear evidence of the nature of developmental change in educational concepts. For this evidence, I turn to phenomenographic and developmental perspectives.

Phenomenographic and developmental research

Several research programs that focus on the development of educational concepts are particularly relevant to the present project. These can be divided into two groups. The first group of studies are directed by a phenomenographic approach (e.g., Giorgi, 1986; Marton, 1994a; Sälgö, 1979; van Rossom et al., 1985). The second group can be labeled cognitive-developmental, and includes early work by Perry (1970), whose research into developing conceptions of knowledge preceded, and often formed the foundation for, numerous investigations into the development of educational concepts (e.g., Ammon, 1984; Ammon et al., 1985; Basseches, 1984; Black, 1989; Black & Ammon, 1992; Loevinger, 1987; McIntyre, 1992; Tarule, 1978; Tarule, 1980; Weathersby, 1980).

Phenomenographic research

One important approach to the study of developing conceptions of learning is known as phenomenography. According to Marton (1994a), "phenomenography is the empirical study of the differing ways in which people experience, perceive, apprehend, understand, [and] conceptualize various phenomena." Phenomenographic researchers conduct lengthy interviews in which respondents are asked to describe their thinking in considerable detail. These interviews are then analyzed through a rigorous, iterative, interpretive process in order to identify

the different ways in which respondents describe their reflections upon the given topic. The phenomenographic approach to studying conceptions of learning was first used by Sälgö (1979, reported in Marton, Dall'Alba & Beaty, 1993), who found 5 distinct conceptions of learning through analysis of a series of in-depth, personal interviews he conducted with a small sample of college students. These 5 qualitatively different conceptions included: (1) *learning as the increase of knowledge;* (2) *learning as memorization;* (3) *learning as the acquisition, retention, and use of knowledge;* (4) *learning as the abstraction of meaning;* and (5) *learning as an attempt to understand reality through a process of interpretation.* Utilizing the same methodology, Giorgi (1986, reported in Marton et al., 1993) identified 5 types of learning concepts that map neatly onto the categories identified by Sälgö. These were: (1) *rote learning;* (2) *learning as the way information is acquired;* (3) *learning as application;* (4) *learning as understanding;* (5) and *learning as gaining perspective on reality.*

Van Rossum, Deijkers, and Hamer (1985) also used the phenomenographic approach to study the educational concepts of Dutch college students. They asked 42 first and second year Arts students to answer a series of open-ended questions, including *What do you mean by learning?* and *What to you mean by good teaching?* Five *categories of description* were identified, which the authors arrange in a developmental hierarchy as follows:

Category 1: Students conceive of learning as the increase of knowledge, where knowledge consists of an assortment of unrelated facts that are

provided by a teacher. These students also think it is important to understand what they learn, but do not elaborate upon their concept of understanding.

Category 2: Students think of learning as memorization, and have a "bag of tricks" or technological view of good teaching. These students stress the memorization of main issues in their studies in order to be able to answer exam questions. Their concept of understanding is linked to success on exams, and insight is defined as a general sense of understanding.

Category 3: Students see learning as the application of knowledge and express a strong dislike for memorization. Their concepts of good teachers include the ideas that teachers should be organized, should appropriately limit the freedom of students, and should provide students with the opportunity to be "literally active" in their learning. At this level, students' conceptions of knowledge include the idea of reproduction; they think that a good exam is one that offers an opportunity to reproduce or apply their knowledge. Understanding means having an overall or general sense of subject matter, and having insight means having a deep enough understanding to be able to apply one's knowledge.

Category 4: Students conceive of learning as insight, which means attempting to see relationships within and across subjects. Their view of good

teaching emphasizes freedom for students to pursue their own learning and opportunities for learning activities that permit them to construct their own knowledge. Knowledge is still seen as reproduction, but the nature of a good test changes from an opportunity to apply knowledge to an opportunity to construct new meanings by exploring interrelationships in subject matter. Understanding of a text is now conceived as the ability to comprehend an author's intended meaning. Insight involves coming to an understanding of the main concepts in a given subject area.

Category 5: Learning is conceived as personal development. Good teaching involves teacher-student discourse and an increased emphasis on independent learning. Students' concepts of knowledge at this level are an integration of the idea of reproduction and insight. The ideal test question requires critical argumentation. Understanding is now seen as the same thing as insight, which is directed toward the creative use of knowledge and study materials.

Van Rossom and Taylor (1985) and Marton, Dall'Abba, and Beaty (1993) identify the same five categories of description and add a sixth. The conception of learning in this sixth category includes the idea that learning involves a new way of seeing phenomena, and that this new way of seeing changes the learner.

Cognitive-developmental research

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) was able to demonstrate that university students' conceptions of knowledge develop through 9 developmental positions, from the absolutist position that knowledge is either 'right' or 'wrong' to the view that knowledge is relative. Descriptions of these 9 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 view 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 (e.g., Benack, 1983; Cleave-Hogg, 1996; Clinchy, Lief & Young, 1977; Kirk, 1986; Knefelkamp & Slepitza, 1976; Kurfiss, 1977; Widick, 1977) For instance, 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 tend to perform at lower levels (Knefelkamp & Slepitza, 1976; Kurfiss, 1976; Perry, 1970). Benack (1983) employed Perry's scale to examine the relationship between the development of epistemological thought and the development of empathy in adulthood. She found that relativistic thinking about knowledge, characteristic of Perry's highest levels, was associated with high sensitivity to the perspectives of other persons, while absolutist thinking, characteristic of Perry's lowest levels, was associated with low sensitivity to the perspectives of others. Kirk (1986) found that stage of epistemological development, in concert with learning style, was a good indicator of student performance.

King and Kitchener's (1994) large body of work on *reflective judgment* includes an

account of the characteristic pedagogical conceptions of learners at each of five stages (2-6, stage 1 is not elaborated). These are reminiscent of Perry's epistemological positions.

Stage 2: The learner views knowledge as certain and believe it resides in authorities.

Stage 3: The learner views some knowledge as certain and other knowledge as uncertain. What "feels right" is a criteria for determining the veracity of some knowledge, while authority is relied upon for other knowledge.

Stage 4: Knowledge is uncertain, but this is because of a lack of adequate information.

Stage 5: No knowledge is certain, because interpretation and context figure in all understanding.

Stage 6: The uncertainty of knowledge is well established, and truth is understood in relation to evidence and context.

Kitchener and King's model has been used to examine the relationship between epistemological development and academic performance (Schommer, 1993; Schommer, 1994; Schommer, Calvert, Gariglietti & Bajaj, 1997; Schommer & Dunnell, 1994) and to structure several successful educational interventions (Brabeck, 1983; Davison, King & Kitchener, 1990; Finster, 1992; Kitchener & Fischer, 1990; Kroll, 1992; Kronholm, 1994; Nevins, 1992; Welfel, 1982; Welfel & Davison, 1986).

In another study that builds on Perry's levels of epistemological development, Belekny, Clinchy, Goldberger, and Tarule (1986; 1987) explored women's conceptions of learning in their longitudinal study of women's epistemology. They describe 4 distinct "ways of knowing," which they view as occurring in a loose developmental sequence.

Position 1: *Received knowledge*. Authorities are viewed as the source of truth and knowledge is received, retained and returned.

Position 2: *Subjective knowledge.* The source of truth and knowledge is the self. Multiple personal truths are possible, and logic, analysis, abstraction, and language are frequently distrusted.

Position 3: *Procedural knowledge*. Individual, intuitive knowledge can be wrong. At this level, some truths are regarded as more true than others, truth can be shared, abstract truth is possible, and genuine expertise is possible. Procedures, skills, and techniques are emphasized, and the ability to look at things from multiple perspectives is valued.

Position 4: *Constructed knowledge.* At this level, women conceive of all knowledge as constructed. Question-posing and problem-solving take center stage, and conflict, stress, and ambiguity come to be tolerated.

The authors discuss the implication of their findings in terms of education for women, which they conclude should stress acceptance of the learner, respect for real-life experience, the need for more differentiated learning opportunities, and

more connected or contextual teaching. Interestingly, these claims are similar to those made by several researchers in adult education, including Cleave-Hogg (1996), whose work is discussed above, and Chickering (1976).

Other evidence of differences between women's and men's epistemologies have been documented by Baxter-Magolda (1988; Magolda, 1990; Magolda, 1992). In one study, Magolda (1990) found that college men viewed learning as an active process, while college women preferred a receiving mode of learning. Women also emphasized demonstrations of learning and environmental comfort, while men valued interaction and exchange with the instructor and learning challenges.

Loevinger's (1987) work on ego development is also a useful source of information about the development of educational concepts. Her sentence completion test includes an "education is..." sentence stem. Ego development, in Loevinger's scheme, moves through seven levels—*impulsive, self-protective, conformist, self-aware, conscientious, autonomous,* and *integrated.* Conceptions of education have been identified for four of these levels:

D Level: *The self-protective level.* Education is seen as a thing one has to get and then has.

I3 Level: *The conformist level.* Education is viewed as important for getting a job.

I4 Level: *The conscientious level.* Education is seen as part of one's inner life.

I5 Level: *The autonomous level.* Education is seen as supporting understanding

and is viewed as intrinsically valuable.

Several researchers have used Loevinger's framework to further explore educational conceptions. For example, Lasker and de Windt (reported in Lasker & Moore, 1980) report the following hierarchy:

D Level: *Self-protective level.* Individuals think of knowledge as a means to a concrete end. It is given or demonstrated by external authorities is motivated by the desire to satisfy immediate needs. Educational institutions inculcate learning with examples and demonstrations. The teacher's role is to show how, and the student's is to imitate.

I3 Level: C*onformist level.* Individuals see knowledge as the general information required for life. Knowledge is viewed as objective truth, revealed by authorities. The value of knowledge is its utility in gaining social advantages and maintaining expected roles, and the motive for gaining an education is to impress others, gain acceptance, and attain expected or desirable social roles. The function of educational institutions is to provide the "right" information and to rubber stamp learning, which is viewed as a revelation of truth. When ideas conflict, one is right and the others are wrong. Teachers are seen as impersonal presenters of information, and the students' role is to internalize and parrot what is taught.

I4 Level: *Conscientious level.* Knowledge is conceptualized as know-how, and is useful as a means for attaining competence, achieving internal

standards, and acting effectively. The source of knowledge is rational inquiry, which involves goal-setting and question posing, and the motive for education is to achieve competence at the level of personal standards of excellence. The role of educational institutions is to promote the development of skills and provide opportunities for analytical thinking. Learning is a process through which the scientific method and logical analysis are applied. Multiple views are possible, but the truth is viewed as unitary. Teachers, at this level, are models who apprentice the learner by posing questions, inviting discussion, and evaluating competencies, while students engage as subordinates in analysis, critique, and practice.

I5 Level: *Autonomous level.* The student views knowledge as subjective, dialectical, and paradoxical. It is a vehicle through which knowledge of the self unfolds, and can be used to transform both the self and the world. The source of knowledge is personal insight, reflection, and judgment; and the motive for seeking it is understanding along with the desire to manage one's destiny. The role of educational institutions is to provide the learner with new experiences, questions, paradoxes, and dilemmas with the intention of fostering insight. Learning is seen as a dialectical process of reorganizing ideas into new systems of relations. The teacher is seen as a facilitator and resource, and works as an equal participant with the student, who is expected to generate her own

interpretations and meanings.

Other researchers have investigated the relationship between Loevinger's ego development scale and educational conceptions. For example, in her case study of 6 elementary school teachers, Witherell (1978) found that teachers' educational constructs and teaching practices were both related to levels of ego development, and Jesness (1972) found that learners at different ego development levels had different learning preferences. Additionally, when Weathersby (1977; 1980) interviewed 67 adult students about their educational values, she found that ego levels predicted study motives, learning style, and learning environment preferences. For example, adults who scored at or above Loevinger's individualistic stage were more likely to appreciate learning from real-life experiences and to report consciously taking control of their own learning processes.

Several researchers have examined the relationship between *post-formal* modes of adult thought and educational concepts. (See Chapter 3 for a discussion of post-formal models.) For example, Sinnott (1994a) utilizes the principles of what she calls the *new post-formal sciences*—relativity and quantum physics, cooperative evolution, and postformal cognition—to hypothesize about the epistemology of adult learners. Sinnott speculates, that from the post-formal learner's perspective, (1) the definition of truth is dependent upon the vantage point and chosen constructs of the viewer *(the new physics;)* (2) individuals are seen both as important in themselves and as part of a larger system *(the new biology);* and (3) one still must commit to action decisions even though realties are relative *(the new*

cognitive science).

In another paper, Sinnott (1994b) examines the implications of differences between formal and postformal thought for conceptions of learning. At the formal level, basic processes such as formal operations and memory structures, and basic facts are the focus of learning. Ultimately, through exposure to diverse perspectives, the thinker comes to understand that there are multiple views of reality. This leads to a restructuring of learning conceptions. At this point, thinkers begin to focus on process and change as constants, rather than isolated facts. This allows them to coordinate multiple realities.

In one of his studies on the development of dialectical thinking—another post-formal mode of thought—Basseches (1984) interviewed a group of 6 college faculty members, 9 freshmen students, and 9 senior students about their conceptions of education. Performances were assigned to one of three groups, *formal, transitional,* and *dialectical.* Formal performances were characterized by linear logic, while dialectical performances were characterized by more complex, discursive logic. Clear differences were found between the educational conceptions of freshmen, seniors, and faculty members. Faculty members had the greatest percentage of dialectical performances, followed by seniors and freshmen. Unfortunately, Basseches does not report on the content of conceptions at each of these levels.

In a departure from most cognitive developmental studies, in which differences between stages are the focus of study, Tarule (1978; 1980) examined college students' reports of their experiences of learning during stage transitions. She found,

that during transitional periods, college students report being conscious of a state of confusion or dissonance when new knowledge and old knowledge don't fit together, and that this confusion gradually disappears as old and new knowledge are integrated. Tarule's strategy was also used by Taylor (Taylor, 1986; Taylor, 1987) who interviewed 12 adult students on several occasions in order to examine their understandings of the learning processes in which they were engaged. She identified 4 phases of learning transitions from an examination of these accounts, including *disorientation, exploration, reorientation,* and *equilibrium.*

Teachers' educational conceptions have also been the focus of developmental investigations. For example, McIntyre (1992) proposes three hierarchical levels of pedagogical reflection by teachers.

Level 1: *The technical level.* Teachers' educational concepts emphasize the achievement of specific goals, maintenance of classroom control, and keeping student attention.

Level 2: *The practical level.* Teachers' concepts are focused on the development of their own educational ideas, objectives, and practice.

Level 3: *The critical level.* The teacher's educational concepts embrace wider ethical, societal, and political concerns.

According to McIntyre, the third level of teacher reflection is rarely found, even among experienced teachers. Using McIntyre's framework, Bramald, Hardman and Leat (1995) found that student teachers classified as traditional in their educational

beliefs were likely to reflect at the technical level, while teachers who were classified as student-centered were more likely to reflect at the practical level.

Several researchers have conducted cognitive-developmental investigations of teachers' pedagogical conceptions (Ammon et al., 1985; Black & Ammon, 1992; Jakubowski & Tobin, 1991; Levin & Ammon, 1992; Oja, 1991; Oja & Ham, 1984). In their longitudinal examination of student teachers' pedagogical conceptions, Black and Ammon (1992) identified 5 levels of teachers' reasoning about instructional goals, learning requirements, and the nature of teaching.

Level 1: The goal of teaching is to provide the student, whose ability and receptivity determine his learning capacity, with a body of facts and procedures through lecture and demonstration.

Level 2: The goal of teaching is to provide, in an appropriate sequence, the skills needed by the student in order to learn and utilize facts and procedures. This is accomplished by giving students the opportunity to practice and offering both positive and corrective feedback.

Level 3: The goal of teaching is to impart, using developmentally appropriate means, correct understandings of the concepts that underlie facts, procedures and skills in a given subject area. This is accomplished by providing the opportunity for students to work with materials that are appropriate to their developmental level.

Level 4: The goal of teaching is to help students to improve their conceptual

understandings. This is accomplished by engaging students in developmentally appropriate activities that stimulate thinking and improve understandings.

Level 5: The goal of teaching is to guide students toward ways of thinking that enhance understanding. This is accomplished by encouraging students to reflect on their own thought processes.

Ammon, Black, and their colleagues (Ammon & Hutcheson, 1989; Ammon & Levin, 1993; Ammon & Peretti, 1995; Black, 1989; Black & Ammon, 1992; Black & Phillips, 1982) have also found that teachers' pedagogical thought develops over the course of a developmental teacher education program and that teachers' pedagogical conceptions are related to teaching practice. The latter finding is corroborated by Oja (1990), who reports, in her review of four studies of teachers' pedagogical knowledge, that teachers who perform at higher stages of development demonstrate greater flexibility, are more likely to view situations from multiple perspectives, and are more effective in the classroom and as supervisors and collaborators. Further, Johnston (1989) reports that teachers who perform at a higher stage of moral reasoning have a more differentiated understanding of individualized instruction. Other researchers have also found a relationship between moral stage and teachers' educational understandings (Johnston & Lubomudrov, 1982; Lubomudrov, 1982; Wheaton, 1984).

The phenomenographic and cognitive-developmental evidence presented here suggests that developmental change explains some differences between educational conceptions. Developmental changes in these concepts can be modeled hierarchically. In Table 1, several developmental hierarchies are displayed. Roughly analogous levels or stages are aligned on the horizontal dimension, while the vertical dimension indicates the order in which various conceptions are expected to appear during development.

Though it is not immediately apparent from the stage and level names in Table 1, qualitative comparison of conceptions along the horizontal dimension reveal similar underlying meanings. For example, at Van Rossum, Deijkers, and Hamer's (1985) and Marton, Dall'Abba, and Beaty's (1993) "learning as memorizing" level, students think of learning as memorization, and view good teaching as the application of proper technique in the distribution of factual knowledge. A similar, though perhaps somewhat more sophisticated concept of teaching, is found at Black and Ammon's "content" level, at which the goal of teaching is to provide the student with facts and procedures through lecture and demonstration. Both of these descriptions are similar to Belenky, Clinchy, Goldberger, and Tarule's (1986) description of learning conceptions at the "received knowledge" level, where authorities are viewed as the source of truth, and knowledge is received, retained and returned; Perry's description of his "basic duality" stage at which the student views knowledge as an accumulation of facts; King and Kitchener's (1994) "knowledge as certain and received" stage, at which the learner views knowledge

as certain and believe it resides in authorities; and Loevinger's (1987) "education as preparation for a job" level, at which learning is seen as a thing that can be inculcated into the learner. This notion is elaborated at Laskar and deWindt's "conformist" level. At this level, knowledge is viewed as objective truth, revealed by authorities, and internalized and regurgitated by students. Across all of the developmental schemes reported here, knowledge at this level is viewed as absolute, held by authorities, and learned through a process of internalization or memorization.

At the following level, Van Rossum, Deijkers, and Hamer, and Marton, Dall'Abba, and Beaty report that students see learning as the application of knowledge rather than memorization; think of knowledge, which is still seen as largely absolute, in terms of reproduction or application; and think that teachers should be organized, maintain appropriate control, provide feedback, and encourage active learning. The conception of teaching at Black and Ammon's "using content" level contains similar elements. Here, good teaching provides students with the skills they need to learn and apply facts and procedures and includes both positive and corrective feedback. At Perry's "multiplicity subordinate" position, the student recognizes that there is diversity and uncertainty of opinion, but believes this is because the "truth" has not yet been discovered. Somewhat differently, at King and Kitchener's stage 2 the learner thinks of some knowledge as certain and other knowledge as uncertain. Both authority and what "feels right" can determine the truth of knowledge. Belenky, Clinchy, Goldberger, and Tarule's "subjective knowledge" level has a

slightly different flavor. At this level, having realized that authorities have different opinions, women come to locate the source of truth in the self's intuitive knowledge.

At this level, as at the first level, knowledge is viewed as absolute, but the belief that authorities hold the truth has been undermined by the realization that authorities do not always agree. Respondents have different ways of explaining this. Some hold to the belief that some authorities are more correct than others, or that the truth has simply not yet been found. Others believe that some truths are more certain than others, or that the self's intuitions are as good a source of truth as any other. Learning is a more active process than at the previous level, primarily due to a new emphasis on practicing with or applying knowledge.

At the next level, according to Van Rossum, Deijkers, and Hamer, students conceive of learning as insight, which is equivalent to finding relationships within and across subjects. They emphasize the independent pursuit of learning as an exploratory activity, and value the opportunity to construct new meanings, but still see knowledge in terms of reproduction. At Perry's, "multiplicity coordinate" level, students believe that everyone has a unique opinion, though they continue hold to an absolutist position on the ultimate nature of right and wrong. Knowledge is viewed as contextual, but only because the "answers" have not yet been discovered. Similarly, at Kitchener and King's stage 4, knowledge is viewed as uncertain, but this is because of a lack of adequate information. At Laskar and deWindt's "conscientious level," as in the previous examples, multiple views are

seen as possible even thought the truth is viewed as unitary. Knowledge is conceptualized as a means to competence and effectiveness and is acquired through rational inquiry, leading, ultimately, to the truth. Learning is seen as an active process. In a similar vein, at the "procedural knowledge" level of Belenky, Clinchy, Goldberger, and Tarule, women come to the conclusion that, though there are multiple perspectives on truth and the ability to take multiple perspectives is valuable, some truths are more true than others. Learning is seen as an active process in which procedures, skills, and techniques are emphasized. Finally, the goal of teaching, at Ammon and Black's "underlying concepts" level, is to impart correct understandings of the concepts underlying facts, procedures and skills. Teachers performing at this level believe that students should have opportunities to work actively with developmentally appropriate materials.

Across all of the developmental schemes at this level, respondents simultaneously hold the views that (1) there are absolute or ultimate truths and (2) that knowledge is contextual. Whereas, at the previous level, respondents explained away the discovery of disagreement among authorities by either attempting to figure out who had the "right" truth or by giving up on authorities and reason altogether and locating a "good enough" truth in personal intuition, respondents at this level view truth-seeking as a process of construction that will eventually lead to truth. In light of this changing understanding, it is not surprising that learning at this level is seen as an active process of investigation that involves taking into account multiple perspectives.

The highest level of performance that is well-represented in Table 1 is aligned with Commons' "systematic" stage. At this level, according to Van Rossum, Deijkers, and Hamer, knowledge is no longer viewed as absolute. Students' concepts of knowledge now integrate the ideas of knowledge reproduction and insight, and are directed toward the creative use of knowledge and study materials to enhance understanding. In this context, *integration* and *creativity* imply knowledge construction, a concept that is more explicitly present in Belekny, Clinchy, Goldberger, and Tarule's (Belenky et al., 1986; Goldberger et al., 1987) description of their "constructed knowledge" level, at which women conceive of knowledge as constructed, and question-posing and problem-solving are valued (and creative) learning strategies. At Perry's positions 6 and 7, the student accepts that all knowledge is relative, and as a consequence, that it is necessary to take responsibility for one's choices and commit to a position within a relativistic world, while at King and Kitchener's stage 5, the individual comes to accept that interpretation and context figure in all understanding. The themes of construction, creativity, and integration are also implicated here. Similar themes are found at Ammon and Black's "improved understandings" level. Here, the goal of teaching is to guide students toward ways of thinking that enhance understanding. This is accomplished by encouraging students to reflect on their own thought processes, thus enhancing students' awareness of their own way of constructing meanings.

At this level, respondents have let go of the idea of absolute truths, still evident at the previous level. They view knowledge as fully contextual, relative, and

constructed. However, this is not the "my truth is as good as any truth" contention present at the second level elaborated above. The present conception incorporates the idea that the way to better understandings involves engaging in creative inquiry and reflection. It is more of a "we are all looking for better understandings" contention.

Though the above discussion demonstrates that it is possible to make meaningful qualitative comparisons of levels from independently constructed developmental sequences, the problem of whether these comparisons are quantitatively verifiable is as yet unresolved (Commons et al., 1989; Commons & Grotzer, 1990; Commons et al., 1995b; King, Kitchener, Wood & Davison, 1989).

The developmental sequences described in this discussion were constructed with a range of methodologies, and were addressed to a number of content areas, including conceptions of learning, knowledge, and good teaching. There are, unsurprisingly, several differences between the *categories of description* (to borrow Marton's phrase; Marton, 1981) depicted by different researchers. However, the above comparisons of these categories of description at analogous levels demonstrate that some common conceptualizations run across schemes. Moreover, some of these common conceptualizations are suggestive of distinctions identified by researchers represented in the first section of this chapter. For example, the early distinction between *traditional* and *progressive* orientations introduced by Peterson (1933), appears to be played out in the extreme ends of these developmental progressions. The lower level conception of knowledge as fixed, factual, and

transmitted is loosely reminiscent of the factors associated with traditionalism that were identified by Kerlinger (1967), *criticism of the schools* (not focused enough on the basics), *learning as storing knowledge,* and *educational conservatism* (children should listen, teachers should maintain discipline), while factors associated with the higher levels are reminiscent of his progressive orientation factors, *experimentalism, reconstructionism, life adjustment,* and *romantic naturalism.*

Several of the developmental schemes presented above are concerned with changes in educational conceptions during adulthood. In most cases, longitudinal evidence establishing the direction of these changes is available (Ammon et al., 1985; Belenky et al., 1986; King, Kithchener & Wood, 1992; Marton et al., 1993; Perry, 1970). Whether the kind of change reported by these researchers is comparable to the developmental change witnessed in children's concepts is one of the questions raised by the present study of evaluative reasoning about education. Further consideration of this question takes place in Chapter 3.

One effect of taking development into consideration is the added explanatory power provided when hierarchical complexity—or level of differentiation and integration—are taken into account. Knowledge of developmental change adds another dimension to our way of framing differences in learner preferences and teacher's pedagogical conceptions. Moreover, the hierarchical nature of developmental sequences suggests the appropriate direction for, and to some extent, even the form of, interventions designed to bring about particular understandings.

Differences in educational conceptions occur for both developmental and contextual (informational, cultural, dispositional) reasons. Consequently, this review of the literature examined research from both developmental and non-developmental perspectives. None of the studies reported above is designed to distinguish between developmental and non-developmental influences, though Belekny, Clinchy, Goldberger, and Tarule (Belenky et al., 1986; Goldberger et al., 1987) do attempt to identify the influence of sex-related contexts on developing epistemologies. One of the objectives of the current project is to begin the process of sorting out the influences on developing conceptions by examining both hierarchical and horizontal dimensions of difference. Relevant questions include: (1) What are the themes or strands of meaning that appear to run across developmental stages? (2) How are particular conceptualizations related to developmental stages? and (3) How are developmental stage and sex, age, and educational attainment related?

Chapter 3

Evaluative reasoning from a structural-developmental perspective

A good education makes you more capable. If you can't do anything yourself then you become really incapable; you're just sort of one of the people in the world that takes stuff but never give anything back.

Student, Female, 12

As demonstrated in the last chapter, there is a significant body of research on the relationship between various educational values, thoughts, and attitudes, and behaviors such as achievement, teaching practices, political orientation, and student response to educational interventions. Moreover, evidence was presented that supports the notion that cognitive development plays a role in both the form and content of respondents' educational conceptions. In the first section of the chapter, the studies presented were largely unconcerned with developmental influences on educational thinking. They focused on identifying individual attitudes and motivations and examining the possibility that these formed themselves into clusters or typologies that could describe different orientations toward elements of the educational domain. These studies, though they are a good source of information about the content of the educational evaluations and dispositions of respondents, are a poor source of information about how concepts change over time. The second half of the chapter was concerned with developmental studies. These studies highlighted the importance of taking the developmental dimension into account, but, in doing so, they lost sight of possible similarities across

developmental levels, such as general dispositions, as well as possible differences within levels, such as the specific content that a given individual might organize at that level. The central issue probed in the present chapter is whether it is possible to conduct a cognitive developmental study in which form and content share center stage—to the extent that both developmental and non-developmental differences can be fruitfully examined.

How are conceptions of good education influenced by developments in cognition? To most developmental researchers, the question posed above would appear circular. This is because most researchers define levels or stages of development in terms of their conceptual content. In order to sensibly ask the question posed above, one must first assume that there is something taking place in the developmental process that is not synonymous with conceptual content. Interestingly, in the light of this assumption, it is defining stage in terms of particular conceptual content that appears tautological—how are we to test our assumption that particular conceptual content is characteristic of a specific stage of development if the stage is defined in terms of that content?

In fact, there is more than one problem with defining stage in terms of particular content. The most significant of these is a measurement problem. Though stages of development are often spoken of as having universal properties—often described in terms of hierarchical complexity (Commons, Piaget)—that hold across domains of reasoning, few efforts have been made to establish a method that can be used to measure these properties across domains. One consequence is that both qualitative

and quantitative efforts to compare developmental stages across domains are fraught with problems (Cavanaugh & McGuire, 1994, Kitchener, King, Wood & Davison, 1989).

To express the problem another way, we presently have a theoretical construct, stage, which has universal properties, in that aspects of the construct are defined similarly across domains, but we have a separate ruler for every domain in which stage is assessed and no satisfactory means for comparing rulers across domains. Using conventional methods we can neither assess whether criteria for determining stage across domains are equivalent, nor whether some developmental changes in reasoning can be successfully modeled as universal "stages."

When developmental stage is assessed in terms of its universal features (hierarchical structure) rather than its contextual features (domain specific content), it becomes possible (1) to meaningfully compare development across domains, and (2) to examine the relationship between developmental stages and conceptual content. The following literature review and discussion examines the cognitive developmental literature to establish a theoretical and empirical basis for an approach to the study of development that differentiates between structure and content.

Because Piaget's (1970; 1985) work exemplifies the cognitive developmental paradigm, a brief sketch of his theoretical perspective opens the discussion. Then, the way in which Piagetian constructs are conventionally operationalized outside of the logico-mathematical domain is investigated by reviewing structural

developmental research on evaluative reasoning, in particular the work of Colby and Kohlberg (Colby & Kohlberg, 1987a), and Armon (1984b). Third, some of the problems confronted by these researchers are discussed, including (1) the limitations of content-dependent stage scoring systems, including cultural bias, restrictions on what constitutes "scorable" content, and the potential for over- and under-estimating ability; (2) the difficulty of comparing stages of performance across domains; and (3) the discovery that some stages of development do not appear before adulthood,. This third "problem" is explored here because the question of whether the same kinds of change characterize development in childhood and adulthood is central to the problem of defining universal developmental criteria.

The chapter concludes with discussion of an alternative approach to stage assessment, and a possible solution to the content/structure conundrum.

Cognitive developmental theory

Piaget thought of his work as both constructivist and structuralist. *Construction,* in the simplest terms, is the activity of meaning-making. *Structures*, as conceived by Piaget, are open systems of transformations, composed of elements and the relations between them. These elements can include operations, and substructures such as schemas, concepts, perceptions, and memories, all of which are interconnected through spacio-temporal, causal, implicative, and other relations (Piaget, 1970). Elements, while they constitute the *content* of more global structures, have a structure of their own. For example, a proposition such as, "A

good education is one in which you learn what you need to make a living," has structural features. One of these is the connection established between two concepts, learning and making a living. However, this proposition is nested within a larger argument with a set of interconnections that, taken together, can provide insight into the over-arching structure employed by an individual reasoning at a given stage of development. In other words, content, such as propositions and individual concepts, can include forms and operations that are incorporated within a more general structure. The many layers of structure present in performances can make it difficult to determine where one ought to look for evidence of the global reasoning system employed by an individual, and can be an obstacle to differentiating between structure and content.

A concept closely related to *structure*, in Piaget's theory, is *procedure* (Inhelder, 1982) and (Inhelder & Piaget, 1979; Piaget, 1977a, cited in Vuyk, 1981a). Procedures are conceived as the mechanisms employed by the psychological subject in order to arrive at specific cognitive goals. Structures and procedures form the poles of a dialectic, in that structures are constructed through the application of procedures, and procedures utilize aspects of structures in the act of construction. At qualitatively different levels of structure, different groups of procedures are available. The deepest level of structure, at the level of cognition, can only be *inferred* from the organization of performances. Most developmental research on evaluative reasoning is predicated on the supposition that this inference is meaningful.

The concept of equilibration is central to Piaget's theory. According to Piaget (Piaget, Grize, Szeminska & Vinh, 1977), thought is organized into structures called schemas. Groups of these are further organized into subsystems, which are, in turn, part of a total system. Each schema is composed of a set of associations. For example, infants develop, during the first few months of life, a reaching and grasping schema that incorporates the set of acts required to reach for an object and hold on to it. During development, such schemas become more differentiated and are coordinated with one another in increasingly complex ways. For successful grasping to occur when the object of interest is in motion, the reaching schema must be coordinated with the visual tracking schema. Subsystems are composed of schemas that are organized and integrated at a new level of complexity. At the subsystem level the reaching and tracking schemas integrate to form a reaching-while-tracking-subsystem. All of these schemas and subsystems are part of the total cognitive system.

Schemas, subsystems, and the total system are organized knowledge systems or structures. They are orderly systems with both dynamic and stationary qualities. Their openness or dynamism results from interactions with the environment; their closed or stationary quality results from the cyclic nature of their functions. When such systems experience external perturbation they must either adapt or they can no longer function. For Piaget, equilibration is the essential process of adaptation. It takes three forms. The first of these operates between objects or events in the environment and schemas; the second involves interactions among schemas and

subsystems; and the third involves interactions between schemas and subsystems and the total system. The two universal processes of equilibration, which work on all three of the above levels, are *assimilation* and *accommodation*. In assimilation, the organism incorporates an object or event into an existing schema, or coordinates existing schemas to perform a function (reciprocal assimilation). Accommodation occurs when the schema or subsystem must be altered to integrate an object or event, or when multiple schemas or subsystems must be altered to perform a function (reciprocal accommodation).

According to Piaget (1977a), stage change is related to accommodation at the third level of equilibration. It takes place when the organizing actions (operations) of a developmental stage become the content of new organizing actions. "The construction of operations on operations is probably the secret of development and of the transition from one stage to the next " (p. 839).

In the structural developmental literature, stage transitions are often conceptualized as qualitative transformations or *hierarchical integrations* of cognitive structures (e.g., Case, 1985; Commons, Richards, with Ruf, Armstrong-Roche & Bretzius, 1983; Demetriou & Efklides, 1985; Fischer, 1980; Kohlberg, 1969). At least one researcher refers to stages as *orders of hierarchical complexity* (Commons, Trudeau, Stein, Richards & Krause, in press), a designation that is much more descriptive than the term *stage.* (I use the terms *order of hierarchical complexity, order,* and *stage* interchangeably in this document.) Moreover, stages are said to follow one another in an *invariant sequence*. Indeed, if the operations of a subsequent stage

organize the operations of the previous stage, then invariant sequence is a logical necessity.

A final, frequently mentioned concept associated with stage is *structure d'ensemble.* This concept is often interpreted as a reference to a homogeneous global structure that governs all thought processes at a given stage. Piaget's use of this construct, however, appears to refer more to the characteristics of a given stage that are common across individuals, rather than a psychological state. Development is a dynamic process in which various cognitive structures are continually differentiated and reintegrated into a perpetually reorganizing global system (Vuyk, 1981a). In fact, for Piaget (1977a) perfect equilibrium, which is a static state, is not a possibility, since as long as learning takes place, equilibrium will be disrupted by the need to incorporate new knowledge. "There is a continual search for a better equilibrium...Equilibration is the search for a better and better equilibrium in the sense of an extended field, in the sense of an increase in the number of possible compositions, and in the sense of a growth in coherence" (p. 840).

Piaget moved further from the concept of stages as his work matured. Even during his early work with stages, he was careful to point out that *stage*, for him, was useful as a taxonomic device only. "I must vigorously insist on the fact that stages do not constitute an aim in their own right." Stages, however, are "indispensable...for the analysis of formative processes" (quoted in Gruber & Vonèche, 1977, page 817). It has been difficult to keep stages in their place,

however, as the enormous literature devoted to stage differences attests.

Despite these problems, the concept of stage continues to provide useful heuristics. Piaget's major stages, sensorimotor actions, early and late preoperations, early and late concrete operations, and formal operations, have been useful to researchers investigating developmental change in a wide variety of domains. For example, Kohlberg (1969) built on Piagetian concepts to construct his moral judgment stages, and Commons' General Model of Hierarchical Complexity (GMHC) (Commons & Miller, in press; Commons et al., 1983), which is based on structural criteria that Commons' claims hold across all domains of reasoning, was strongly influenced by Piaget's stage theory. Moreover, the concept of stage has permitted researchers to examine the development of reasoning in diverse domains, with several researchers reporting qualitative similarities between Piagetian logico-mathematical stages and stages of reasoning in other domains (e.g., Armon, 1984b; Commons & Grotzer, 1990; Commons, Miller & Kuhn, 1982; Damon, 1989; Fischer & Ayoub, 1996; Fowler, 1981; Kegan, 1982; Kohlberg, 1969; Noam, 1988), suggesting that the stage model formulated by Piaget has a useful role to play in efforts to understand the development of knowledge.

Some notes on the invariant sequence criterion and *structure d'ensemble*

Though the contention that stages follow one another in an invariant sequence has been supported by a large body of research across multiple reasoning domains (Armon & Dawson, 1997; Case, Okamoto, Henderson & McKeough, 1993; Colby, 1981; Colby, Kohlberg, Gibbs & Lieberman, 1983; Damon, 1980; Davison, King,

Kitchener & Parker, 1980; Fischer, 1980; Flavell, 1992; Holstein, 1976; Selman, 1977; Siegler, 1981; Turiel, Edwards & Kohlberg, 1978; Vasudev & Hummel, 1987), it has also been the target of criticism. Most of this criticism has taken one of two forms. Some authors, like Brainerd (1978), reject the notion that invariant sequence is evidence of the existence of stages, claiming that because invariant sequence is part of the definition of stage, the argument that evidence of invariant sequence is evidence in support of the existence of stages is tautological. A second criticism of the invariant sequence criterion results from rejection of the stage model as conceived by Piaget. Those who make this criticism deny that epistemological development necessarily occurs in a uniform hierarchical fashion and assert that a more adequate explanation must take into account the variety of dispositional, contextual, and informational contingencies that lead to a range of developmental pathways (see, for example, Fischer & Silvern, 1985; Flavell, 1971; Pascual-Leone & Goodman, 1979).

It seems clear that no stage model can adequately explain everything about development. However, if we accept the definition of stage change as the hierarchical integration of operatory structures, invariant sequence is a logical necessity. (Given this, Brainerd's argument, above, is well-taken.) Accepting this definition does not reject the argument that dispositional, contextual, and informational contingencies affect the trajectory of development or the content of thought. As will become apparent, the research design employed here is intended to detect both stage-related and stage-independent aspects of cognition.

Piaget's concept of *structure d'ensemble* has frequently been criticized (e.g., Brainerd, 1978; Flavell, 1971; Pascual-Leone & Goodman, 1979). One line of argument holds that the empirical evidence supports a continuous model of individual development over a stage model. In this view, the concept of *structure d'ensemble* is incompatible with the notion of continuous development, because it implies that development progresses through a series of qualitative shifts followed by stable periods at which all or most reasoning is organized in the same way. This type of criticism, unfortunately, assumes a mechanistic conceptualization of stage that is unlike the model proposed by Piaget in two important ways. First, it assumes that periods of perfect equilibrium take place, whereas Piaget clearly argued that equilibrium is never attained. Second, it applies a model used to describe development in the epistemic subject to description of the individual psychological subject, something Piaget warned against.

As Piaget argued, the concept of stage is a heuristic device, not something to be construed as factually true or not true. Viewed from this perspective, the idea of *structure d'ensemble* might also be a useful construct. For example, it seems reasonable to propose that knowledge, at least within a single domain, might be organized such that individuals are unable to integrate knowledge at a level beyond their current modal stage before they have organized their existing knowledge according to the principles of that stage. This view is consistent with Flavell's (1971) second model of stage change (which he rejected in favor of a much less restricted model), and certainly seems like a reasonable possibility if it is

agreed that stage change involves operations on operations and that knowledge structures seek equilibrium. It is difficult to conceive of an equilibrating system in which qualitative restructuring at a global level can proceed more rapidly for some knowledge than for other knowledge coordinated *within the same system*. All of the knowledge coordinated within a dynamic system is interconnected such that structural changes in one part of the system must simultaneously propel and restrain the organization of the system as a whole.

Because of the dynamic nature of Piagetian knowledge systems, evidence for a *structure d'ensemble,* as conceptualized here, would be difficult to observe in the individual subject. First of all, it seems unlikely that agreement could be reached about what would constitute evidence of a structured whole within individuals. Some have argued that more or less consistent performance at a given stage is an adequate determinant (Walker, 1984a), but this brings up several questions. (1) Does performance have to be consistent within individuals across domains of reasoning or only within a single domain? (2) What constitutes a domain? and (3) If consistency across domains is required, are stage assessments in different domains comparable? To add to the difficulties, *Structure d'ensemble* has been an extremely awkward construct to operationalize, primarily because of the difficulty of comparing development across domains. One of the aims of the present research is to develop a means for determining stage that permits meaningful cross-domain comparisons of developmental progress.

The concept of *structure d'ensemble* was intended to apply to the *epistemic* subject

rather than the psychological subject. As mentioned earlier, the epistemic subject represents the common characteristics of stage across individuals at the same developmental level. Estimates of the relative difficulty of stages and stage transitions, constructed on the basis of data from many individual subjects, can be used to characterize the epistemic subject more effectively than estimates of individual functioning. Analytical methods exist that can be used to test the hypothesis that it is more difficult to construct new structures than to extend the use of existing structures. Evidence of this kind, presented below in Chapter 6, is consistent with the hypothesis that knowledge systems seek a state of relative equilibrium by employing available structures before entirely new operations are constructed.

This approach still leaves open the question of how global this tendency must be to support the Piagetian conception of *structure d'ensemble*. This is less problematic than it seems. Piaget's concept of *structure d'ensemble* was never intended as a claim that reasoning in individuals is at a single stage across domains. Instead, it refers to the common logical structure of the operations characteristic of a stage. If we accept this interpretation of Piaget, then the question of whether all the thought of individuals is ever equilibrated at a single stage should not be associated with *structure d'ensemble*.

According to van Geert (1993), "the more one knows [in a given domain], the easier it becomes to acquire more of that knowledge" (p. 278). If this is the case in the additive accumulation of knowledge, might it also be the case in the qualitative

reintegration of knowledge? Might one assert that once an individual begins to restructure knowledge at a new stage, it becomes easier to restructure other knowledge in the same way? Such a phenomenon would be consistent with the conception of cognitive systems sketched in the previous paragraphs, as well as with a domain-sensitive conception of *structure d'ensemble*.

To explore Piagetian constructs like structure d'ensemble and the distinction between the epistemic and psychological subject, it is helpful to differentiate between structure and content. Attempts to study the continuity of reasoning structures across domains or to examine the universal features of reasoning embodied in the stage construct are clouded when content and structure are confounded. This is apparent in the following review of cognitive developmental research in evaluative reasoning.

The Piagetian constructs examined above hint at ways in which the distinction between content and structure can be operationalized, even though they are interrelated in complex ways. Piaget's distinction between levels of structure, in which he defines substructures as content with respect to global structures (stages), points to one way to operationalize these two constructs. Stages can be defined as orders of hierarchical complexity, and their presence can be inferred from the organizational structure of arguments. The particular propositional content of these arguments can then be defined as content. In this way, the most stage-like, universal features of performance are differentiated from the more context-dependent features of performance.

Cognitive developmental research on evaluative reasoning

Numerous researchers and theoreticians working within the cognitive developmental paradigm have conducted research into the development of evaluative thought (Armon, 1984a; Armon, 1984b; Colby & Kohlberg, 1987a; Gilligan, 1982; Keller & Wood, 1989; Rosenberg, 1988; Selman & Schultz, 1990; Turiel, 1983). Arguably, the best known of these investigators was Kohlberg (1969) whose controversial approach to the study of moral development revolutionized developmental research outside of the logico-mathematical domain. Kohlberg's work on moral judgment and Armon's (1984b) work on evaluative reasoning about the good are important to the present project for several reasons. First, they exemplify cognitive-developmental research in this area. Second, there is overlap between the research methodologies employed by Kohlberg and Armon, and those used in the present project. For instance, Kohlberg's "bootstrapping" process, an iterative approach to developing stage definitions, resembles aspects of the relational approach employed in this research, and the clinical interview method used to collect the data for the present project was developed by Armon from Kohlbergian and Piagetian interview procedures. Third, Armon and Kohlberg's work can be used to highlight both the strengths and weaknesses of cognitive developmental research on evaluative reasoning, including problems stemming from the failure to differentiate between structure and content in the definition and determination of developmental stages. Finally, several of the respondents involved in the present project also participated in Armon's life-span longitudinal study of

moral reasoning and evaluative reasoning about the good, which has made it possible to examine the relationship between measures of moral reasoning, evaluative reasoning about the good, and evaluative reasoning about education.

Moral judgment

Kohlberg (1971) combined a Piagetian understanding of developmental processes with a philosophical analysis of the moral domain to develop his theory of moral development. His instrument, the Moral Judgment Interview (Colby & Kohlberg, 1987a; Colby & Kohlberg, 1987b), consists of a series of moral dilemmas that focus on a variety of moral issues, most of them concerned with problems of rights, responsibilities, and justice. Participants in Kohlberg's studies were not only asked what was the right thing to do, but they were required to provide justifications for their responses. This probing of respondents' reasoning was designed, not only to generate adequate material for scoring, but also to elicit participants' highest stage of performance or competence.

Kohlberg's stage-scoring system was developed through a dialectical "bootstrapping" process between philosophical/theoretical constructs and empirical findings. Early in this process, philosophical concepts held sway over structural concepts (like hierarchical complexity), in that his scoring rubric almost exclusively relied on the philosophical content of responses. By the time the third version of the scoring system was developed (Colby & Kohlberg, 1987a; Colby & Kohlberg, 1987b), structural criteria for stage determination had become more apparent, but the scoring system itself continued to rely in concept-matching.

Kohlberg (1969; 1984) claimed that his moral stages met stage criteria he attributed to Piaget, including invariant sequence, hierarchical integration, *structure d'ensemble*, and universality. Since the development of his most recent scoring system (Colby & Kohlberg, 1987a; 1987b), the invariant sequence criterion has received substantial empirical support (e.g., Armon & Dawson, 1997; Colby et al., 1983; Dawson, 1996d; Nisan & Kohlberg, 1982; Snarey, Reimer & Kohlberg, 1985; Turiel et al., 1978; Walker, 1982; Walker, 1989; Walker & Taylor, 1991). Good empirical evidence for the concepts of hierarchical integration and *structure d'ensemble* are more difficult to obtain than evidence for invariant sequence, due to difficulties in operationalizing these constructs, but some supportive evidence has been collected (Turiel, 1977; Walker, 1979; Walker & Richards, 1979; Walker & Taylor, 1991). Claims of universality have received both positive (Nisan & Kohlberg, 1982; Snarey, 1985; Turiel et al., 1978) and negative (Gibbs, Arnold & Burkhart, 1984; Gilligan, 1982) support, depending on the particular features of Kohlberg's stages under investigation.

The good life

Armon's (1984) work focuses on the development of evaluative reasoning about the Good, including the good life, good work, the good person, and good friendship. The scoring systems for each of Armon's Good Life domains was developed independently. Her research approach differs from Kohlberg's in that she does not use dilemmas. Instead, she developed an interview technique that employs open-ended questions such as, "What is the good life?" and "Why is that good?" A

second distinction between Armon's and Kohlberg's work is that the development of her stage scoring system was guided by more formal structural criteria. Like Kohlberg, however, Armon was guided in her stage definitions by predetermined philosophical categories, and her stage scoring method relies, as does Kohlberg's, on content descriptions, though she is careful to point out that the conceptual content included in her stage scoring manual is not exhaustive.

Armon's stages are intended to conform to the same hard stage criteria as Kohlberg's. Her longitudinal study (Armon & Dawson, 1997) provided convincing empirical evidence of invariant sequence, and some evidence in support of hierarchical integration, but there have been no longitudinal replication studies that could provide additional support for these or the other hard-stage criteria. A description of Good Life and Moral stages can be found in Appendix 1.

The development of evaluative reasoning in adulthood

This research involves a life-span sample of participants. A central hypothesis is that development in adulthood involves the same processes as development in childhood. The cognitive-developmental approach is a relatively new but promising paradigm for the study of adult development. Research in this area began in the moral domain, and probably owes its origin to the disconfirmation of Kohlberg's (1969) hypothesis that post-conventional reasoning emerges in late adolescence. In the early 1980's, Colby and Kohlberg (1987a) rescored Kohlberg's longitudinal data, using their third and most reliable scoring system. They found no postconventional moral reasoning before mature adulthood—even in their sample

of well-educated males. Several studies of adult moral development have since been conducted using Kohlberg's dilemmas and scoring manual, all resulting in similar findings (Armon & Dawson, 1997; Commons et al., 1989; Pratt, Golding & Hunter, 1983; Walker, 1989).

To date, the body of work on positive adult development outside of the moral domain has been largely theoretical (e. g. Basseches, 1984; Fischer, Hand & Russel, 1984; Souvaine, Lahey & Kegan, 1990; Sternberg, 1983). However, there are a few instances of empirical research. For example, Armon (1984a; 1984b) found, in her longitudinal study of 42 children and adults, that higher stages of evaluative reasoning about the good were confined to mature adulthood, and Demetriou and Efklides (1985), based on their study of 400 adolescents and adults, concluded that post-formal abilities were more likely to be found in adulthood. Further, Sinnott (Sinnott, 1989) has distinguished forms of adult thought, as have Kitchener and King (1981).

Common's General Model of Hierarchical Complexity (Commons et al., in press) provides the most compelling evidence that development in adulthood involves changes in the structure of performance that are identical to changes that take place in childhood. Each of his 12 stages (described below) is strictly defined as the integration of the operations of the previous stage. Scoring with his system involves identifying the level of hierarchical integration present in performances. Commons systematic stage (which immediately follows his formal stage) has rarely been found before late adolescence or early adulthood, and and his metasystematic

stage has rarely been found before mature adulthood (Commons et al., 1989; Commons & Richards,). I report similar findings in Chapter 6.

Contextual versus universal features of development

The research design of this project was informed by a desire to distinguish between what Piaget calls *operations* or *global reasoning structures* (or what Commons calls *orders of hierarchical complexity)* and the elements that are operated on with these structures. In other words, stages are conceptualized in this project as orders of hierarchical complexity. Elements are propositions constructed within informational, contextual, and stage boundaries. Though the elements have their own structure, in the present context they are treated as content. This classification is referential. Concepts are content with respect to the larger structures in which they are embedded.

The possibility of distinguishing structure from content cannot be taken for granted. In fact, even the wisdom of making such an attempt has been challenged in the literature. The debate is characterized by at least three distinct positions: The first of these asserts that form and content are part of a dynamic whole that is irreducible (Colby & Kohlberg, 1987a; Vuyk, 1981a; Vuyk, 1981b). Universal processes may exist, but these cannot be studied independently of their content. A second position is that developmental differences in both structural and content features of behavior are an artifact of social transmission, not the result of internal processes (Bandura, 1986; Schweder, 1987). In this view, there is no basis for distinguishing between them, because they both emerge from the same source. A third position, and the

one I am taking here, is that structure and content, though they are part of a dynamic whole, have distinct features. Structures, at the level of operations, undergo a series of universally observable hierarchical integrations. It is the *content* of these operational structures that varies. Structure and content, looked at in this way, can be identified and isolated for the purpose of understanding both their distinct qualities and their interrelationship.

The stage scoring system developed by Commons and his colleagues was chosen for this project primarily because stage scoring in this system does not rely upon content descriptions. As noted above, in Colby and Kohlberg's (1987b) and Armon's (1984) stage scoring systems, scoring involves the identification of specific conceptualizations. In fact, most traditional stage assessment systems suffer to some extent from bias introduced by an over-dependence on content, because scoring is largely a process of matching concepts in a scoring manual. As Levine (Levine, 1979) points out, it is impossible to study content or context effects when these are conflated with what he calls general cognitive abilities. Several attempts have been made to overcome this problem (see, for example, Edelstein, Keller & Wahlen, 1984; Keasey, 1975; Keller & Reuss, 1984; Rosenberg, 1988; Selman & Damon, 1975; Selman, 1971b; Simpson, 1976; Stuart, 1967), but none of these involved a means of independently coding the hierarchical complexity and content of the same material.

The GSSS differs from traditional systems in that scoring is based upon the organization of conceptual elements in a scorable segment of text (ideally, a

"complete" argument). The elements are central to scoring, but only in terms of their hierarchical complexity and the hierarchical complexity of the statements in which they are embedded. They could be replaced with other elements of the same hierarchical complexity without altering the stage of the statement. (See chapter 4 for more on the GSSS.)

One advantage of this system is that propositions or other content can be examined with respect to the stages of the larger arguments in which they are embedded. This is possible because stage scoring criteria are independent of the specific conceptual elements contained in texts. Moreover, the content can be examined for patterns unrelated to stage, and stage-related effects and non-stage effects can be examined in conjunction.

A second advantage is that such a system does not, in itself, impose limitations on the range of conceptual content that can be coded. Content-laden scoring systems are frequently criticized because the range of conceptualizations they include is restricted (Coggins, 1978; Commons et al., 1984; Langford, 1995; Puka, 1991; Rosenberg, Ward & Chilton, 1988). Much of the work in stages of social development traces its origin to Kohlberg and his colleagues, whose contribution to our understanding of development in the social realm has been enormous. A common criticism of their early work, in particular, is that they defined the moral realm too narrowly. This problem is particularly evident in Kohlberg's (1969) early work, in which moral reasoning is defined predominantly as the realm of rights, obligations stemming from those rights, and issues of justice. The often intense

negative response to Kohlberg's work has primarily—admittedly in many guises—been a rejection of this narrow conception of the domain of moral reasoning.

For example, Gilligan (1982) argued that Kohlberg's emphasis on justice caused him to ignore or marginalize care considerations. One of the reasons for Kohlberg's neglect of moral categories like care was his approach to stage scoring. The parameters of his coding system were not only predetermined by philosophical categories, but the coding itself relied upon the identification of particular content at least as much as it relied upon the explicitly structural/organizational features of justifications.

Though conventional philosophical categories may provide a starting-point for the delineation of domains of reasoning, a *psychological* domain must ultimately be defined in terms of the reasoning of a wide range of persons across both developmental levels and cultures. Domain definition should be the result of an open-ended dialectical process with its beginnings in theory, rather than a formally articulated *a priori* delineation of the domain. By embedding a set of content descriptions in coding protocols, Kohlberg restricted the applicability of his scoring system outside of Western cultures, and unnecessarily limited the delineation of the moral domain.

A third advantage is that the GSSS, because it can be applied in any domain of reasoning, permits the comparison of the hierarchical complexity of reasoning across domains. Commons and his colleagues (Commons et al., 1989; Commons et al., 1983) suggest that one can legitimately compare the level of an individual's

reasoning across tasks only when scoring systems are based upon the same structural criteria, or, alternatively, when they tap the same dimension of reasoning. Even when stage sequences in different domains are based on similar conceptions of structure, comparing stage achievements across domains is not a simple task. In a study in which they compare reasoning across the domains of intellectual, moral, and ego development, King, Kitchener, Wood, and Davison (1989) point out that no method exists for determining just how stages in different domains should be related to one another.

A fourth advantage of a content-independent method of stage scoring is that such a system may provide a more accurate measure of stage than one that relies on the identification of particular content. There are three reasons for this hypothesis. First, with a content-independent system it is possible to score any performance—not just those that match performances in a coding manual—which means that more material from a given interview can be scored, offering more opportunities to assess stage. Second, because it is possible for low-stage individuals to "parrot" higher stage concepts, content-based scoring systems can sometimes give a higher stage score to a performance than it actually merits. It is far more difficult, arguably impossible, for an individual to consistently fake a higher order of hierarchical complexity than their own highest level. Third, it is quite possible for an individual performing at a higher stage of development to utilize concepts that are associated with a lower stage in a content based scoring system, which could lead to an underestimation of stage in some cases.

In a study comparing the performance of three stage-scoring systems, the GSSS, the Standard issue Scoring System (SISS) (Colby & Kohlberg, 1987a) and the Good Life Scoring System (GLSS) (Armon, 1984b) Dawson (1998) found not only that the three scoring systems appeared to tap the same dimension of performance, but scores for performances on the GSSS were more uniform than on the other instruments. Scores for individual GSSS performances, with only one exception, were either at a single stage or two adjacent stages—were either consolidated at a single stage or transitional—while scores for performances on the other two instruments often spanned three stages. It is notable that GSSS scoring was conducted on randomly sorted (alphabetized) segments of text by a blind rater. Moral judgment and good life scoring were not conducted blindly. Further comparison of the GSSS and the SISS and GLSM is underway.

One problem with the idea of content-independent stage scoring is that the description of a stage within a domain lacks meaning if it is bereft of conceptual content. Armon (1984b) discusses this problem:

> ...[a] stage represents both the formal processes by which the individual decides what is ideal or good, and the outcome of that process, which consists of what is thought to be ideal or good. Only the former is purely structural, however. That, is, only formal evaluative processes, independent of their outcomes, can withstand a rigorous structural analysis. For the structure to be more psychologically and empirically meaningful, the outcomes must then be connected to these processes (p.

152).

The implication is that stage can be determined without dependence on *specific* conceptual content, though stage definitions within a psychological domain must include descriptions of the content associated with a particular stage within that domain. For the present work, descriptions of reasoning at each stage are constructed on the basis of the conceptual content associated with particular stages of performance, integrated at the level of structural complexity characteristic of that stage. The stage descriptions constructed in this way are not intended to be used as guides for stage scoring, and their content should not be construed as comprehensive. In fact, because stage can be determined without reference to these stage descriptions, they can be viewed as open-ended, subject to alteration as additional empirical data becomes available.

According to Blasi (1987), "The real issue [in the structure/content debate] is one of integration; namely of understanding how concrete...judgments and abstract...structures go together both at the level of individual consciousness and in functionally affecting behavior" (p. 89). The advantage of the present approach is that it can help to bring about such an understanding. The differentiation of operational structure from conceptual content provides an opportunity to better understand their interrelationships.

This chapter lays out the cognitive-developmental background of the present research. By providing a sketch of relevant Piagetian constructs, and showing the

ways in which these constructs have been operationalized in previous research, I have highlighted both strengths and weaknesses of the cognitive-developmental paradigm. Its strengths include (1) a strong theoretical framework for research into the development of knowledge, (2) constructs that can be fruitfully operationalized to demonstrate how evaluative understandings change with development across the lifespan, and (3) methods of research like Kohlberg's "bootstrapping" procedure, that have resulted in rich insights into knowledge development. Its major weakness, in terms of the present project, is in most researchers' failure to adequately explore and elucidate the relationship between the content of knowledge and its structure. Instead, these are often confounded, imposing limitations on the range of research questions that can be meaningfully addressed and bringing some research results into question.

In the chapters that follow, I employ commonly used cognitive-developmental techniques, including (1) the probed clinical interview, (2) stage-scoring, and (3) an iterative research method similar in spirit to Kohlberg's "bootstrapping" process, in concert with more innovative methods such as (4) independent analyses of hierarchical complexity and propositional content, (5) Rasch modeling, and (6) a new method for constructing descriptions of the meanings available at different stages of development, to examine a long list of hypotheses.

Hypotheses

The first set of hypotheses involve developmental stages of evaluative reasoning about education. They are (1) that patterns of stage performance support the

notions of *structure d'ensemble* and invariant sequence; (2) that development within individuals is both continuous and discontinuous; (3) that the systematic and metasystematic stages of the GMHC are present predominantly in adult performances; (4) that educational attainment and stage have a strong linear relationship across the lifespan; (5) that gender is not significantly related to stage after educational attainment and age are taken into account.

The second set of hypotheses involve the relationship between the conceptual content of evaluative reasoning about education and the hierarchical complexity of performances. They are (6) that the presence of particular concepts does not, in and of itself, predict the stage of performance of any individual, because many concepts are explicitly present at multiple stages and because performance can reflect more than a single conceptual domain; (7) that fundamental concepts with roots identifiable early in development can be traced as threads of meaning (metaphorical strands) that traverse several stages of development; (8) that these can be shown to represent unidimensional constructs, both quantitatively and interpretively—in the sense that all of the meanings present in the concepts available at lower stages are implicated in the meanings constructed at higher stages; (9) that the meanings constructed at each stage—along individual metaphorical strands—can be understood and described in terms of the organizing principles of each stage; and (10) that the *descriptions* constructed in this way are conceptually analogous to the stage *definitions* of Armon and Kohlberg.

Chapter 4

Participants, Data Collection, and Coding

A good teacher is fun and is good at learning...I guess it's the same things that make a good student.

—Student, male, age 11

Participants

Participants in this study represent a convenience sample. They range in age from 5 to 86 years of age, and are predominantly Caucasian, well-educated, and from the middle class. They were recruited through word-of-mouth and through advertisements placed in a newspaper and on an e-mail listserve. In addition to long clinical interviews, demographic information was collected, including age, educational attainment, and sex.

Good education interviews

A total of 153 good education interviews were conducted. Ninety of these interviews were collected by Armon for her longitudinal, life-span study of evaluative reasoning about the good (Armon, 1984a; Armon, 1993; Armon & Dawson, 1997; Armon & Dawson, in prep). Thirty-three participants from Armon's study were administered the Good Education interview up to three times at four-year intervals between 1980 and 1989. A total of 90 Good Education interviews were administered to these 33 participants. (An analysis of the longitudinal results is not reported here.) For the purpose of this analysis, each of these interviews is treated as a separate case, to give a total of 90 (permissable when test-times are

separated by several years, Willet, 1989). I collected the data for the remaining 63 cases. All interviews were recorded on audio tape and transcribed verbatim by professional transcribers.

Data collection was conducted as follows: After demographic information was collected and the respondent appeared to be at ease, the Good Education Interview was administered. It includes the following questions:

1. What is a good education? or What is your concept (idea) of a good education? or What makes an education good? Why is that good (important)?
2. What are the aims (goals, purposes) of a good education? Why are those good (important)?

The clinical interview method, as adapted by Armon (1984) from an approach employed by Piaget (1929) and Kohlberg (1969), was employed. Questions and probes are chosen to encourage participants to expand upon their conceptions of good education and elicit their highest level of reasoning. Responses are probed with requests for further elaboration— "Why is that good?" "Why is that important?" "Why should education include both of those things?"—until the interviewer is satisfied that a given participant has presented as full an account as possible of his reasoning on each question. The interviewer does not introduce concepts of her own unless the subject is unable to respond to her initial questions. Instead, she notes the elements of good education that are mentioned by the participant and probes for explanations of why these are important. Interviews vary in length from 20 minutes to over an hour in

duration. I have been trained in this method by Armon and have used it for several projects.

The following example is offered as an illustration of the interview process. Because of space limitations, I have selected a portion of the interview of an unusually reticent 18-year-old college freshman. Interviewer questions are in italics.

> *The first question is what is your concept of a good education?*
>
> What do you mean by concept?
>
> *Your idea of a good education?*
>
> To...basically learn what you want to learn, and to learn..., get a job, maybe when you graduate. I guess that's it.
>
> *Why is it important in a good education that you can learn what you want to learn?*
>
> 'Cause if you're just learning things that you're going to get a job with, there's no interest. There's nothing interesting about that. I think a good education should..besides...it should give you the knowledge that you find interesting. So...
>
> *And why is it important that it give you a kind of knowledge that you find personally interesting?*
>
> I guess if you find something interesting, when you find something interesting, you learn it better than if you're just told to learn it.
>
> *Okay, and why is it important also that a good education prepare you to get a good job?*
>
> Well, I think that's what most people get an education for. Right? I guess when you get an education, you're differentiating yourself from other people who are not, so a good education should provide you the knowledge you need to get a job that, you know, other people can't get without an education, perhaps.

Is there anything else that is important in a good education?

I can't think of anything right now.

As this example illustrates, the probe questions refer to claims made by the respondent. Her responses to these questions elaborate on her original response and extend it. It is not unusual for additional elements of a good education to emerge in these responses. When they occur, these are also probed.

Stage scoring

Stage scoring for this project was conducted with Commons' General Stage Scoring System (GSSS) (Commons et al., 1984; Commons et al., 1995). Sonnert and Commons (1994, p. 40) offer the following summary explanation of the General Stage Model (recently renamed *The General Model of Hierarchical Complexity, or GMHC*):

> At the heart of the [General Stage] model is the notion of hierarchical (vertical) complexity—as opposed to non-hierarchical (horizontal) complexity. Hierarchical complexity is the order of logic required to complete a task (Commons et al., 1984; Commons & Rodriguez, 1990)...The successful completion of a task requires an action of some order of hierarchical complexity. At each order of hierarchical complexity, actions, including reasoning, are defined in terms of the actions of the previous order. Actions at a higher order of hierarchical complexity organize and transform lower-order actions. We say that higher-order actions coordinate the actions of the next lower order. This organization of lower-order actions is new and unique and cannot be accomplished by those lower-order actions alone.

The GSSS has been employed in several studies (Commons et al., 1991a; Commons, Meaney & Weaver, 1989; Commons & Richards, 1984; Commons,

Sonnert, Gutheil & Bursztajn, 1991b; Gutheil, Commons, Rubin, Morley & Applebaum, in preparation; and Rodriguez, 1992; Sau-Ching Lam, 1995).

Commons' conception of development differs from those of Piaget, Kohlberg (1969), Fischer (1984), and Pascual-Leone (1979), among others, in its focus on behavior rather than the mechanisms and mental constructs that underlie behavior. Most developmental theories, including Piaget's, are mentalistic, in the sense that mental structures analogous to the operations performed at each stage of development are the subject of both research and theory. For Piaget (Vuyk, 1981), the elements forming the content of these mental structures are perceptions, memories, concepts, and operations. In a similar fashion, perceptions, memories, and concepts are the content of operations. In Piaget's system, the nature of mental structures is inferred from operations, which are observable in behavior. Commons' makes no such inferences. His work examines only the level of structure exhibited in performance. Both Piaget and Commons, however, have only interpreted texts or behaviors to work from and to offer as evidence in support of their theories.

Despite this difference between Commons' approach and those of other structural developmentalists, the General Stage Scoring System (GSSS) and other developmental scoring systems appear to measure the same dimension of performance. Evidence for this has been reported by Sau-Ching Lam (1995) who found no significant difference between stage scores generated by Armon's (1984) Good Life Scoring System and the GSSS, and Dawson (1998) who found correlations from .90 to .97 among stage (ability) estimates generated from data

scored with the SISS (Colby & Kohlberg, 1987b), GLSS (Armon, 1984), and GSSS, even when they were applied to different interview material and scored by different raters. Moreover, the results of the analysis of GMHC scores presented in Chapter 6 are remarkably consistent with those reported in research involving several instruments designed in accord with Piagetian principles (Dawson, 1997; Dawson, 1996b; Draney, 1996; Goodheart, 1996a; Goodheart, 1996b; Müller, Reene & Overton, 1994; Müller, Winn & Overton, 1995; Noelting, Coudé & Rousseau, 1995; Noelting, Rousseau & Coudé, 1994; Wilson, 1989c).

The GMHC proposes 14 stages of development. These include (1) sensory and motor actions, (2) circular sensory-motor actions, (3) sensory-motor, (4) nominal, (5) sentential, (6) pre-operational, (7) primary, (8) concrete, (9) abstract, (10) formal, (11) systematic, (12) metasystematic, (13) paradigmatic, and (14) cross-paradigmatic. The present project is concerned with stages 7 through 12. These stages correspond hypothetically to Moral Judgment and Good Life stages (Commons et al., 1995a; Dawson, 1998) as shown in Table 2.

Table 2

A Comparison of Three Developmental Numbering Sequences

	GMHC Stage	Kohlbergian stage	Good Life Stage
Primary	7		
Concrete	8	2	2
Abstract	9	3	3
Formal	10	3.5	
Systematic	11	4	4
Metasystematic	12	5	5

To reduce confusion, GMHC stages will be referred to by name, rather than by their numbers.

Stage Descriptions: At the primary order of the GMHC, individuals can accurately order sequenced events that are connected solidly to concrete reality with some specified time, place, act, or actor (statements in quotes are 'ideal' responses, composed from multiple actual examples): "I go to school and play at recess," or "I like spelling. I have a spelling test tomorrow." They can take the perspective of either self or other, but cannot coordinate perspectives: "I like to play," or "Billy likes to play."

At the concrete stage, respondents can report about things, incidents, events, actors, actions, and places in the context of an interaction between self and other: "The teacher yells when we act up in class." They can discriminate how their own actions affect others' behavior. One's own sequence of causes, resulting behavior, and outcome is related to another's sequence, and vice versa: "If I talk without raising my hand the teacher gets mad and yells at me." Concrete reasoners can list several external features and report rudimentary internal states such as feeling happy or feeling sad, generally with repect to some specified time, place, act, or actor: "The teacher was nice when I did my work and mean when I talked in class," or "I get scared when the teacher yells at me, " or "Spelling is fun because I like it."

At the abstract order individuals operate on abstract variables such as times, places, acts, actors, states, types. These variables have a stereotypcal quality and are

subject to logical quantification with the use of words like *all, none, some, never, sometimes, always, no one, somebody,* and *everyone*: "Teachers always get mad when kids fool around in class," or "Some teachers are mean, but mostly, they're nice," "Kids who like to go to school get good grades," or "Everyone should go to college so they can get a good job." Clusters of stereotypical attributes and traits can be used in descriptions: "Mean people yell at you and don't play with you and get mad for no reason." There is beginning awareness of an inner self, composed of feeling states and personal preferences: "I'm the kind of person who likes to go to school." Individuals can construct a third-person or "neutral" other by generalizing cause-and-effect chains of two individuals' behaviors. The neutral other determines which side in a conflict is correct by selecting the outcome that represents the norm or mode: "You ought to go to school, even if you don't like it, 'cause everybody wants to get a good job." This ability to generalize experience allows for the narrative representation of local group norms and archetypical characters embodying stereotypes: "A good student studies hard, listens to the teacher, likes her classes, and gets good grades." Abstract stage third person perspective-taking leads to the formulation of aphorisms: "The more you learn, the smarter you'll be."

At the formal order, respondents employ abstract variables connected with "if... then" inferences and other kinds of logical relations, with single variables on the input side. Reasoning is linear, logical, one dimensional: "If you talk to your teacher like a person, she'll get to know you, and that will make her more interested in you. Teachers who are interested in you are more fun to learn from."

Arguments are supported with empirical or logical evidence: “Getting to know my students helps me figure out what they need to learn next. Last year, when I was substitute teaching, I had no idea what the students knew and it made it very hard to figure out what to teach them.” The self, now a personality, is composed of clusters of traits or societal roles: “I try to be a good student and ask lots of questions, but I’m shy and a little afraid of being embarrassed, and that makes it hard.” Behavior is seen as rule-governed (i.e., governed by univariate causality) and its rules can be tested by empirical evidence. Rules are followed because of their empirically proven benefit. “Nine times out of ten, you’ll fail if you don’t study.” “Students want to learn if the subject is made interesting.” Simple rules can be strung together. “If the teacher makes the subject interesting, then you’ll feel like studying. You’ll learn more.”

At the systematic order, systems words appear frequently in utterances, words like *legal system, society, our company, the economy, the country.* Arguments are constructed with systems of relations—relations among relationships among variables. The individual constructs multivariate systems and matrices, coordinating more than one variable as input: “Students have different learning styles and interests. One aspect of a good teacher’s role is to learn as much as she can about her students so she can individualize instruction enough to engage all of her students.” Events and ideas can be situated in a larger context: “A good education isn’t just about learning for personal success or for your own personal family, it’s about what each of us contributes to society.” The self is composed of a system of

sub-personalities. “Though my core self seems the same no matter what I’m doing, in some ways I’m a different person as a teacher than I am as a student. As a student, I’m more in touch with my own ambitions and interests. As a teacher, I’m more in touch with the kind of world I want to help create.” At the systematic order, the behavioral framework of the other is viewed as an integrated system of tendencies and relationships: “Students aren’t bad or good. Every student has a combination of strengths and weaknesses. We need to bolster their strengths and help them overcome their weaknesses in order to prepare them for life in soceity.” They also coordinate linear causality with hierarchical organization, and order different perspectives in a hierarchy of preference. “When I think about a good education, I ask myself, ‘from what point-of-view—the student, the teacher, the society?’ I think you really have to look at it from the societal perspective, because that encompasses the interests of everyone.”

At the metasystematic order, metalogical, meta-analytic words appear. Properties of systems can be considered: *homomorphic, isomorphic, incomplete, inconsistent system, consistent system, incomplete system, complete system, incommensurable system, orders of complexity of the system, commensurable systems, etc*. The comparison of systems and perspectives occurs in a systematic way (i.e., across multiple domains): “A good educational system is based both on the needs of individuals within the society and on the needs of the society itself. Fortunately, our personal interests and shared interests are commensurate—especially when we take a developmental view.” The metasystematic individual reflects on systems,

constructing supersystems of systems. Treating systems of vertical and horizontal causal relations as the objects allow systems to be compared and contrasted in terms of their properties: “The self system and the social system are not closed to one another. They are interdependent. The existence and well-being of one is dependent on the existence and well-being of the other. When we fully grasp this, then we can go about determining what kind of educational system we want to put into place.” The focus is on the similarities and differences in each system’s form, as well as constituent causal relations and actors within them: “This is not to say that the individual will always perceive his interests to be commensurate with the interests of the system. The desire, for example, to accumulate personal wealth may be incompatible with a social need to distribute wealth equitably. Ultimately, however, the inequitable distribution of opportunity that results from the unbridled accumulation of personal wealth by a few individuals leads to a decrease in the quality of life at every level of society. We see this in increased crime, deterioration of services, and degradation of the quality of the workforce.”

Scoring with the GSSS: To score stage with the GSSS, one must first identify a scorable statement. A scorable statement is defined as a segment of text that contains a complete argument, or a justification for including an element or group of elements in the definition of a good education. Here, for example, is an extract from one of the interviews conducted for the present project:

A. The ideal education would be to have a school system, okay, first of all, that you have teachers that are aware of human needs, and have a balanced perspective. Because nowadays we have one type of school where all they care about are the emotional needs and social interaction,

> but they have no intellectual growth for the kids. And you have the others that will make them into little computers. So what you have to do is realize that both of those have some equal weight in maximizing potential in human beings. I think where our problem lies, in devising the system, is that we're in the age of specialization, which I'm completely appalled by. Because I feel it makes all these lopsided people in society. So that we have to start reeducating the educators about the uniqueness of the human being and his potentiality and the fragileness of it. That he has—not only is he an animal, emotional but he is also an intellectual, he has intellectual potential. And these things are intertwined so he can seek his happiness. So we have to have educators that are more interdisciplinary, that are more aware of all the different repercussions, the factors that influence the human being.

The general argument stated here is that a good education promotes both the emotional and the intellectual development of students, and that good schools systems and teachers should promote this development. Several justifications (or arguments) are offered:

f1. Teachers would be aware of human needs.
f2. Teachers would have a balanced perspective.
g1. Teachers would not emphasize emotional needs over intellectual needs.
g2. Teachers do not emphasize intellectual needs over emotional needs.
h. Intellectual and emotional growth have equal weight in maximizing potential.
j. It appalls me that people are so specialized.
k. Specialization makes lopsided people.
l1. Educators should be reeducated about the uniqueness of human beings.
l2. Educators should be reeducated about the potentiality of human beings.
l3. Educators should be reeducated about the fragility of human beings.
m1. Human beings have an animal part.
m2. Human beings have an emotional part.

m3. Human beings have an intellectual part.

n. The emotional and intellectual parts are intertwined.

p. Ability to seek happiness results from [appropriate?] intertwining of parts of self.

q1. Educators should be more interdisciplinary.

q2. Educators should be more aware of the different repercussions.

q3. Educators should be more aware of the factors that influence the human being.

There are several examples of formal propositions in the above list. They are considered minimally formal because they either (1) include formal concepts in the form of phrases or words, or because they (2) coordinate pairs of abstract variables. An example of (1) is seen in *m1, m2,* and *m3,* where the respondent employs the term *human beings.* This minimally represents a formal concept, in that it is a broad abstract category that is made up abstract traits rather than an accumulation of persons. An example of (2) occurs in *k,* where the variable *specialization,* is coordinated causally with the variable *lopsided people.*

In addition to formal propositions, several systematic propositions are present in the text sample. Some of these are concepts such as *factors that influence the human being*. This is a system because one can infer from the types of factors listed in the text as a whole that these *factors* are, at the least, a coordinated set of formal concepts such as the *potentiality*, *fragility*, and *uniqueness* of human beings, and their *intellectual* and *emotional* needs. Other systems present in the text are constructions containing several concepts that are related to one another through causal or other logical links. The following four systems are formed in this way:

System A = proposition (f1 + f2) where proposition (g1 + g2) because

proposition h.

System B = proposition j because proposition k, \ proposition (l1 + l2 + l3)
System C = proposition (m1,m2 = m3) where proposition n
System D = proposition p, \ proposition (q1 + q2 + q3)

The final System, *D*, is actually a metasystem, because it coordinates System A with Systems B and C. Educators, in a good education, must balance the intellectual and emotional needs of students, and, in order to do this, they must understand their students' potentiality, fragility, and uniqueness. In other words, they must coordinate these two knowledge systems. In making his case, this participant has demonstrated his ability to think metasystematically by advocating that educators should reason metasystematically. Based on this analysis, the stage score for this participant's performance would be stage 5, or metasystematic.

For the present analysis, all 153 interviews were divided into scorable segments *(statements)*. Because the interviews were open-ended, there was no predetermined content-guided basis for segmentation. Consequently, the following criteria, similar to those outlined by Stinson, Milbrath, and Reidbord (1993) were employed:

1. A scorable segment, should, as much as is possible, represent a complete argument for a given proposition or related set of propositions, including all of the "why" probes and responses associated with that argument.

2. When two or more arguments are intertwined in the same text, the text is left intact and scored only once; and

3. Arguments must include responses to "why" probes or spontaneous justifications, because these, much more than the propositions

themselves, reveal the structure of participants' thinking. When these are not present, the argument is not scorable, and is dropped from the analysis.

Retest data were collected for 11 participants in order to examine the stability of performances on the good education measure. Two weeks following their original interviews, 11 participants were reinterviewed, employing the same procedures as in their original interviews. All statements, including those from the retest, were sorted alphabetically and scored by a single blind rater who was personally trained by Commons. The correlation of the measures (ability estimates from a partial credit analysis) from the first and second interviews of these participants was .96 (not disattenuated for error).

Good education content analysis: Proposition coding

The goal of the content analysis is to identify all of the distinct propositions about good education made by each participant, and to classify these according to the stage of the *statements* in which they are embedded, so that the relationship between conceptual content and stage can be explored.

Identification and classification of propositions are guided by several criteria. First, the concept must apply to the subject of good education. The following criteria are used to determine if this is the case:

a. From the point-of-view of the participant, this proposition is a feature, result, or element of a good education;

b. From the point-of-view of the participant, this proposition is a feature,

result, or element of a bad education.

Second, the proposition must have an irreducible meaning. A series of propositions, strung together without a claim that they are part of a single, larger proposition are treated as separate, while a proposition that represents a non-additive synthesis of less complex elements, such that it has a meaning distinct from its elements, is treated as a proposition in its own right. The elements of this larger proposition, if explicitly listed by the participant, are also coded as separate propositions. Including all elements in this way ensures that no particular level of conceptualization is privileged in the content analysis.

Third, the proposition is assigned a proposition code. The code may be one of a list of existing codes, or it may be created for the current proposition. Criteria used here are:

a. A proposition should not be assigned an existing code if doing so would require ignoring some aspect of its meaning.

b. A proposition should not be assigned an existing code if it lacks any aspect of the meaning present among the other propositions assigned to that code.

Finally, the proposition is linked, in the database, with the statement in which it is embedded as well as the stage of that statement. The stage and content analyses yielded 739 stage-scored statements and 2424 propositions. Each proposition was assigned one of 644 proposition codes. These were numerous as a consequence of the decision to preserve subtle differences in meaning. The number of propositions assigned to each of the proposition codes ranged from 1 to more than 40. Table 3 provides examples of proposition coding.

The qualitative reintegration of structure and content

The goal of the qualitative analysis is to describe evaluative reasoning about education at each stage of development. To accomplish this end, the propositions identified in the content analysis are reintegrated into descriptions of reasoning at each stage, employing the organizational principles of the stages at which they are found in performances. In a sense, this is the inverse of the process employed in stage-scoring.

For example, at the abstract stage the following social benefits of education were identified in the data:

1. You could help people out with problems;
2. You're interesting to other people;
3. You can make a lot of friends;
4. It helps you work together;
5. You understand each other more.

At the GMHC abstract stage, the conceptual elements of statements are organized into abstract propositions, which consist of two or more generalized concepts (such as *being interesting, having lots of friends, working together, understanding one another*) that are associated with one another additively or sequentially. These propositions take the form of assertions, and can be accompanied by supporting examples. Given the concepts listed above, here are some constructions that would be possible at the abstract stage:

1. A good education helps you work together. You understand each other more.
2. In a good education you could learn to help out people with problems.

3. A good education makes you interesting to other people. You can make a lot of friends.

In these constructions, though some kind of relationship between the elements present in propositions seems to be implied, this relationship is not explicated. At the abstract stage, individuals do not construct *explicitly* logical antecedents and consequents. The following construction is formal operational rather than abstract.

1. In a good education you learn to work together. This teaches you social skills you need in order to understand each other. *[If children work together, then they will learn social skills, and if they learn these, then they will come to understand one another.]*

Here are some examples of abstract reasoning from actual interviews.

1. *Why is it important that a good education makes you smart?*
   People like smart people. You can make a lot of friends.

2. *Why is it important to work together in a good education?*
   So that you get along with the other people, and that you do a whole bunch of stuff together.
   What kinds of things are you talking about doing together with the other people?
   Writing out a project, or doing a garbage thing where you get a big garbage bag and you pick up some garbage that's on your street.
   Why are those kinds of things important to a good education?
   It helps you work together and you understand each other more.

As these examples illustrate, articulation of the social benefits of education at the abstract stage appear, in the present sample, to be limited to rather immediate benefits of socialization, being able to work together, having friends, and understanding others. These are presented as loosely linked sets of elements that are not yet organized through explicitly logical relations.

In Chapter 7, this process is used to trace the development of reasoning in a more detailed examination of two metaphorical strands, *good education as play, and good education as preparation for the good life.* Variations on the theme of *education as play* were spontaneously and explicitly generated in 70% of the interviews and were identifiable in some form at every order of hierarchical complexity identified in the data. Its prevalence, as well as its relevance to the spirit of this research made it an appealing choice for in-depth analysis. Variations on the theme of *education as preparation for the good life* were present in 91% of the interviews, and were identifiable in some form from the abstract to the metasystematic orders.

Analysis of the hierarchical complexity of performances

The stage scoring process produced a total of 739 scored statements. The data fields for the quantitative analysis were set up so that each participant had 7 opportunities to be scored for GMHC stage. Because interviews were of different lengths, some interviews included enough scorable statements to fill all 7 columns. Others did not. When participants had more than 7 scored statements, which occurred in a few cases, scores were randomly deleted with an SPSS subroutine until the maximum allowable number were left. When participants had fewer than 7 responses, columns with no response were coded as missing. Following this procedure, the data were set up as follows:

```
3433434
2122
444444
```

```
54455
011
```

Though moderate amounts of missing data do not seriously compromise the results of Rasch analyses (See Chapter 5 for more about Rasch analysis), the fact that missing data were clustered at the end of the data strings poses a problem. In positions 6 and 7 in the data string, there were large amounts of missing data. A great deal of missing data can result in unreliable estimates. To correct this problem, an experimental procedure was devised. First, the data were examined to determine whether there were any systematic increases or decreases in stage scores as interviews progressed. None were found. Then, scores for each participant were randomly distributed across all 7 positions in the data string by employing an SPSS subroutine. This was done to create data analogous to traditional items in which, barring systematic effects, missing data is likely to be distributed at random. The resulting data were set up as follows:

```
3434343
21 2  2
444 444
4 55 44
 1 0 1
```

The “items” in this analysis are, perhaps, better thought of as bundles of responses than as independent items. The procedure described above is exploratory and experimental, designed so that a maximum amount of the data from the stage scoring process can be included in analyses. Basing estimates on most or all of the data available for each case permits sensitive estimates of relative participant ability (stage), and of the relationships between stages.

Chapter 5

Data Management and Analytic Tools

We think of education principally as intellectual pursuits, or going to school, and that's not the sum total of it. Education involves a great deal more than just going to school. You can go to school all your life and never really learn anything, or how to adapt those learnings to positive social input.

—Male, age 62

Interview data, demographic information, and the results of the stage and concept analyses were entered into an interactive database.

Sequencer

Because of the very large number of proposition codes generated in the content analysis, it was necessary to develop an efficient method for examining their distribution in the sample. To accomplish this end, I designed a means for creating figures that permit preliminary visual inspections of proposition code distributions, and developed a software application, *Sequencer* (Dawson & Wilson, 1997), to produce these figures, called CSQs (concept sequences). They permit visual identification of both developmental and non-developmental features of proposition use. For example, from a cursory examination of the CSQ in Figure 1, one can immediately ascertain that some propositions first appear at the earliest stages and either drop out or continue to be used at almost every stage, while other propositions are used only at the higher stages. Second, CSQ's can aid in reducing the number of proposition codes to a more manageable number. With CSQs, one

can quickly determine whether proposition codes with apparently similar meanings occur within the same age, stage, or educational range, facilitating the dialectical process of determining whether the meanings expressed across developmental stages, ages, or educational levels are similar or distinct. This helps to prevent the loss of valuable developmental information, which would result if one inadvertently collapsed proposition codes across developmental levels. Third, CSQs can be used to compare how propositions are distributed with respect to any variable with two or more categories of response, and finally, CSQs can be used to how propositions are distributed within individual performances.

In the present set of analyses, CSQs are used primarily to examine the distribution of propositions across stages, and to aid in the process of tracing the development of one commonly employed metaphor, *education as play.*

Rasch analysis

The partial credit model, a member of the Rasch family, is used to model the stage performances of participants in this sample. Though well-known in psychometric circles, Rasch's models for measurement have been employed by developmentalists only recently (Andrich, 1986; Andrich & Styles, 1994; Bond & Bunting, 1995; Bond, 1994; Dawson, 1996d; Demetriou, Efklides, Papadaki, Papantoniou & Economou, 1993; Draney, 1996; Goodheart, 1996a; Goodheart, 1996b; Hautamäki, 1989; Kelley & Schumacher, 1984; Müller et al., 1995; Noelting et al., 1995; Noelting et al., 1994; Wilson, 1984; Wilson, 1985; Wilson, 1989a; Wilson, 1989b; Wilson & Draney, 1997). One area of

application for these models is the examination of behavior on measures intended to capture hierarchies of difficulty, which makes them highly suitable for developmental applications. Rasch's models test the extent to which data meet the requirement that performances and items (or levels of items) form a hierarchical sequence (in a probabilistic sense) along a single continuum (Andrich, 1989; Fisher, 1994; Masters, 1988; Wright & Linacre, 1989).

Rasch analysis is also useful for examining stage transitions. In their raw ordinal form, little can be said about the amount of difficulty associated with transitions between stage scores. When participants are ordered by the likelihood that they will perform at a given stage, the persons whose raw scores are high will be closer to the top of the developmental continuum, and the persons whose raw scores are lower will be closer to the bottom of the continuum. Rasch's models convert these likelihoods into quantitative estimates of item difficulty and person ability, expressed in the same equal-interval metric. The common metric along which both stage difficulty and respondent measures (ability estimates) are arranged is referred to as a logit scale, in reference to the log-odds unit employed (Ludlow, 1985; Wright & Masters, 1982; Wright & Stone, 1979). When two stages differ in difficulty, it is less likely that a participant with a measure near their average will score at the higher stage and more likely that he or she will score at the lower stage.

Much can be learned about stage transitions through the analysis of these probabilities. An ability estimate for a given individual means that the probability of that individual performing accurately on an item at the same level is 50%. There is a 73% probability that the same individual will perform

accurately on an item whose difficulty estimate is one logit easier, an 88% probability that he or she will perform accurately on an item whose difficulty estimate is two logits easier, and a 95% probability that he or she will perform accurately on an item whose difficulty estimate is three logits easier. The same relationships apply, only in reverse, for items that are one, two, and three logits harder. (For more on Rasch's models, see Andrich, 1988; Fisher, 1992; Fisher, 1994; Masters, 1997; Masters, 1982).

The logit estimates of item difficulty and person ability are but one of the three statistics essential to measurement in any context. Reliability and validity assessments require (1) that item and case estimates be associated with an error term, which makes it possible to establish confidence intervals for all item and person estimates, and (2) model fit statistics so that data from both items and persons can be examined for their conformity with the requirements of measurement. Data from individual items and persons or from sub-groups of items and persons can be examined in this way. As demonstrated in Chapter 6, the results of the analysis provide a great deal of useful information about patterns of behavior, and can assist in bringing a construct (latent variable) like developmental stage into focus.

Chapter 6

The Hierarchical Complexity of Performances

The aim of education should be to teach us rather how to think, than what to think
—rather to improve our minds, so as to enable us to think for ourselves,
than to load the memory with the thoughts of other men.

—Bill Beattie

Rasch analysis of the hierarchical complexity of performances

The partial credit model (Masters, 1982; 1994), an extension of the original Rasch model, designed to permit the estimation of multiple levels unique to each item in an instrument, was employed in this analysis. The partial credit model is appropriate here because each of the items in the analysis has 5 possible levels, concrete, abstract, formal, systematic, or metasystematic. One hundred and fifty-three cases were included in the analysis, which was conducted with Bigsteps software (Linacre & Wright, 1997). The data were set up as described in Chapter 4.

Case analysis I

The proportion of the observed case estimate variance considered to be true (reliability of case estimates) was .86. (See Wright & Masters, 1982) for more information about reliability of case estimates.) Given the small sample size (n = 153) and the number of levels (5) and items (7), .86 is remarkably good. Reliabilities for case estimates of .85 and up are considered acceptable in ability testing (Wilson, January 1997, personal communication). All items have acceptable infit (weighted) and outfit (unweighted) statistics (Wright & Masters, 1982),

indicating that they all belong on the same scale, or, in other words, that they measure the same dimension of ability. Infit and outfit statistics are expressed as mean squares and *z* scores. Mean squares are expected to be close to 1.0, and *z* scores are expected to be close to zero with standard deviations of 2.0 or less.

Figure 2 shows the stage (item) estimates and measures (case estimates or ability estimates, in this case, measures of the overall stage performance of respondents). Stage estimates are displayed on the left of the figure, and measures are displayed on the right (by case number). Measures are presented in three different type styles, to highlight different types of performances. The heavy black case numbers represent persons who performed at a single stage, the heavy gray case numbers represent people who performed at 3 adjacent stages, and the light black case numbers represent people who performed at two adjacent stages. Only 7 out of 158 (< .05%) of performances were at more than two stages, indicating that this pattern of performance is unusual and may be due to measurement error. However, the fact that 6 of these performances are clustered within a 10-point range is curious. These performances may have a significant impact on the outcome of the analysis. For this reason, I conducted an investigation into the performances associated with these cases, shown in Table 4.

Table 4

Six Performances that Span 3 Stages

Case #	Statement Scores						
115		**abstract**	system	formal	system	formal	system
020	**abstract**		formal	system	formal	system	formal
005			**abstract**			system	
114	formal	formal		**abstract**	formal	system	
148	formal	formal	**system**	formal	formal	abstract	abstract
132	formal	formal	formal	**system**	abstract	abstract	

A re-examination of the interview data revealed that the systematic score in 148 is the result of a typographic error. In the database, the statement associated with this score is coded formal. Cases 115, 020, 148, and 132 each have a single performance that appears out of place. These are highlighted in bold type. For cases 005 and 114 it was necessary to return to the interviews to determine which, if any, of the scores were questionable. In each case, it was the statement coded as abstract. All 5 of the questionably coded statements are shown in Table 5.

Table 5

Statements Receiving Unexpected Scores for 4 Performances that Spanned 3 Stages

005 He's had a lot of college. It also can be-someone that's real well read, that's read a lot of books and is informative. Somebody that has a doctorate degree...

115 RIGHT. THAT MAKES SENSE. LET ME THINK. NOW, IS THERE

ANYTHING ELSE? I THINK WE'VE DONE REALLY WELL HERE, ACTUALLY. OH. WHAT ABOUT THE ENVIRONMENT? WHAT'S A GOOD ENVIRONMENT FOR THAT GOOD EDUCATION?

One thing I noticed is that classes that aren't so huge like the ones I went to. It should be a much smaller classroom where it's easier to interact with the students and teachers.

WHY WOULD THAT BE IMPORTANT?

It makes it more comfortable for the students to get involved than if there's 500 students. It's very hard to sit there and yell out your questions to this tiny little dot on the stage. And the teacher just spends more time with you.

020 BUT IS IT IMPORTANT TO HAVE AN EDUCATION WHEN YOU GROW UP?

Yeah.

WHY?

Well, if you're going to take a job like a veterinarian, yes. There's no other way you're going to learn what it's about.

DO YOU THINK THE MOST IMPORTANT REASON TO GET AN EDUCATION IS TO GET A GOOD JOB?

Yeah, teaching also gets you ready for life ahead.

IN WHAT WAY?

Let's say, the driver's ed—that will get you ready for driving. And that would be when you're eighteen, nineteen. Oh, I have, I have play production. And if I got interested in acting, that would teach me how to do it, and teach me how to use my skills so I could do it when I get older.

114 WHAT KIND OF AN ENVIRONMENT IS A GOOD EDUCATIONAL ENVIRONMENT?

Smaller group. I always liked best when I knew the people in my class quite well.

THAT'S INTERESTING, JUST KNOWING PEOPLE IS IMPORTANT, FEELING COMFORTABLE. IF YOU'RE NOT COMFORTABLE HOW

CAN YOU ASK QUESTIONS?

When I was in 7th or 8th grade I would sometimes, even when the teacher, the last teacher was really funny, he was great, and sometimes when he would try to explain something I understood it, but the other kids didn't. And I sort of said, well why don't I explain it, and I got up and I did it. He was great. I loved him.

132 CAN YOU TELL ME ABOUT WHAT YOU THINK MAKES A GOOD TEACHER?

A good teacher wouldn't speed through the things. You have about 4 different things in the same class for homework, say you have to go back and do something on India and then to Egypt, they don't pile on the homework or give you so much, they gradually give it. They put more on to your homework load gradually so then you get used to it.

WHAT IS YOUR IDEA OF A GOOD SCHOOL?

They have a good system going. I'll give an example of my school. We have families and a homework hot line, everything's all organized. You know what you have to do, you won't get behind, you can keep your grades up. A good school that has good organization is pretty important. They pick the right teachers, not teachers that don't pay attention to their jobs.

WHY IS IT IMPORTANT THAT TEACHERS PAY ATTENTION TO THEIR JOBS?

If they don't, they're probably not going to teach the kids anything, and if the kids don't learn anything they won't get good grades for that class.

The statement associated with case 005 in Table 5 is a fragment that should not have been scored. There is no probe, and no reasons are provided, revealing very little about the form of thinking employed by this individual. The same is true for the statement associated with the abstract score in case 114. These statements are rescored as missing in subsequent analyses.

The statement associated with case 115 was coded abstract. However, a causal proposition present in this statement brings this decision into question. At the abstract level, several abstract concepts can be strung together in additive, multiplicative, or sequential relations, but they are not related to one another in a causal relation. Here, four abstract concepts—teachers and their students, class size, student comfort, and interaction—are related to one another in a clear causal relationship: Small class size, by making students feel more comfortable, contributes to interaction among students and teachers. This statement should have been coded formal and will be recoded as formal for future analyses.

The statement associated with case 020 was coded abstract, but once again, a causal relationship exists that was overlooked by the rater, probably because the individual illustrated this concept with several concrete examples. Here, the participant asserts that education, because it provides a range of skills, prepares us for life (viewed as a group of life-tasks). This is a causal proposition that should have been coded formal, and will be recoded as formal for future analyses.

The statement associated with case 132 was coded systematic, probably because the participant used two words often associated with the systematic stage, *organization* and *system*. Yet, upon closer examination, the meaning associated with these word does not appear to be at the level of hierarchical complexity required at the systematic level. It is not enough to state that a system exists. Such terms can easily be employed without complete comprehension of their meaning. A systematic reasoner will demonstrate some understanding of the way in which

components of a system interact. There is no indication of that here. This statement, instead of being systematic, is formal, in that it contains a series of causal propositions, one of which is the assertion that if a school is organized, then students will know what they have to do, won't get behind, and will be able to keep their grades up. This statement will be recoded as formal for future analyses.

Case analysis II

Following recoding of the six cases described above, the data were resubmitted to partial credit analysis. This time, the proportion of the of the observed case estimate variance considered to be true was .89. As shown in Table 6, all items have acceptable infit (weighted) and outfit (unweighted) statistics, indicating, once again, that they measure the same dimension of ability.

Table 6

Fit Statistics for Good Education Data

ENTRY NUMBR	RAW SCORE	COUNT	MEASURE	REALSE	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTBIS CORR.	ITEMS	G
3	234	88	48.3	1.2	.97	**-.2**	1.32	**.8**	A .79	Item 3	0
1	233	91	48.9	1.3	1.05	**.4**	1.05	**.2**	B .77	Item 1	0
2	217	87	50.9	1.2	1.04	**.3**	.88	**-.4**	C .77	Item 2	0
5	201	81	51.8	1.3	.95	**-.4**	.74	**-.8**	D .78	Item 5	0
6	247	95	51.0	1.2	.98	**-.2**	.79	**-.9**	c .78	Item 6	0
4	224	86	48.7	1.2	.93	**-.7**	.71	**-.9**	b .76	Item 4	0
7	224	88	50.3	1.2	.80	**-1.6**	.61	**-1.7**	a .85	Item 7	0
MEAN	226.	88.	50.0	1.2	.96	-.3	.87	-.5			
S.D.	13.	4.	1.2	.0	.08	.6	.23	.7			

Figure 3 shows the stage estimates and measures. Stage estimates are displayed on the left of the figure, and measures are displayed on the right (by case number).

Measures are presented in three different type styles, to highlight different types of performances. Once again, the heavy black case numbers represent persons who performed at a single stage, the heavy gray case number represents one person who performed at 3 adjacent stages, and the light black case numbers represent people who performed at two adjacent stages.

To the left of the center of Figure 3 is the logit scale, which ranges from 0 to 100 in 10 point increments. (This scale is a transformed logit scale. It is important to keep in mind that there is no true zero in a logit scale, which is not a true ratio scale.) The meaning of this scale, in terms of performance, is illustrated by the horizontal bands to its right. Each band represents one logit (10 points, on the present scale). The type of performances generally found at the range encompassed by each band is described in reverse type[1]. More distinct patterns are identified in the formal/ systematic and systematic/metasystematic ranges than in the abstract/ formal range. This may be due to a difference in the nature of stage transitions at the lower stages. Alternatively, it may be an artifact of the relatively small number of participants performing in the abstract/formal range and the fact the average number of statements produced by individuals in this range was lower than at higher stages, resulting in more missing data. The mis-scored responses for the 6 cases identified in the original analysis suggests a third possible explanation. It is conceivable that, due to coder error or a lack of clarity in the scoring criteria for the abstract and formal stages, more coding errors have been made at the abstract/ formal level than at the formal/systematic and systematic/metasystematic levels.

This might contribute to the 'noise' in the results at this level. However, if this was true, higher standard errors would be associated with person measures at the lower stages than at the higher stages, and this was not found to be the case.

All of the performances fit the model.

Standard errors for individual person estimates range from 3.5 to 30.9, with a mean of 8.2. This means that the 95% confidence interval for an average ability estimate spans a 32.8 point range (plus or minus two errors). With an average error of 8.2, the measurement standard deviation of 25.72, and the associated .89 measurement reliability, we can distinguish 4.3 statistically distinct strata (ranges) in the measurement continuum. Strata are ranges with centers separated by three errors. The presence of four strata demonstrates that each of the four stages represented by the person measures in the figure is statistically distinct from the others (for more on strata, see Wright & Masters, 1982).

Interestingly, the persons with the largest standard errors are of two kinds. They are persons with perfectly uniform performances—at a single stage—and one person with a performance that spans more than two stages. If the 55 people with perfect performances (the large group strung out just above the center of the figure) are eliminated, the mean error for the remaining measures is 5.18, which yields an average confidence interval of 20.72 points, still large, but only 63% of the mean confidence interval with these cases included. In contrast, the mean error for the perfectly consistent performances is 12.77, resulting in a confidence interval of over 51 points.

What could account for the association between large standard errors and uniform performances? Recall that each stage is defined as an hierarchical integration of the operations of the previous stage, such that the operations of the earlier stage become the content of the current stage. Once new operational structures are available, it is expected that they will be applied, relatively rapidly, to an increasing range of content, and that further stage change will not occur prior to the extension of existing operational structures to a wide range of content. Because the shift from one stage to the next involves a qualitative transformation in the organization of thought (a major alteration in thought structures), the initial shift from one stage to another may be more "difficult" than the application of existing operational structures to a wider range of content. This leads to the expectation that persons who have organized a wide range of knowledge at a given stage of hierarchical complexity will consistently perform at that stage. Such performances are referred to as *consolidated*. Further, individuals who are in transition from one stage to another are expected to use the operational structures of two adjacent stages, but are not expected to use the operational structures of more than two stages. Performances of this kind are referred to as *transitional*. Performances that span more than two stages or are consolidated at one stage are both associated with large standard errors. While larger standard errors are expected when performances span a wide number of stages, large standard errors for consolidated performances are counter-intuitive.

In a Rasch analysis, evidence of a large pecentage of consolidated performances

emerges in statistically significant gaps appear between the item difficulty estimates of items at adjacent stages. This tendency for items at the same stage to cluster is called segmentation, and reflects a strong propensity for individuals to reason predominantly at the highest stage available to them. The stronger this propensity, the greater the gaps between clusters. Prolematically, this pattern of segmentation means that all of the items are virtually the same (in terms of difficulty). When this happens, it is as though measures are based on only one item. For performances that receive the same stage score on every item, this is literally the case. The high standard errors associated with these performances reflect the lower confidence that can be placed in estimates based on too little information. Consequently, while the pattern of performance is consistent with Commons' theory, it is not entirely captured by the Rasch model. The high errors associated with uniform performances are counter-intuitive. A model is needed that can take into account this tendency for individuals to employ the highest stage available to them, assigning error terms that reflect the actual confidence that can be placed in measures based on uniform and near-uniform performances.

Item analysis

The reliability of item estimates is .01. This low coefficient stems from the fact that a large percentage of performances are at the maximum expected level for their stage. In the present study, 36% of respondents (55 out of 153) performed at a single stage. From the perspective of developmental theory, these uniform performances are expected, but their rigidity effectively reduces the amount of

information on which to base item estimates. Rasch's probabilistic model expects more random variation than is found in this data. This leads to several problems in interpreting fit and error statistics, including the problems of variation in standard errors noted above in the case analysis.

Standard errors for the item step estimates (stage estimates) range from 1.70 to 2.81 logits. Two standard errors are shown in Figure 3 as grey areas around each group of stage estimates. These are calculated from the standard errors for the easiest (the lower range of the error range) and most difficult items (the higher end of the error range) at each stage. These can be thought of as 95% confidence intervals. Stage estimates for the systematic and metasystematic stages cluster within a 7-8 point range. This is expected because of the random assignment of scores to each of the 7 'items.' The wider range of estimates at the formal stage are very likely the result of the large number of empty cells for performances in the abstract/formal range.

The gaps between stage estimates provide evidence that upholds the invariant sequence criterion. This can be illustrated by examining probabilities of performance at various stages of hierarchical complexity for an individual case. For example, the probability that participant 012, located at the 50-point level on the logit scale, will perform at the systematic stage is 50%, while the probability that he will perform at the metasystematic stage is close to 0%. Gaps between stages also, as mentioned above, provide evidence for the concept of consolidation. They reflect the propensity for extensive horizontal *decalage* to precede vertical *decalage* in individual development.

The pattern of measures in Figure 3, taken in concert with evidence from the stage estimates, tells a compelling story. The systematic/metasystematic transition provides the clearest illustration. Locate the band to the right of the 25 to 35 point range. This band represents the consolidated formal performances, those for which the overwhelming preponderance of responses were scored as formal. Just above this band is the range in which individual performances include about 25% systematic statements. Above this is the band that includes performances that are about 50% formal and 50% systematic.

Consider the triangular shape of the distribution of measures in the three bands between 35 and 65 points. As one scans up from 40 to 60 points—the range in which performances go from 86% formal and 14% systematic to 86% systematic and 14% formal—a generally increasing number of performances is found. A similar pattern is seen in the 75 to 95 point range—the range in which performances go from 14% metasystematic and 86% systematic to 86% metasystematic and 14% systematic. This is yet another illustration of the strong tendency for individuals to apply the same form of thinking to a range of content, at least within a given subject area. Very few individuals do only a little reasoning at their highest stage. Though this is not directly implied by the model, from a temporal perspective, this may mean that horizontal *decalage* is rapid, relative to stage change.

The pattern of performance shown in Figure 3 is consistent with reports from analyses of stage of performance in the domains of moral reasoning and evaluative

reasoning about the good (Dawson, 1997; 1996d; 1998) and other domains (e.g., Draney, 1996; Goodheart, 1996a; Goodheart, 1996b; Müller, Reene & Overton, 1994; Müller, Winn & Overton, 1995; Noelting, Coudé & Rousseau, 1995; Noelting, Rousseau & Coudé, 1994; Wilson, 1989).

Age, education, and sex effects

Table 7 shows the relationship between GMHC stages of performance and age groups. Though two performances with more than a 23% probability of formal reasoning were found in the 5-9 year-old group (both gifted 9-year-olds), formal reasoning does not generally appear until the early teens. More than a 23% probability of systematic reasoning is not found before 20 to 24 years of age, and more than a 23% probability of metasystematic reasoning (with the exception of a single very bright 19-year-old college student) was not present until 25 to 29 years of age, and not common in this sample until 35 through 39 years of age. These results mirror those in previous studies (Armon, 1993; Armon & Dawson, 1997; Colby et al., 1983; Commons et al., 1989; Walker, 1989).

Table 7

Age and Stage in Evaluative Reasoning about Education

(n = 153)

	Stage				
	Concrete	**Abstract**	**Formal**	**Systematic**	**Meta**
Age					
5 through 9	2	6	2		
10 through 14		1	20		
15 through 19			10		**1**
20 through 24			1	22	
25 through 29			1	8	1
30 through 34				4	3
35 through 39			2	3	9
40 through 44				8	11
45 through 49				7	5
50 through 54				1	2
55 through 59					4
60 through 86			2	12	4

Correlations of measures with age, educational attainment, and sex are shown in Table 8.

Table 8

Correlations of Measures with Age, Educational Attainment, and Sex

	Age	Edu	Sex
Measure	.66	.86	-.13
	$p < .05$	$p < .05$	$p = .111$

Both age and educational attainment correlate strongly with the measures, but sex does not. Because sex differences in developmental attainment have been the subject of considerable debate in the literature, and have been associated with differences in age and educational attainment, sex was not dropped from the remaining analyses. Instead, it was entered into two multiple regressions on the measures, first with age, and second, with educational attainment.

The relationship between stage measures and age and sex was examined with a series of regression analysis. First, logistic regressions of stage measures by age were run for each sex. A logistic model was chosen because it explained more of the variance stage measures than a linear model. (The logistic model has been employed in other studies of age and stage relationships, with similar results (Armon, 1984a; Armon & Dawson, 1997) As depicted in Figure 4, the curves for males and females are largely congruent up to about age 20, then begin to diverge. Note also that the amount of the variance explained by age is less for women (54%) than for men (71%).

To examine these apparent differences in the relationship between stage and age for men and women, a multiple regression of sex and age effects on measures was conducted using the log of age. It revealed that older males tend to perform at somewhat higher levels than females, explaining less than 2% of the variance in measures. The multiple regression of the log of age and sex on the case estimates resulted in the following equation:

$$R=.80,\ F(2,147) = 128.80,\ p<.001,\ \text{Stage Measure} = -61.45 + 84.14_{\text{logage}} - 5.89_{\text{sex}}$$

$$t_{\text{logage}} = 15.83,\ p<.001,\ t_{\text{sex}} = -2.04,\ p<.05.$$

Moderate to strong correlations between age and stage are common in the literature, particularly in studies involving young and life-span samples (for examples, see Armon, 1993; Armon & Dawson, 1997; Colby et al., 1983; Dawson, 1997; Dawson, 1996c; Demetriou et al., 1993; Nisan & Kohlberg, 1982; Oser, 1994; Selman, 1971a; Welfel & Davison, 1986). The fan-shaped distribution seen in Figure 4 has also been reported eleswhere (Armon, 1993; Armon & Dawson, 1997; Dawson, 1997; Dawson, 1996a). It has been suggested that this distribution reflects the increasing diversity in life experiences—education, work, and personal relationships—encountered as we age, each of which has a different impact on development.

Second, the relationship between stage measures and educational attainment and sex was examined. A linear model provided the best fit, as depicted in Figure 5. Females appear to perform slightly less well than males across the educational spectrum, but the effect size for sex is non-significant. One year of education explains a difference in ability of about half of a logit (5 points on the present scale), or, in other words, individuals gain about one stage for every six years of education: (Figure 2 shows the relationship between logits and stages.) The equation for the multiple regression of educational attainment and sex on measures is as follows:

$$R=.87,\ F(2,147) = 220.37,\ p<.001,\ \text{Stage Measure} = -9.74 + 4.84_{ed} - 4.05_{sex}$$

$$t_{ed} = 20.75,\ p<.001,\ t_{sex} = -1.69,\ p < .10.$$

Strong linear relationships between educational attainment and stage have been reported elsewhere in the literature (see, for example, Armon, 1993; Armon &

Dawson, 1997; Bakken & Ellsworth, 1990; Colby et al., 1983; Dawson, 1997; Dawson, 1996a; Pratt et al., 1983; Walker, 1986). As can be seen in Figure 5, the relationship between educational attainment and stage is relatively stable across the life-span, though there is increasing noise with advancing age.

The sex effects seen here are similar to those reported in the literature, in that they are small when present, and are affected by factors such as educational level and age, suggesting that sex differences have more to do with contextual factors than an inherent tendency of one gender to out-perform the other (see Walker, 1984b, for a review of the relationship between sex and moral stage. Also, see Armon, 1993; Armon & Dawson, 1997; Bakken & Ellsworth, 1990; Dawson, 1997; Dawson, 1996a; Demetriou et al., 1993; Keller & Wood, 1989).

Chapter 7

Meaning and orders of hierarchical complexity

Good learning is not just a theory; it's not just something you read in a book, it's something that could be applied in some way to any situation.

—Female, 24

The goal of this chapter is to analyze and describe continuities and discontinuities in meaning from one developmental stage to another by conducting a detailed investigation of the development of meaning along two metaphorical strands, *education as play*, and *education as preparation for the good life,* employed by 70% and 91% of respondents, respectively.

The analysis represents an explicit effort to apply a relational approach to investigating the relationship between the construction of meaning and the hierarchical complexity of performances. Here, the relational dialectic is played out in two important ways. First, distinguishing between the hierarchical complexity of linguistic performances and their conceptual content makes it possible to examine the ways in which they interact with one another to produce particular meanings. Second, both quantitative and qualitative methods of analysis are applied in a deliberately iterative, "checks and balances" (Fisher, 1998; Fisher, 1994; Overton, in press) process, in which theory is employed in interpreting analyses and analyses are employed in testing theory.

The methodology is complex and tends to overshadow the substantive findings. Consequently, before turning to the analysis itself, I provide (1) a synopsis of the

findings, and (2) an interpretive discussion of the relationhip of these findings to those from other studies of the development of reasoning about educational concepts. These discussions are followed by several analyses of the distribution of propositions by stage, employing both interpretive and psychometric methods. Finally, I reintegrate the propositions associated with each stage with the organizing principles of that stage, forming descriptions of the kind of conceptualizations expected at each order of hierarchical complexity. These are referred to as *categories of description*, a term borrwed from Marton (1981, 1994a). Brief summaries of these categories are shown in Figures 8 and 9, which are described in more detail on pages 147-170.

The development of evaluative reasoning about education

An important distinction between the approach taken in this analysis and those taken by earlier developmental researchers in related domains is that descriptions of reasoning associated with each of the stages are not intended as stage definitions. Stages are conceptualized as orders of hierarchical complexity and identified with reference to attribites of performances that do not rely on an analysis of their *particular* conceptual content. It is assumed that the conceptualizations associated with these stages are affected not only by the hierarchical complexity of constructions, but by informational, contextual, and dispositional factors. I do not make the claim that there are stages of evaluative reasoning about education that are structurally distinct (at the level of structure associated with stage in the definitions offered in Chapters 3 and 4) from stages of

reasoning in any other domain. The following descriptions of evaluative reasoning about education should be viewed in this light.

Education as play: As shown in Figure 8 (page 150), at the concrete order, a good school (for concrete children, education equals school) is one in which you get to play and have fun. The child knows that she likes to play and have fun, so these activities should take place at school. She does not connect concepts of fun and play to learning.

> *What kinds of things are good to do at school?*
>
> Play recess. I play with my friends (Respondent 151).

Having fun (or playing) is a specific school activity, separate from other activities.

> *What do you think a good education should be?*
>
> Passing grades, having fun, being nice (Respondent 88).

At the abstract order, respondents connect learning (as knowledge acquisition) with having fun. A concept of interest appears for the first time at this stage, though it appears to be a relatively undifferentiated alternative to the word *fun*. Certain fun or playful activities, such as learning games, are explicitly seen as educational. A beginning notion of motivation is present in respondents' assertions that they learn more when they are interesed or having fun. It is the teacher's job to make learning fun or interesting. The following constructions are typical.

> "If you don't like school, which most people don't, if you had fun during it you would like it more...[You] want to have fun, but not too much fun like bouncing off the walls, but fun with learning"

(Respondent 135).

> "When you're bored [not interested], you don't really want to do [the boring thing], you want to do a game which is fun. You don't learn when you're in the game as much as you do when you're working. [It would be good to find] games that have work so you can learn things too, like a math game, or computer game" (Respondent 111).

At the formal order, active engagment in learning is a central theme. The concept *interest,* is differentiated into *involvement, engagement,* and *inspiration,* all of which point to the absorbtion of the learner into the learning process. Inspiration, stimulation, involvement, and engagement are generated by others (teachers), however, suggesting that the potential engagement of the learner does not imply that learning is intrinsically involving. Social interaction is important insofar as it enhances engagement. The following constructions are typical of this order:

> You don't learn anything if it isn't fun or creative. The students aren't focused on boring stuff (Respondent 131).

> [As a student] I needed someone to inspire me. Some teachers I found just can't [do that]. They're just very bland and you can't really learn a lot from them. And others are really inspiring and you can learn more (Respondent 114).

> [If the students are really interested in something] they're more likely to really understand and want to understand and to really actually work on it because they'll want to know it (Respondent 146).

> [It is better to learn by doing than to learn from books because] it's more interactive. It's funner for the brain and it's more stimulating than

> if you just read a book all day (Respondent 148).

At the systematic order, the concept of active participation in learning is coordinated with the notion that intellectual engagement can be increased through social interaction to produce the idea that good learning takes place in a discursive, participatory context. Though this is not yet viewed as the defining context for learning, it is seen as the most enjoyable.

> I enjoy it when my teachers make me realize that what I'm saying doesn't really have anything to back it up...I think one of the key things is when someone questions someone else. That's when you can learn a lot about how to come up with answers...Usually I have come up with answers by starting off with a series of questions—like things that make me curious (Respondent 83).

> I believe that it is inherent in us, if it isn't distorted, to enjoy learning. I mean, we have curiosity—some more than others. In some it's blunted. Now it's true that there are emotional aspects to this in terms of what you get from other people. So part of what is necessary in education is the right emotional climate of your being appreciated, and to some extent, you're motivated to please. Because that pleases people. And you find all these rewards in [pleasing others] (Respondent 41).

Learning is now sometimes seen as intrinsically enjoyable or fun.

> [Learning is important] because it's fun!...You feel expanded, you feel that you're growing as you take in more and more and more and as you develop more and more capacities. And that in itself is gratifying (Respondent 41).

However, it is more often viewed as enjoyable when it is done in a way that is appropriate to the particular individual. People are seen to have different styles of

learning:

> [You should study] things you're interested in, in a way that you're interested in. If you approach certain subject matters in the wrong way you'll completely say, 'I hate this.' But if you approach it in another way you might realize that you love it. It's important to study everything in a way that's suitable for you. (Respondent 14).

At the metasystematic order, the notion that good learning takes place in social interactions is coordinated with the idea that learning is discursive. Education is viewed as a dialectical process in which teacher and student (or student and student) get caught up in the playful activity of learning. Testing—as a continuous spiral of feedback—is one way of conceptualizing this playful back-and-forth. This dialectic defines the learning process and produces positive consequences for individuals and for humanity as a whole. The inherent playfulness of this dialectic is embodied in the following:

> Good education not only builds on an awareness of learning through what is already known, but realizes that the teachable and the learnable are...idealizations that emerge from the convergence of question and answer, teacher and learner, text and reader. These idealizations absorb each of the various partners in the dialogue into themselves as they play themselves out through those partners, recreating those involved to the extent that they lose themselves in the playful back-and-forth motion of the activity of the thing itself (Respondent 125).[2]

> One of the main values of a good education is creating human beings who are engaged, who are proactive, people who are, like in skiing, leaning down the hill, they're having fun, they're not afraid or intimidated. It doesn't necessarily mean they just have incorporated

> some silly notion they can do anything—because they can't—but they've acquired a set of skills in terms of learning and a set of reasonable expectations and they know who they are and what they can do in general. There comes this notion or attitude, [that learning] is fun, or can be fun (Respondent 130).

The coordination of societal and individual perspectives is also apparent:

> The educational process of convergence, absorption, and separation is good because through it individuals and societies are transformed into fuller realizations of their productive potentials. Good education...uses what is known about learning to help students learn about the world and about learning itself, with the aim of bringing people into the world who can fulfill themselves and improve society by putting their contributions and needs into play in a way that is cognizant of ecological, psychosocial, spiritual, cultural, economic, and political balances and interdependencies (Respondent 125).

While the theme, *education as play,* is found across all orders of heirarchical complexity identified in this sample, the meanings associated with this metaphorical strand change from the simple notion that a kid wants to have fun to the complex construction of an ideal education as one in which learner and teacher are absorbed into a playful dialectical spiral of learning and testing to the mutual benfit of individuals and society.

Education as preparation for the good life: Concrete operational constructions of *education as preparation for the good life* were not identified in this sample. This does not mean (as discussed below) that there are no concrete constructions on this strand, but in the absence of empirical evidence, a description of reasoning at the concrete order on this metaphorical strand would be premature.

As shown in Figure 10, at the abstract order, repondents make a connection

between education and life skills. The purpose of a good education is to prepare students to get a job and to be able to perform other adult tasks.

> *(Interviewer) Why might it be important to get a good education?*
>
> It just helps you get a job when you're older (Respondent 139).

In addition, a good education provides specific skills, such as mathematics, reading, and writing, that are useful for accomplishing particular life tasks.

> *(Interviewer) Why do you think that people have to go to school?*
>
> Because they have to learn.
>
> *(Interviewer) Why do people have to learn?*
>
> If they didn't learn then they wouldn't know anything about the world.
>
> *(Interviewer) Is it important for people to know about the world?*
>
> Yes.
>
> *(Interviewer) Why is that so important?*
>
> If somebody didn't know what a fire was then they could just light a match and they could start a fire and they might get killed (Respondent 159).

Abstract performances also indicate an understanding of the connection between having a job and taking care of basic personal needs. For example:

> If you don't learn it you won't really be able to get a good job.
>
> *(Interviewer) And why is it important to get a good job?*
>
> To be able to have some money to buy things, so you can get food (Respondent 111).

At the formal order, the notion established at the abstract order, that education provides skills for adult life, is further differentiated. Skills are sub-categorized into

types, such as those required for effective social interaction, basic survival, and goal attainment.

> Education should prepare you to live in the real world, [people should] be able to read and write and speak correctly, to express themselves, to communicate...I've met people who are supposedly well educated who can't communicate with a turtle. They're lacking any social skills. So an education should include that. (Respondent 5).

The acquisition of a range of skills and knowledge is important because it leads to a good life in which individuals are prepared to be good people who are able to contribute to their own well-being and the well-being of significant others (such as in supporting one's family).

> [A good education is] important in everyone's life, because you want to learn in your life, you want to learn how to do the best you can do in your own life, for your family and for your friends, relationships, and all that. So it's important that you learn how to deal with things (Respondent 114).

At the systematic order, the goals of a good education are to enhance psychological and material well-being as important and different aspects of the good life, and to enable the individual to function in and contribute to society. Quality of life is, for the first time, related to how one functions within the larger society.

> [Liberal arts] gives you roundedness, wholeness. You need [skills] to survive. You need liberal arts to be happy, to be fulfilled (Respondent 62).

> [People need a broad education] for background and evaluating things all through life, some basis for thinking about whatever they're doing. I hate to use [the term] well rounded...but I think [a broad education] also gives more satisfaction in life. [People] can appreciate more of the

> things that are available to them (Respondent 32).

Psychological well-being and material well-being are similarly coordinated. Though the concept of self-esteem appears at the formal order, it is employed only with reference to students' need to feel good about themselves in order to learn. Not until the systematic stage is this concept coordinated with the idea that education is preparation for a future good life in which one's psychological well-being is of central importance.

> Self-esteem is feeling good about yourself and being confident about yourself. When I was young I was such a goal oriented person. I felt that having goals for myself and achieving those goals made me feel better about myself. That gave me self-confidence (Respondent 138).

> [People] should be able to participate in [learning] something that's going to make them feel good about themselves and be exciting and interesting for them, inherently interesting for that individual. (Respondent 90).

The enhancement of practical and psychological well-being are strong themes at the systematic order.

> [A broad education makes you] more aware of options, possibilities. You're not so easily intimidated by things that are foreign to you or strange because you're aware of them. There are a lot of things just by being familiar with them to a degree you're not so intimidated by them, you're more willing to try them (Respondent 40).

> So the more education you have, in the general broad sense, the more options you have as far as living is concerned, meeting people is concerned, doing things, doing things for people, getting things done. It

> just makes sense that the more knowledge you have, the more you have available to you (Respondent 046).

> [A good education makes] possible a much wider range of experience. That primarily. Secondarily, there's the thing about skills, about your using your own capacities to the fullest which gives you more gratification (Respondent 43).

Personal well-being and the notion of making a contribution to society are seen as complementary.

> It's important to feel that you're contributing to society. All I can say is in terms of your concept of your self, and of what the good life is to you, that's part of it, is contributing something to society that is more positive than negative. (Respondent 43).

> [Education should prepare you to do] something that will be socially useful and which will fulfill your own capacities for growth (Respondent 42).

At the metasystematic order, personal well-being and social welfare are fully integrated. Psychological and material well-being are viewed as part of a larger picture in which education is a vehicle for the simultaneous and non-contradictory promotion of both social justice/welfare and personal growth/empowerment/autonomy.

> [The aim of a good education is] to promote an autonomous person who is able to integrate [individual and societal perspectives] into a universal perspective (Respondent 8).

> The aim of a good education is education for autonomy, which is the choice

> of the principle of justice for society, upholding justice (Respondent 73).

> One of my concepts of the good life is individual growth and development and opportunities for that. And those opportunities are in education. In order for the society better itself, in order for it to create a better world for all people to live in, it has to have those things that are included in education, including self-reflection. Not just as individuals, but self-reflection on the society, the programs, the development, the justice structure. You know, it has to be self-critical. You can't have that without educating individuals (Respondent 1).

> A good education is an initiation. It's an initiation into the institution of being human. To me,...the result of a truly good education is a full understanding of the institutions of mankind, which involve morality, intellectual achievements, historical events, writing, all the things that man has developed over time... I see that the ultimate outcome of this is sort of the fully actualized person, or a powerful character, minimally, at least, capable of critically viewing the problems, and objects and experiences that are happening in the present with the historical perspective as well as a critical and objective perspective in the present (Respondent 68).

Though the general theme of education as preparation for the good life is present at every stage, the particular meanings constructed along this strand change from the idea that a good education prepares one with specific skills needed for stereotypical life tasks, to the complex notion of a good education as one that supports the development of individuals and the societies in which they live, to their mutual advantage.

Evaluative reasoning about education and other evaluative and education-related developmental sequences

While the GLSS, SISS, and GSSS have been shown to measure the same underlying

variable—hierarchical complexity—through a comparison of performances on each of these instruments (Dawson, 1998), the relationship between meanings constructed at analogous stages on these scales has not yet been explored. Does an analysis in which hierarchical complexity and conceptual content are initially analyzed independently produce descriptions of the meanings constructed at each stage that are similar to stage definitions produced through other methods? At the end of Chapter 2, I discussed interpretive similarities between several developmental sequences relevant to the education theme. Here, I explore similarities between the categories of description for *education as play* and *education as preparation for the good life* and some of the sequences described in Chapter 2 as well as Armon's *good life* and Kohlberg's *moral judgment* sequences.

Table 9 shows a summary of levels in 7 sequences and their hypothetical relationships with one another. Included are Kohlberg's stages of justice reasoning (or moral judgment), Armon's stages of good life reasoning, Marton's categories of description for conceptions of learning, Ammon and Black's levels of instructional goals, Loevinger's levels of educational thinking, and Dawson's categories of description for *education as play* and *education as preparation for the good life*.

The concrete level (stage 2 in Kohlberg's sequence, and the first level discussed in Chapter 2) is represented in Table 9 by Marton's *conceptions of learning,* Ammon and Black's *instructional goals,* and Loevinger's *educational thinking.* Knowledge at this level is viewed as absolute, held by authorities, and learned through a process of internalization or memorization. The teacher's job is to present knowledge, and

the learner's job is to absorb this knowledge. There is no further coordination of the perspectives of teacher and learner, no concept of interaction. Reasoning at Armon and Kohlberg's stage 2 is characterized by this failure to coordinate perspectives. At stage 2, individuals are aware that everyone has interests to pursue and even that these can conflict, but what is right or good is seen entirely from the point-of-view of the individual. This failure to coordinate perspectives is also seen in concrete conceptions of *education as play*. Conceptions of play/fun and conceptions oflearning are not coordinated. The good thing about school is playing or having fun. The concrete conception of learning as memorization is not compatible with a conception of play/fun, which makes it unlikely that the two could be connected at the concrete order.

In Table 9, the abstract level is shown between stages 2 and 3 in Kohlberg's sequence (level 2 in Chapter 2), and is represented in Marton's scheme by learning for application, in Ammon's sequence as a focus on content knowledge, and in Dawson's sequence as the notions that learning can be fun and interesting, and that a good education can prepare you with the specific skills needed for life. As shown in Chapter 2, at this level, learners understand that even authorities disagree on what constitutes truth. This realization undermines the notion that knowledge is absolute. A notion of good knowledge as useful knowledge emerges at this level, which, to some extent allows the individual to evaluate the truth value of knowledge in terms of what works in concrete applications. This new emphasis on using or applying knowledge is associated with a conception of learning as a more

active process than at the concrete level. To learn skills, one needs to be involved in practice. The emphasis on applying knowledge also leads to specification of the appropriate content of learning, as seen in Ammon's content oriented level and the abstract level of *education as preparation for the good life.* Here, the appropriate content of learning is the set of skills needed for work and other life tasks. An emphasis on learning through doing and for application are also seen at the abstract level of *learning as play,* where learning and enjoyment are linked for the first time via the concept of interest. When students have their own concepts of why and what they are learning, and experience being caught up in the activity of learning, they can conceptualize learning as a fun activity.

Individuals at the formal level (Kohlberg's stage 3 in Table 9, and the third level discussed in Chapter 2) are caught up in an apparent contradiction, asserting both that (1) there are absolute or ultimate truths and that (2) knowledge is contextual. Whereas, at the previous level, respondents attempted to solve this conundrum by seeking correct or "good enough" truths (e.g., in the successful application of learning), respondents at this level view truth-seeking as an active process that will eventually lead to the real truth. Different truths, though inadequate in themselves, are useful in working toward ultimate truth. This kind of attempt to coordinate perspectives to the end of making right decisions is reflected in Kohlberg' stage 3, in which the perspective of the individual in relationships with other individuals takes prominence. In justice conflicts, primacy is given to shared feelings, agreements, and expectations, relating points-of-view through the concrete Golden

Rule—putting yourself in the other guy's shoes. Stage 3 of Armon's good life sequence is also characterized in terms of this kind of mutuality or perspective-taking, which spills over into conceptions of good work in which personal enhancement, supporting the well-being of significant others, and helping others are central. This is very similar to the description of conceptions of *education as preparation for the good life* at the formal level, in which a good education provides a range of skills that permit one to provide for the well-being of the self and significant others.

The theme of engagement is common at the formal level. Because truth is now seen as something that remains to be discovered, the learner must participate in its pursuit. Though formal reasoners do not assert that *all* learning ought to be interesting, engaging, or fun, they clearly articulate the ways in which learning can be made more enjoyable. Most importantly, they assert that learners should be actively engaged in the learning process. Engagement is enhanced by social interaction in the form of group projects and discussions, and in teachers' ability to inspire, involve, and engage students.

At the systematic level (Kohlberg's stage 4 in Table 9), knowledge is no longer viewed as absolute. Earlier conceptions of learning as knowledge reproduction and learning as personal insight are integrated into a systematic conception of understanding, in which all knowledge is seen as relative to interpretation and context. Consequently, personal understanding and evaluation are emphasized. The goal of teaching is to guide students toward ways of thinking that enhance

understanding by encouraging them to reflect on their own thought processes. Students should be caught up in the activity of learning. This new conception of learning is compatible with the systematic conception of *education as play,* in which good learning is viewed as inherently enjoyable, ideally taking place in a discursive, participatory context, and motivated by the self's desire for understanding.

Individuals are able to take the perspective of the social system at the systematic level, but there is a tension between the perspective of the social system and the individual. Societal and individual well-being are not yet fully coordinated and consequently are often viewed as contradictory. In Kohlberg's moral judgment sequence, the right is seen in terms of what is good for the system, often at the expense of the individual. In contrast, stage 4 in Armon's good life sequence is focused on the pursuit of individual goods. Enjoyment, physical comfort, and the financial freedom to enhance and maintain one's life-style are among the goods valued at this level. The social system is a concern, but only in a limited sense. For example, individuals performing at this stage often assert that one's individual pursuits should not conflict with the good of the system as a whole, or that achieving the good life involves making a contribution to society, but these are not seen as goods for the self. Rather, they are goods for society. Similarly, at the systematic level of *education as preparation for the good life,* respondents assert that education should enhance material and psychological well-being and prepare one to take one's place as a contributing member of society, where enhancing

material and psychological well-being is good for the self, and contributing to society is good for the social system.

True coordination of individual and societal perspectives takes place at the metasystematic level. In Table 9, this level is represented by Kohlberg's stage 5, in which individuals focus on issues of prior rights, social contract, and utilitarianism; Armon's stage 5, in which personal autonomy is central; Marton's *learning as personal transformation* level; and Dawson's metasystematic levels of *education as play* and *education as preparation for the good life.* At the metasystematic level, the coordination of systems—the social system, the self system, knowledge systems, etc.—is a prominent theme. At stage 5 in Kohlberg's sequence, individuals solve moral conflicts by coordinating interpersonal and societal perspectives through over-arching principles like the utilitarian principle of maximizing the good for all, universal human rights, or the social contract. At stage 5 of evaluative reasoning about the good life, individuals coordinate societal and individual perspectives by developing a conception of autonomy that views the good of the individual as inseparable from the good of the society. In a similar way, metasystematic performances coordinate a concern for social justice/welfare and personal empowerment/autonomy, and invoke an educational system that simultaneously promotes both.

The type of coordination employed in metasystematic performances is dialectical. The good is not seen as something that can be defined from an objective viewpoint, but as an evolving understanding that continually incorporates new

knowledge from multiple perspectives. In the same way, learning itself is viewed as transformative. In Marton's sequence, learning is defined in these terms, and in Dawson's *education as play* sequence, education is viewed as a fundamentally playful activity in which knower and known, teacher and student are caught up in and transformed through a dialectical spiral of meaning making.

As was found at the end of Chapter 2, it is possible to compare conceptions of learning at analogous levels across a variety of developmental sequences. The fact that data were gathered and analyzed differently does not prevent a meaningful interpretive analysis. Moreover, this type of analysis can be carried out across a range of related domains. In order to make these comparisons, however, it is necessary to look at deeper levels of meaning than the particular content of conceptions. Here, I compared sequences on the basis of shared conceptions, many of which underlie the surface content of performances. For example, a direct comparison of Marton's *learning as personal transformation*, at the metasystematic level and Dawson's *education as the playful activity of meaning making* does not reveal any surface commonalities (at least, to this researcher). It is deep concepts (or ways of conceptualizing) that give these categories of description something in common—the fact that both invoke the ongoing dialectical coordination of systems, the knowledge system and the self-system.

The following section presents the process employed in developing the categories of description presented in the first part of this chapter.

General patterns in the distribution of propositions

As reported above, 2,424 propositions were identified in the interviews and coded into 644 categories—an unwieldy mass of data. To solve this problem, I divided the propositions into categories representing individual strands of meaning. These included *education as play, education as preparation for the good life, education as an initiation into humanity, education as work, education as learning,* and *education as social interaction.* Because the analytical approach employed here is laborious, an examination of all of these themes was not feasible within the constraints of the present project. It was possible, however, to conduct exhaustive developmental analyses of *education as play* and *education as preparation for the good life.*

First, proposition categories with meanings that appear, on interpretive grounds, to contribute to each of these metaphorical strands of meaning were identified in the dataset. I recruited a colleague to assist me in this process. (All of the propositions each of us selected are included in the initial phases of this analysis.) The selected propositional categories were first examined by ordering them according to the estimated global stage of performance of each respondent, as can be seen in Figures 1 and 6. This initial attempt to bring the results of the structural and content analyses together reveals some interesting patterns. First, some propositions appear at an early stage and then disappear. For example, in Figure 1, the *unelaborated* notion that a good education includes play occurs only at the concrete stage. Other propositions emerge early and continue to be found across a range of stages—sometimes across the entire spectrum of development examined here. The notion

that a good education stimulates a desire to learn, for instance, occurs from the abstract to the metasystematic stages. There are several possible reasons for these patterns. (1) Some propositions may fall into disuse at higher stages because their meanings have been integrated into new constructions such that the root meanings are no longer interesting or attractive. Perhaps, for example, a child who connects learning as a good with certain *kinds* of play (at the abstract level) is less likely to assert that playing, in and of itself, is what makes an education good. (2) Propositions that continue to appear in explicit form across a range of stages may do so because they continue to be useful as a means for explaining the meanings of new constructions. For example, one metasystematic performer (a political analyst) illustrated all of his more complex assertions with a series of concrete examples. When I later asked him about this pattern in his performance, he explained that in his work the art of communicating complex ideas often requires extensive illustration with examples that can be understood by people with a range of reasoning styles. (3) Some propositions may occur across several stages because they represent strands of meaning akin to the *root metaphors* employed by Lakoff (1994). This phenomenon is traced at some length in the following analyses. (4) Another possibility is related to the fact that one of the requirements of the coding process was that propositions were to be coded into a given category only when *explicitly* present. Failure to take into account implicit meanings may mask the presence of some concepts. To remedy this (and to transform the data into a form that could be modeled quantitatively) the data were recoded to include implicit meanings as described below. Finally, (5) some propositions may occur across

multiple stages because of a failure, during the coding process, to distinguish subtle differences in meaning. It is probable, for instance, that the assertion that a good education stimulates a desire to learn does not carry precisely the same meaning for abstract performers that it does for metasystematic performers.

A second pattern that emerges upon examination of the distribution of propositions is that some propositions occur more often than others. They are, in a sense, more "popular." Some of the propositions in Figures 1 and 6 were exhibited in only one performance, while others were exhibited in up to 28. There are several possible explanations for this: (1) The more populated categories may be more central to defining the construct; (2) cultural or informational factors may influence what aspects of a thread of meaning have more salience; or (3) the interviewer may have inadvertently influenced performances by emphasizing some aspects of a construct over others in follow-up questions.

The graphic representations of the data in Figures 1 and 6 do not, in and of themselves, take us much further than these observations, however. We can neither determine from this level of analysis how propositions might be related across stages, nor whether the propositions selected actually contribute to delineating the strands of meaning we're attempting to describe.

Proposition recoding

To investigate how particular meanings unfold across stages, a second layer of analysis was conducted. First, the pool of propositions for each of the metaphorical strands was reduced by collapsing the most specific propositions occurring at a

given stage into more general propositions that share much of the same meaning. For example, “A good education is one in which teachers give parties” and “A good education includes games/projects was,” were collapsed into a new category “A good education includes playing games/doing projects/doing fun things.”

Second, beginning with propositions found at the metasystematic stage, each proposition was analyzed to determine which propositions in the list, at the same stage or lower stages, were necessary to construct the higher order proposition. For example, a concept of fun is required to construct the formal proposition, “A good education is one in which learning is fun/enjoyable/entertaining.” Consequently, all respondents who employed this proposition were also given credit for the proposition, “A good education is fun.”[3] These “imputed observations” are included in the data displayed in Figures 7 and 9.

Education as play

Rasch analysis 1

Figure 7 shows the results of collapsing some of the proposition classes and including imputed observations for *education as play.* Note that now there is only one empty cell to the right of the diagonal, the metasystematic level of “A good education is one in which teachers are nice/caring/not too strict.” The grid has been filled in, indicating that there is, in fact, continuity of meaning across stages even though new constructions appear at each stage. However, we are left wondering how much continuity there should be. Is there enough to consider all of the propositions here as contributing to the same latent variable, respondents’

understanding of *education as play*?

To pursue this question, the imputed data from the second layer of the content analysis were submitted to a Rasch analysis. The imputed data are treated in this analysis as though they were directly observed. Each proposition was treated as an item, scored 1 for present and 0 for not present. The first time the analysis was run, all of the items were included. The item fit statistics indicated that several of the items were not working with the other items to describe the same thread of meaning (infit $z > 2.0$). These included: a good education... (a) is based on student likes/preferences, (b) includes friends/friendship, (c) is one in which teachers are nice/caring/not too strict, (d) is one in which teachers are understanding/open/ listen, (e) requires interested/ motivated/curious/students, (f) is one that stimulates curiosity/interest/desire to learn, (g) is not frightening, (h) is one in which people are nice/get along, and (i) supports learning for the sake of learning.

Recalling that the metaphor under investigation is *education as play,* it is not too difficult to see how these propositions, though they may partake of this metaphor, might include other meanings. First, for example, the idea that education should be based on students' educational preferences (a) incorporated, in several interviews, the notion that following one's preferences is enjoyable and engaging. However educational preferences were more often associated in the interviews with the pursuit of a self selected career. Second, though friends (b), may make education more fun, friends are valued for other reasons, such as the social development of students. Similarly, nice or understanding teachers and schoolmates (c, d, & h) may

make learning more pleasurable, but some respondents wanted nice teachers or schoolmates for reasons perhaps peripheral to learning and playfulness, such as getting good grades.

It is also interesting, but not too surprising, that the only propositions in the list that refer to the qualities of students (e, f) do not fit this particular construct. *Education as play* focuses on a quality of the learning experience, not the participants involved. Finally, the negatively worded proposition, "A good education is not frightening," (g) is ambiguous. A lack of fear does not automatically imply enjoyment or playfulness.

The failure of learning for the sake of learning (i) to work with the remaining items in the scale is more troubling than the failure of items (a) through (h). The idea of learning for its own sake, on interpretive grounds, seems like a logical consequence of defining learning in terms of playfulness. Yet this proposition fits the model poorly. A closer examination of Figure 1 reveals that this proposition was explicitly present in only 5 of the interviews, and it only shows up in 6 interviews in Figure 7, after the addition of implicitly coded cases. Perhaps it does not occur frequently enough (isn't popular enough in this sample) to register consistently on the continuum being described by the other items, though additional research on the issue would be required for more decisive rejection.

Neither model misfit nor the interpretive analysis of items are, on their own, adequate bases for determining whether items in a scale contribute to the same strand of meaning. Items that are included on strong interpretive grounds can misfit

for reasons unrelated to their meaning. Similarly, it is possible that some items fit the model well despite some murkiness in their meaning. For example, "A good education includes social interaction," which fits the model, partakes of strands of meaning other than *education as play,* including *education as preparation for social life*. Here, information about model fit and interpretation are employed in conjunction to determine which propositions do a good job of contributing to *education as play*. Though this process is subject to error at every level of analysis and interpretation, the fact that multiple propositions are included in the delineation of *education as play* is a buffer against the loss of important elements of meaning. With the possible exception of *learning for the sake of learning* (i), the notions of enjoyment, engagement, and playfulness that were thought to be present in the misfitting items are present in other items (with good fit) found at the same developmental stages.

Rasch analysis 2

Propositions (a) through (h) were eliminated from the final Rasch analysis because of their failure, on both statistical and interpretive grounds, to work with the remaining items to delineate the latent variable, education as play. Proposition i was also eliminated because doing so significantly improved the reliability of the case estimates. (Reliability for case estimates for the analysis that eliminated only items a through h was .84. It went up to .88 following the elimination of item i.)

A Rasch analysis was conducted with Bigsteps (Linacre & Wright, 1997) on the reduced dataset. The proportion of the observed item estimate variance considered

to be true (reliability of item estimates) was .98 (item separation = 7.41). (See (Wright & Masters, 1982) for more about reliability of item estimates.). All items (Table 10) have acceptable infit (weighted) and outfit (unweighted) statistics ($z < 2.0$) (Wright & Masters, 1982), indicating that they all belong fit the model adequately. One proposition, "A good education is one in which testing is an opportunity to show off learning," that fit the model before misfitting items were deleted, had to be dropped from the analysis because the one performance that included it became a perfect case (scored 1 on every remaining proposition). The interview in question was, ironically, part of the original inspiration for exploring the metaphor, education as play. In this interview, play, in the form of a dialectic between teacher and learner, knowable and known, was a dominant theme.

Table 10

Item Statistics: Education as Play

95 Persons 14 Items

ENTRY NUMBR	RAW SCORE	COUNT	MEASUR	REALSE	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTBIS	ITEMS
1			DELETED							likes/prefer
2	68	95	27.4	1.4	.99	-.1	1.09	.0	.49	learning fun
3			DELETED							friends
4			DELETED							teachers nice
5			DELETED							teachers under
6	24	95	45.0	1.6	1.21	1.2	1.47	.0	.46	involv/engag
7			DELETED							interested stud
8	69	94	26.6	1.5	1.04	.2	.89	.0	.48	includes play
9	8	95	73.0	11.3	.01	-1.0	.00	-.1	.60	convers/discuss
10	52	95	33.9	1.3	.69	-2.2	.44	.0	.66	games/fun
11			DELETED							stimulat curios
12	27	94	43.6	1.4	.94	-.4	.59	.0	.57	social interac
13	43	95	37.3	1.3	.75	-2.0	.58	.0	.64	active/learning
14			DELETED							not frightening
15			DELETED							learn own sake
16	11	95	55.6	2.9	1.03	.1	.37	-.1	.59	group activities
17	14	95	51.7	2.4	1.27	.8	2.67	.1	.46	stdnts question
18	1	95	117.4	5.5	.97	-.1	.03	.0	.21	involves test
19			DROPPED							show off learn
20	65	95	28.7	1.6	1.35	1.9	1.25	.0	.41	interesting
21	53	95	33.6	1.3	.58	-3.2	.38	-.1	.69	through play
22	89	95	14.8	2.3	1.18	.6	2.73	.1	.13	have fun
23			DELETED							people are nice
24	2	95	111.4	5.3	.90	-.2	.03	.0	.31	dialectical
MEAN	38.	95.	50.0	2.9	.92	-.3	.90	.0		
S.D.	28.	0.	29.7	2.7	.33	1.3	.86	.1		

As shown in Figure 8, the propositions order neatly by the order of hierarchical complexity of performances in which they first appeared. On the left of the figure is the logit scale. Relative distances between propositions (to the right of the logit scale) reflect their relative difficulties. In the present case, an examination of response patterns reveals that at least two factors appear to influence the relative difficulty of the propositions. One of these is order of hierarchical complexity, the

other is the relative "popularity" of propositions across stages (literally, the number of times they appeared either implicitly or explicitly in performances). The relative popularity of items does not confound the between-stage order of items, but it does have within-stage effects. "A good education is one in which students have fun," is easier than "A good education includes play," because it was identified both explicitly and implicitly in more performances. One possible interpretation of this pattern is that informational, cultural, and situational factors have an impact on the content of particular propositions constructed at a given stage, but not on the order in which more global *ways* of thinking about a given construct emerge. Another possible interpretation is that some propositions within a given stage-range appear "harder" because they are more difficult to construct at that stage (or are present only because they are being parrotted) than other propositions in the same stage-range.

Hierarchical complexity and education as play

The categories of description shown on the right of Figure 8 describe the way in which conceptions at each order of hierarchical complexity integrate constructions available at the previous order, such that a single strand of meaning is preserved *and* transformed across orders. These descriptions were constructed by examining the propositional content associated with each order in light of the organizing principles of that order, and integrating the results into general descriptions of the new understandings of *education as play* that result from these coordinations. This process was guided by the principles of task analysis described by Commons and

his colleagues (Commons et al., in press) and discussed here in Chapters 3 and 4.

New propositions occur at each order of hierarchical complexity. At the concrete level, two propositions are found that relate to *education as play*. These are that a good education: (1) is one in which students have fun, and (2) includes play. At the concrete order of the General Model of Hierarchical Complexity, performances are characterized by the ability to manipulate concrete variables. Abstract words like education and fun are in the concrete operational child's vocabulary, but their meanings are undifferentiated from concrete instances. This means that for concrete operational children, education is synonymous with school. Similarly, having fun and playing are synonymous. The meaning of the propositions that a good education: (1) is one in which students have fun, and (2) includes play must be viewed in this context. What the concrete-operational child means is that you should get to play (have fun) at school. Concrete reasoners can view school as a place to play and/or a place to learn things, but cannot coordinate play and learning into the idea of learning through play. This notion does not emerge until the abstract stage.

Abstract performances are characterized by the coordination of concrete operations into new, abstract variables and propositions. The propositions found at the abstract order include assertions that a good education (3) is one in which learning is fun, (4) is one in which subjects/teachers are interesting, (5) includes learning through play, and (6) includes playing games/doing fun things. Not only are notions of play (or fun) and learning integrated, a new concept, *interest,* is present. Play and fun

are no longer synonymous. Fun has been abstracted from its original context (play as a concrete activity) and is now imbued with a more generalized meaning. Being interested, in the sense of not being bored, is one of the things that makes things fun, so learning can be fun by being interesting (or not boring).

Formal performances are characterized by the coordination of abstract operations into formal propositions. Abstract categories are increasingly differentiated, enabling the formal performer to specify sets of abstract qualities. These are seen as exclusive categories, with different applications. The propositions first identified at the formal order include the assertions that a good education (7) includes active/experiential learning, (8) includes social interaction, (9) is stimulating/involving/engaging, and (10) is one in which students are encouraged to ask questions. Student absorption into the activity of learning is the central theme at this stage. Interest is differentiated into stimulation, involvement, inspiration, and engagement. These are seen as things that can be given to the student by good teachers. They are not yet viewed as inherent in the learning process.

Systematic performances are characterized by the coordination of formal propositions into systems of relations. Systematic performers understand that the actions that produce a particular outcome are causally related, and they are able to coordinate these actions into a system. At the systematic stage, the subject can see a network of interactions among actions, variables, and individuals within the system and is capable of recognizing that there is more than one possible system, but doesn't yet compare or integrate multiple systems. The systematic order adds

two new propositions, that a good education (11) includes group activities, and (12) includes conversation/discussion. Student absorption into the activity of learning is no longer viewed as a perk provided by exceptional teachers, but inherent in the learning process. The notion that good education involves being absorbed into the process of learning is now coordinated with the idea that ideal education takes place in social contexts. These contexts are inherently playful (enjoyable, engaging), in that they involve interaction.

Metasystematic performances are characterized by the coordination of multiple systems of operations. At the systematic order, two or more systems of formal relations are coordinated through a deliberately dialogical process, which takes into account multiple perspectives. New propositions at this stage are that a good education (13) is one in which teaching involves constant testing, and (14) requires a dialectical engagement with the learning process. At the metasystematic order, respondents take a dialectical approach to the notion that ideal learning is an inherently pleasurable activity that takes place in social interactions. Learning is now viewed as discursive, involving a continuous spiral of feedback between teacher and learner. This dialectic produces positive consequences for individuals and for humanity as a whole:

Education as preparation for the good life

Rasch analysis

Figure 9 shows the explicit and imputed observations for *education as preparation for the good life.* Note that the grid has been filled in, as occurred

above in the analysis of *education as play,* indicating that there is continuity

of meaning across stages. However, it remains unclear whether there is enough continuity to consider all of the propositions as contributing to the same latent variable, respondents' understanding of *education as preparation for the good life*?

In order to pursue this question, the explicit and imputed data from the second layer of the content analysis were submitted to a Rasch analysis. The imputed data are, once again, treated as though they were directly observed. Each proposition is treated as an item, scored 1 for present and 0 for not present.

A Rasch analysis was conducted with Bigsteps (Linacre & Wright, 1997). The proportion of the of the observed item estimate variance considered to be true (reliability of item estimates) was .99 (item separation = 8.56). All items (Table 11) have acceptable infit (weighted) and outfit (unweighted) statistics ($z < 2.0$) (Wright & Masters, 1982), indicating that they fit the model adequately.

Table 11

Item Statistics: Education as Preparation for the Good Life

139 Persons 22 Items

ENTRY NUMBR	RAW SCORE	COUNT	MEASUR	REALSE	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTBIS CORR	.ITEMS
17	5	139	70.2	3.1	1.45	1.0	.41	.0	.32	good citizens
20	7	139	67.7	2.7	1.49	1.2	.60	.0	.33	autonomy
22	7	139	67.7	2.2	.53	-1.7	.10	.0	.52	is empowering
21	8	139	66.7	2.1	.68	-1.1	.16	.0	.52	social justice
18	20	139	59.1	1.5	1.12	.7	.77	.0	.52	increases choices
19	23	139	57.8	1.4	.63	-2.5	.25	-.1	.72	effective people
15	23	134	57.7	1.4	.98	-.1	1.44	.0	.57	enhances self-worth
7	24	139	57.4	1.3	.87	-.8	.49	-.1	.62	gratification
9	27	139	56.2	1.4	1.09	.6	.59	.0	.57	function in society
10	29	139	55.5	1.5	1.35	2.1	1.03	.0	.46	advances society
16	31	139	54.8	1.3	1.05	.3	.67	.0	.59	personal fulfillment
6	30	136	54.2	1.4	1.31	1.9	.89	.0	.45	rounds out student
14	35	139	53.5	1.3	1.17	1.2	1.01	.0	.54	success
1	48	139	49.7	1.1	1.06	.5	.93	.0	.60	make a living
12	57	139	47.3	1.1	.94	-.5	.75	-.1	.64	personal goals
2	59	139	46.8	1.1	.93	-.6	2.07	.2	.64	contribute
13	63	139	45.8	1.2	1.23	1.9	.93	.0	.54	betters life
8	67	139	44.8	1.1	1.11	1.0	.85	-.1	.56	better people
11	69	139	44.2	1.1	.64	-3.6	.37	-.3	.71	survival skills
4	75	139	42.7	1.1	.80	-1.9	.55	-.2	.63	social skills
5	131	139	9.2	3.4	.09	-2.6	.02	-.1	.26	work skills/knowledge
3	138	139	-8.8	5.5	1.27	.3	7.00	.1	.07	work/a job
MEAN	44	139.	50.0	1.8	.99	-.1	.99	.0		
S.D.	35	1.	17.8	1.0	.33	1.5	1.38	.1		

As shown in Figure 10, the propositions order neatly by the order of hierarchical complexity of performances in which they first appeared. On the left of the figure is the logit scale. Relative distances between propositions (to the right of the logit scale) approximate their relative difficulties (for display reasons, some items are not shown at precisely accurate locations on the scale, but are never more than one line away from their accurate locations). As was the case with *education as play,* at least two factors appear to

influence the relative difficulty of the propositions. One of these is order of hierarchical complexity, the other is the relative "popularity" of propositions across stages.

Hierarchical complexity and education as preparation for the good life

The categories of description shown on the right of Figure 10 describe the way in which conceptions at each order of hierarchical complexity integrate constructions available at the previous order. Propositions occur at each order of hierarchical complexity with the exception of the concrete level. Because there were so few concrete performances in this sample, it is not possible to say whether the absence of any concepts related to this construct is due to the small sample size or a complete absence of the notion, at the concrete level, that education is preparation. It seems quite possible that concrete thinkers could at least parrot the often repeated statement that going to school prepares students to get a job, and it seems plausible that they would understand that specific skills, such as reading, are needed in jobs[4]. Further, as conversation with almost any six-year-old will substantiate, young concrete operational children clearly have a beginning notion that they will grow up to "be" something. Given these observations, it seems that further research with young children is needed to determine whether the idea of education as preparation for aspects of future life is constructed at the concrete order.

The propositions found at the abstract order include assertions that a good education (1) prepares you for work/a job, and (2) provides life/work skills/

knowledge. Abstract reasoners clearly understand that doing a job or living an adult life require skills that can be learned at school. They also understand the connection between having a job and taking care of basic personal needs.

The increasing differentiation and coordination of abstract concepts at the formal order produces several new propositions. A good education (3) prepares you to succeed, (4) prepares you to make a living (5) prepares you to achieve personal goals, (6) enhances/enriches/betters your life, (7) prepares students to contribute, (8) produces better/respectable/responsible people, (9) teaches survival skills, and (10) provides interpersonal/social skills. The notion established at the abstract order, that education provides skills for adult life, is further elaborated. Instead of referring to specific, concrete skills, the formal performer focuses on categories of knowledge (or skills), such as those required for effective social interaction, basic survival, and goal attainment. These categories of knowledge perform specific functions in bringing about the good life. The acquisition of a range of skills and knowledge is important because it leads to a good life in which individuals are prepared to be good people who are able to contribute to their own well-being and the well-being of significant others (such as in supporting one's family).

At the systematic order, the categories of knowledge (now often spoken of in terms of understanding) operated on at the formal stage are integrated into a system of knowledge in which relationships among different forms of knowledge are coordinated. For example, social, academic, and practical learning are coordinated in constructing the proposition that (11) a good education should round out the

student.

Psychological well-being and material well-being are similarly coordinated in assertions like: A good education (12) produces effective/competent/self-reliant people, (13) enhances self-worth/self esteem/confidence/respect. Though the concept of self-esteem appears at the formal order, it is employed only with reference to students' need to feel good about themselves in order to learn. Not until the systematic stage is this concept coordinated with the idea that education is preparation for a future good life in which one's psychological well-being is of central importance.

The enhancement of practical and psychological well-being are strong themes at the systematic order. A good education (14) increases choices/opens doors, (15) leads to personal fulfillment/meaning, and (16) increases gratification/satisfaction.

The assertions that a good education (17) prepares students to function in society and (18) improves or advances society appear for the first time at this order. Personal well-being and the notion of making a contribution to society are seen as complementary. For example, respondent 90, above, asserts that building self-esteem in individuals has societal benefits, in that individuals who confidently pursue their interests improve society. Other systematic performances link the idea of contributing to society with the promotion of personal psychological and material well-being.

The coordination of self-interest and societal interests at the systematic order is characterized by the idea that a good education prepares one to choose work that

is both personally fulfilling and makes a contribution. The conception that making a contribution can be inherently fulfilling, in that it contributes to your self-esteem, is also constructed at this stage. But the self-system and the social system are not yet fully coordinated, in the sense that the interpenetration of societal and individual interests is not yet fully articulated. This occurs at the metasystematic order, in which multiple systems of operations, in this case the self-system and the social system, are coordinated through a deliberately dialectical process. These systems are compared with respect to universal principles. The metasystematic performer thinks about the aims of education in terms of optimizing human life. To do this it is necessary to attend to the perspective of the social system as well as the perspective of the individual—to coordinate the system of education for individual fulfillment with the system of education for the advancement of society under the metasystem of education that optimizes human life.

The propositions related to education as preparation for the good life that appear for the first time at the metasystematic stage are that a good education (19) promotes autonomy/self-actualization/self-examination, (20) is empowering, (21) promotes social justice/welfare, and (22) produces good citizens.

In metasystematic performances, personal well-being and societal well-being are never spoken of without reference to their interdependence. The primary aim of a good education is to promote the development of an optimal person, one whose aims, abilities, and satisfaction are coincident with the ultimate well-being of a dynamic and healthy social system.

The goal of this chapter was to analyze and describe continuities and discontinuities in meaning from one developmental stage to another by conducting a detailed investigation of the development of meaning along two metaphorical strands, *education as play*, and *education as preparation for the good life*. A relational (iterative, dialectical) methodology was employed to investigate the relationship between particular meanings and the hierarchical complexity of performances. To make this possible, I first distinguished (in Chapter 4) between the hierarchical complexity of interview performances and their propositional content by coding these two aspects of performance separately. In Chapter 6, I conducted an analysis of the relationships among orders of hierarchical complexity and produced ability estimates for 153 respondents with a partial credit analysis. These ability estimates (measures) were employed to sort the propositional content of performances by the hierarchical complexity of interviews in which they were identified. Then, an interpretive analysis was conducted to examine the relationships among propositions thought to contribute to *education as play* and *education as preparation for the good life* at different orders of hierarchical complexity and to give individuals "credit" for both the explicit and implicit presence of concepts in their performances. The resulting data was submitted to a Rasch analysis to see how well propositions at different orders of hierarchical complexity worked together to delineate the same strand of metaphorical meaning. The results of this analysis suggested the elimination of several propositions from

the *education as play* strand, but left the *education as preparation for the good life* strand intact. The remaining propositions in these metaphorical strands were then submitted to further interpretive analysis, reintegrating the meanings generated at each order of hierarchical complexity with the principles of that order in categories of description. Finally, these categories of description were compared interpretively to analogous descriptions of reasoning in two stage sequences and three educational reasoning sequences.

One logical direction for future research is to conduct a check on the last analysis by employing the present research method to analyze the interview data upon which the other sequences are based. This would make it possible to determine whether the categories of description constructed with this method are, in fact, similar to the stage descriptions developed with other methods. It would also, given common participants across studies, facilitate the direct comparison of reasoning across domains. This work is currently underway.

A second direction for future research is to complete the analysis of the propositional content of the interviews collected for this project. Identifying several strands of meaning and submitting them to a rigorous analysis will make it possible to look more closely how individual strands develop. It will also make it possible to examine how meanings are shared across strands to produce new meanings, as occurred when a concept of learning became incorporated into the education as play strand at the abstract level, or when a concept of the social value of education became part of its value as an aspect of the good life at the systematic level.

Completion of this phase of the analysis will also make it possible to examine sex differences in the conceptual content of performances as their hierarchical complexity increases.

Chapter 8

Discussion and directions for future research

Good education is wanting to learn more.

—Student, female, 15

Invariant sequence and *structure d'ensemble*

An elementary analysis of the hierarchical complexity of performances supported postulates of invariant sequence and *structure d'ensemble*. Both were demonstrated by the tendency for individuals to perform at only one stage, or two adjacent stages. *Structure d'ensemble* was further supported by the partial credit analysis through the distribution of respondents' measures. Though these are strung out along the entire continuum of development, they cluster in a fashion that suggests individuals tend to construct most of their arguments at the highest order of hierarchical complexity available to them.

Moreover, these patterns of performance hold for all orders, including those rarely identified prior to adulthood, suggesting that stage development is a coherent process throughout the lifespan. This brings into question theories that strongly link developmental stage change to physiological changes (e.g., Epstein, 1986).

Age and education effects

Both age and education predict developmental stage as assessed with the GSSS. The relationship between age and stage is strong in childhood, becoming weaker

and less deterministic over the course of the lifespan, as suggested by the curvilinear, fan-shaped distribution in Figure 4. This pattern suggests one of three possibilities: (1) that the rate of development slows with advancing age, (2) that the aging process causes regression in stage, or (3) that the older participants in this sample have lower scores due to generational effects. Because some of the data collected for this project were longitudinal, it will be feasible in the future to examine whether the rate of development slows with age as well as the possibility of stage regression with age.

The distribution of educational attainment by stage is fan-shaped (Figure 5), indicating that this relationship, like the relationship between age and stage, becomes weaker as the number of years of education increases. However, the relationship between educational attainment and developmental stage is linear, not curvilinear as is the case with age and developmental stage. The expected advancement in developmental stage for each year of education remains constant at about one stage for every six years of education. If stage advancement became more difficult as we progress along the developmental continuum, we would expect to find a curvilinear relationship between educational attainment and stage. A linear relationship indicates the opposite—that the task demands of moving from one stage to the next are in some way the same across the entire spectrum of development.

Sex differences

The much-debated relationship between sex and developmental stage hinges on

the interpretation of small effect sizes like the ones seen here. Sex and stage are weakly related, but this relationship diminishes when age is taken into account and disappears when education is taken into account. The latter result suggests that sex differences in stage development are contextual rather than innate. However, even though there is no significant effect of sex after education is taken into account, the divergence of the slopes for males and females shown in Figure 5 remains somewhat disturbing. Are life opportunities for men and women so different that they support developmental progress differently? This question is rarely posed in the gender difference literature, which focuses more on the possibility of different "ways of knowing" than on the possibility of uneqitable social practices (e.g., Baumrind, 1986; Chodorow, 1990; Eisenberg, Fabes & Shea, 1989; Gilligan, 1982; Gilligan, Ward, Taylor & Bardige, 1988; Haste & Baddeley, 1991; Magolda, 1989; Okin, 1990).

Though I expect to find sex differences in the conceptual content of the interviews collected for this project, these have not yet been systematically examined. Sex differences in conceptual content will be studied once the content analysis is complete.

Content and structure

The attempt to differentiate and then reintegrate the conceptual and structural features of reasoning, as they have been defined here, was successful. Much was learned about these constructs. First, the independent analysis of the hierarchical complexity of performances revealed a markedly stage-like pattern. Analysis of

their propositional content, however, did not expose the same developmental pattern. Though new propositions emerged at each order of hierarchical complexity, the distribution of their difficulties was more continuous (with no systematic gaps between groups of propositions) than the distribution of stage performances. This indicated that though these aspects of performance are interrelated, they are not identical in what they have to teach us about development.

Second, though it was found that the hierarchical complexity of interview performances predicts, to a limited extent, their conceptual content, the inverse is less true. Propositional content, in and of itself, is not a good predictor of stage. This is partially because many propositions that occur at lower orders also appear at higher orders.

Third, though propositional content and developmental stage do not stand in a one-to-one relation, propositions that first appear at a given order can be reintegrated with the organizing principles of that order, creating general descriptive categories structurally analogous to the stage descriptions and definitions developed by other researchers.

The results of this research have several implications for developmental research in socio-moral domains. If it is possible to meaningfully employ a single stage scoring system across a range of texts from a variety of domains, then the problem of comparing development across domains is solved. As noted in Chapter 5, I have already produced evidence (Dawson, 1998) that the GSSS assesses the same

dimension of reasoning performance as the SISS and the GLSS. All three assess hierarchical complexity or stage. Combined with the evidence that descriptive categories analogous to stage descriptions in other systems can be produced by reintegrating the propositional content of texts with the organizing principles of the stage at which they occur, this finding has profound implications. First, developmental progress across domains can be sensibly compared; second, relationships between the propositional content of constructions at the same stage in different domains can be examined; and third, stage-scoring itself is simplified, with only one system to learn and no need to develop a scoring manual for each new domain investigated.

What is a good education?

So far, the answer to this question is incomplete and tentative. The existing literature provides some clues to the ways in which individuals value education, and the present analysis of two metaphorical strands gives some idea of trends. First, there are numerous recurrent themes in the literature, beginning with the idea that there are multiple conceptions of good education, as in the traditional and progressive orientations described by (Peterson, 1933). However, the attempt made by Peterson and his colleagues to identify coherent and distinct conceptions was unsuccessful in the sense that replications of Peterson's original work did not yield the same typologies. To anyone who has studied concept formation from a cognitive developmental point-of-view, this is not surprising. When developmental changes are ignored, their effects on concept formation can easily obscure

conceptual coherence. The developmental studies examined in Chapter 2 demonstrate that conceptualizations alter radically over the course of development, and as demonstrated in the present research, these changes are strongly associated with differences in the hierarchical complexity of performances.

In the metaphorical strands identified and studied in this project, each order of hierarchical complexity is associated with both increasing differentiation in the content brought to bear in individual propositions, and the increasing integration of this content to produce new meanings. For example, whereas the concrete conception of *education as play* does not yet link play with learning, at the abstract level learning and play are integrated into the idea that learning is a potentially fun activity. The metaphorical strand *education as learning* has not yet been examined, but it is clear from this result that at the abstract stage, individuals have a conception of learning that is generalized enough to include evaluations of different learning experiences. At each stage, concepts from previously unobserved (with respect to *education as play*) strands of meaning appear in performances. For example, at the formal stage *education as doing* is incorporated, at the systematic stage *education as communication* is incorporated, and at the metasystematic stage *education as testing* is incorporated.

To construct a more complete understanding of the development of evaluative reasoning about education, these and other metaphorical strands must be identified and evaluated with particular attention to their interrelationships. Such an analysis should reveal not only the way in which individual strands of meaning unfold

across orders of hierarchical complexity, but should make it possible to identify the informational and conceptual content required to produce particular constructions. At present, I envision a network such as that rather simplistically depicted in Figure 11. If the kind of relations suggested in the figure hold true across metaphorical strands, a developmental account of reasoning will have a web-like quality, integrating an increasing number of metaphorical strands into conceptions within each strand as reasoning increases in complexity.

One of the implications of such a web of dependencies is that reasoning at a higher order of complexity on one metaphorical strand is contingent upon the development of concepts on other strands. The absence of certain conceptions could temporarily halt development while reasoning along related strands "catches up." Could this explain, in part, why so few individuals in this study demonstrated only a small amount of reasoning at their highest stage, making it look as though we leap from one stage to the next? Note how the distribution of measures within stages form right triangles, with the wide edge at the top and the point at the bottom in Figure 3. This suggests that stage transitions seem to require integration of a sufficient number of strands before reasoning can advance to another level of hierarchical complexity. That is, people advance as far as they can within a stage with only a few strands, and if they never have the opportunity to work in more strands, they never make the next stage. When they do work in enough strands (How many is enough?), this precipitates a developmental leap because the complexity managed increases by an order of magnitude, not simply incrementally.

Importance

I have already made several comments on the scientific importance of this work, but have not yet addressed its practical educational importance. The most immediate advantage of this research is what has been and will be learned about the development of evaluative reasoning about education. This has many implications for educational policy and practice. Knowledge of this kind can be used (1) to look at how the educational conceptions of individuals impact learning in different environments; (2) to improve teacher training by providing more detailed knowledge of the content and forms of pedagogical thought associated with high-quality teaching practice, (3) to specifically inform the development of adult education programs, and (4) as a part of the knowledge brought to bear in forming educational policy.

A more general practical contribution of this research is methodological. First, it introduces a means for investigating the development of conceptions in any domain that permits direct comparisons across domains. Because the GSSS can be employed to score in any content domain, it provides a general means for evaluating stage. Combined with a content analysis of conceptions at each stage, the results can be used to develop open-ended domain-specific descriptions of reasoning at each stage. Comparisons of developmental progress across domains is made possible, as is the comparison of the conceptual content of multiple domains within stages.

Another methodological contribution is in the use of measurement technology as a

check on interpretive findings. Though sophisticated measurement technology has been employed in test construction for decades, it has been used less frequently in interpretive research. This project demonstrates that it is possible to increase confidence in interpretive findings by double-checking these with appropriate quantitative tools.

Finally, the CSQ software, *Sequencer,* developed for this project has broad practical applicability. Researchers who work with large numbers of variables, particularly those who conduct content analyses, including cognitive, ethnographic, and educational researchers, find graphic representations of data especially valuable. Sequencer provides a means for visually inspecting the distribution of large sets of dependent variables (concepts, propositions, words) in their relation to any independent variable. CSQs can be used either descriptively or as a means for examining data prior to other interpretation and analysis.

References

Ames, C., & Ames, R. (1984). Goal structures and motivation. Elementary School Journal, 85, 39-52.

Ammon, P. (1984). Human development, teaching, and teacher education. Education Quarterly, 11, 95-108.

Ammon, P., & Hutcheson, B. P. (1989). Promoting the development of teacher's pedagogical conceptions. The Genetic Epistemologist, 17, 23-29.

Ammon, P., Hutcheson, B. P., & Black, A. (1985, April). Teacher's developing conceptions about children, learning, and teaching: Observations from a clinical interview. Paper presented at the American Educational Research Association, Chicago.

Ammon, P., & Levin, B. B. (1993). Expertise in teaching from a developmental perspective: The developmental teacher education program at Berkeley. Learning and Individual Differences, 5, 319-326.

Ammon, P., & Peretti, D. (1995). Preparing constructivist teachers for parent involvement: The developmental teacher education program at Berkeley . Berkeley, CA: University of California at Berkeley.

Andrich, D. (1986, September). Intellectual development of pre-adolescent and adolescent children from a psychometric perspective. Paper presented at the International Conference on Longitudinal Methodology, Budapest, Hungary.

Andrich, D. (1988). Rasch models for measurement. Newbury Park, CA: Sage Publications.

Andrich, D. (1989). Distinctions between assumptions and requirements in measurement in the social sciences. In J. A. Keats, R. Taft, R. A. Heath, & S. H. Lovibond (Eds.), Mathematical and theoretical systems, (pp. 7-16). North-Holland: Amsterdam, Netherlands.

Andrich, D., & Styles, I. (1994). Psychometric evidence of intellectual growth spurts in early adolescence. Journal of Early Adolescence, 14, 328-344.

Armon, C. (1984a). Ideals of the good life. In M. Commons, F. Richards, & C. Armon (Eds.), Beyond formal operations, Volume 1: Late adolescent and adult cognitive development. New York: Praeger.

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

Armon, C. (1993). Developmental conceptions of good work: A longitudinal study. In J. Demick & P. M. Miller (Eds.), Development in the workplace (pp. 21-37). Hillsdale, NJ: Lawrence Erlbaum.

Armon, C., & Dawson, T. L. (1997). Developmental trajectories in moral reasoning across the lifespan. Journal of Moral Education, 26(4).

Armon, C., with Dawson, T. L. (in prep). A longitudinal study of adult value reasoning about the good life.

Bakken, L., & Ellsworth, R. (1990). Moral development in adulthood: Its relationship to age, sex, and education. Educational Research Quarterly, 14, 2-9.

Bandura, A. (1986). Social foundations of thought and action: A social cognitive theory. Englewood Cliffs, NJ: Prentice-Hall.

Basseches, M. A. (1984). Dialectical thinking as a metasystematic form of cognitive organization. In M. L. Commons, F. A. Richards, & C. Armon (Eds.), Beyond formal operations: Vol. 1. Late adolescent and adult cognitive development (pp. 216-238). New York: Praeger.

Baumrind, D. (1986). Sex differences in moral reasoning: Response to Walker's (1984) conclusion that there are none. Child Development, 57, 511-521.

Baxter-Magolda, M. B. (1988). Measuring gender differences in intellectual development: A comparison of assessment methods. Journal of College Student Development, 29, 528-537.

Belenky, M. F., Clinchy, B. M., Goldberger, N. R., & Tarule, J. M. (1986). Women's ways of knowing: The development of self, voice, and mind. New York, NY: BasicBooks.

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

Biggs, J. B. (1984). Student motivation and strategies of learning and studying: A handbook for teachers, counselors, and researchers. Hawthorn,

Victoria: Australian Council for Educational Research.

Bird, T., Anderson, L. M., & Swidler, S. A. (1993). Pedagogical balancing acts: Attempts to influence prospective teachers' beliefs. Teaching and Teacher Education, 9, 253-267.

Black, A. (1989). Developmental teacher education. Genetic Epistemologist, 17.

Black, A., & Ammon, P. (1992). A developmental-constructivist approach to teacher education. Journal of Teacher Education, 43, 323-335.

Black, H., & Phillips, S. (1982). An intervention program for the development of empathy in student teachers. Journal of Psychology, 112, 159-168.

Blasi, A. (1987). The psychological definitions of morality. In J. Kagan & S. Lamb (Eds.), The emergence of morality in young children (pp. 83-90). Chicago, IL: The University of Chicago Press.

Bond, T., & Bunting, E. (1995). Piaget and measurement III: Reassessing the méthode clinique. Archives de Psychologie, 63, 231-255.

Bond, T. G. (1994). Piaget and measurement I: The twain really do meet. Archives de Psychologie, 63, 71-87.

Brabeck, M. (1983). Critical thinking skills and reflective judgment development: Redefining the aims of higher education. Journal of Applied Developmental Psychology, 4, 23-34.

Brainerd, C. J. (1978). The stage question in cognitive-developmental theory. Behavioral Brain Science, 2, 171-213.

Bramald, R., Hardman, F., & Leat, D. (1995). Initial teacher trainees and their views of teaching and learning. Teaching & Teacher Education, 11, 23-31.

Bunting, C. E. (1984). Dimensionality of teacher education beliefs: An exploratory study. Journal of Experimental Education, 52, 195-198.

Case, R. (1985). Intellectual development: Birth to adulthood. New York: Academic Press.

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, (Vol. 23, pp. 71-100). Basel, Switzerland: S. Karger.

Cavanaugh, J. C., & McGuire, L. C. (1994). Chaos theory as a framework for understanding adult lifespan learning. In J. D. Sinnott (Ed.), Interdisciplinary handbook of adult lifespan learning (pp. 3-21): Greenwood Press/Greenwood Publishing Group.

Chickering, A. W. (1976). Developmental change as a major outcome. In M. T. Keeton (Ed.), Experiential learning: rationale, characteristics, and assessments. San Francisco: Jossey-Bass.

Chodorow, N. J. (1990). Gender, relation, and difference in psychoanalytic

perspective. In C. Zanardi (Ed.), Essential papers on the psychology of women (pp. 420-436). New York: New York University Press.

Cleave-Hogg. (1996). Learning perspectives of older students enrolled full-time in a regular undergraduate program. In M. L. Commons, J. Demick, & C. Goldberg (Eds.), Clinical approaches to adult development. 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. (1981). Invariant sequence and internal consistency in moral judgment : Harvard Univ., Cambridge, Mass. Graduate School of Education.

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.

Colby, A., Kohlberg, L., Gibbs, J., & Lieberman, M. (Eds.). (1983). A longitudinal study of moral judgment. (Vol. 48).

Commons, M. L., Armon, C., Richards, F. A., Schrader, D. E., Farrell, E. W., Tappan, M. B., & Bauer, N. F. (1989). A multidomain study of adult development. In D. Sinnott, F. A. Richards, & C. Armon (Eds.), Adult development, Vol. 1: Comparisons and applications of developmental models (pp. 33-56). New York:

Praeger Publishers.

Commons, M. L., & Grotzer, T. A. (1990). The relationship between Piagetian and Kohlbergian stage: An examination of the "necessary but not sufficient relationship". In M. L. Commons, C. Armon, L. Kohlberg, F. A. Richards, T. A. Grotzer, & J. D. Sinnott (Eds.), Adult development, Vol. 2: Models and methods in the study of adolescent and adult thought. (pp. 205-231). New York: Praeger Publishers.

Commons, M. L., & Miller, P. M. (in press). A quantitative behavior-analytic theory of development. Mexican Journal of Experimental Analysis of Behavior.

Commons, M. L., Miller, P. M., & Kuhn, D. (1982). The relation between formal operational reasoning and academic course selection and performance among college freshman and sophomores. Journal of Applied Developmental Psychology, 3, 1-10.

Commons, M. L., Richards, F. A., with Ruf, F. J., Armstrong-Roche, M., & Bretzius, S. (1983). 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., Sonnert, G., & Rodriquez, J. (1995a, November). The general stage scoring system: How to score anything. Paper presented at the Annual meeting of the Association for Moral Education.

Commons, M. L., Straughn, J., Meaney, M., Johnstone, J., Weaver, J. H., Lichtenbaum, E., Sonnert, G., & Rodriquez, J. (1995b, November). Measuring stage of development across domains: A universal scoring scheme. Paper presented at the Annual meeting of the Association for Moral Education.

Commons, M. L., Trudeau, E. J., Stein, S. A., Richards, S. A., & Krause, S. R. (in press). Hierarchical complexity of tasks show the existence of developmental stages. Developmental Review.

Cook, W. W., Leeds, C. H., & Callis, R. (1952). Manual of the Minnesota Teacher Attitude Inventory. New York: Psychological Corporation.

Cooney, W., Cross, C., & Trunk, B. (1993). From Plato to Piaget: The greatest educational theorists from across the centuries and around the world. Lanham, Maryland: University Press of America.

Cooper, P., & McIntyre, D. (1993). Commonality in teachers' and pupils' perceptions of effective classroom learning. British Journal of Educational Psychology, 63, 381-399.

Courtney, S. (1991). Why adults learn: Toward a theory of participation in adult education. New York: Routledge, Chapman, and Hall.

Cross, K. P. (1978). The adult learner, Current issues in higher education, . Washington, DC: National conference series, American Association for Higher Education.

Damon, W. (1980). Patterns of change in children's social reasoning: A two-

year longitudinal study. Child Development, 51, 1010-1017.

Damon, W. (1989). Child development today and tomorrow. , Jossey-Bass Publishers.

Davison, M. L., King, P. M., & Kitchener, K. S. (1990). Developing reflective thinking and writing. In R. Beach & S. Hynds (Eds.), Developing discourse practices in adolescence and adulthood. Advances in discourse processes, Vol. 39 (pp. 265-286). Norwood, NJ: Ablex.

Davison, M. L., King, P. M., Kitchener, K. S., & Parker, C. A. (1980). The stage sequence concept in cognitive and social development. Developmental Psychology, 16, 121-131.

Dawson, T., & Wilson, M. (1997). The development and use of Sequencer, A tool for the management of qualitative data (Spencer Foundation Small Grants Report). Berkeley: University of California at Berkeley,.

Dawson, T. L. (1996a, September). An examination of intraindividual and interindividual developmental trends in moral and evaluative reasoning. Poster presented at the Growing Mind, Geneva, Switzerland.

Dawson, T. L. (1996b, September). Integration and differentiation in longitudinal moral development data. Paper presented at the Growing Mind, Geneva, Switzerland.

Dawson, T. L. (1996c, September). Moral reasoning and evaluative reasoning about the good life—A Rasch Analysis of Armon's 13-year lifespan

investigation. Paper presented at the 26th Annual Symposium of the Jean Piaget Society, Philadelphia, PA.

Dawson, T. L. (1996d, June). Rasch analysis of moral and evaluative reasoning across the life-span: An examination of stage transition. Paper presented at the 11th Annual Adult Development Symposium of the Society for Research in Adult Development, Boston, MA.

Dawson. (1997, June). New tools, new insights: Kohlberg's moral reasoning stages revisited. Paper presented at the Twenty-Seventh Annual Symposium of the Jean Piaget Society, Santa Monica, CA.

Dawson, T. L. (1998, June). A stage is a stage is a stage: An empirical look at a modern metaphor. Paper presented at the Annual Symposium of the Jean Piaget Society, Chicago.

Demetriou, A., & Efklides, A. (1985). Structure and sequence of formal and postformal thought: General patterns and individual differences. Child Development, 56, 1062-1091.

Demetriou, A., Efklides, A., Papadaki, M., Papantoniou, G., & Economou, A. (1993). Structure and development of causal-experimental thought: From early adolescence to youth. Developmental Psychology, 29, 480-497.

Doyle, J. F. (Ed.). (1973). Educational judgments. Boston, MA.

Draney, K. L. (1996). The polytomous Saltus model: A mixture model approach to the diagnosis of developmental differences. Unpublished doctoral

dissertation, University f California at Berkeley, Berkeley, CA.

Dweck, C. S. (1991). Self-theories and goals: Their role in motivation, personality, and development. In A. D. Richard (Ed.), Nebraska Symposium on Motivation, 1990: Perspectives on motivation. Current theory and research in motivation, Vol. 38., (pp. 199-235). Lincoln, NE: University of Nebraska Press.

Edelstein, W., Keller, M., & Wahlen, K. (1984). Structure and content in social cognition: Conceptual and empirical analyses. Child Development, 55, 1514-1526.

Eisenberg, N., Fabes, R., & Shea, C. (1989). Gender differences in empathy and prosocial moral reasoning: Empirical investigations. In M. B. Mary (Ed.), Who cares?: Theory, research, and educational implications of the ethic of care., (pp. 127-143): Praeger Publishers, New York, NY.

Elam, S. M. E. (1978). A decade of Gallup Polls of attitudes toward education,1969-1978.

Finster, D. C. (1992). New pathways for teaching chemistry: Reflective judgment in science. Liberal Education, 78, 477-531.

Fischer, K. (1980). A theory of cognitive development: The control and construction of hierarchies of skills. Psychological Review, 87, 477-531.

Fischer, K. W., & Ayoub, C. (1996). Analyzing development of working models of close relationships: Illustration with a case of vulnerability and violence. In G. G. Noam & K. W. Fischer (Eds.), Development and vulnerability in close

relationships. The Jean Piaget symposium series (pp. 173-199). Mahwah, NJ: Lawrence Erlbaum.

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

Fischer, K. W., & Silvern, L. (1985). Stages and individual differences in cognitive development. Annual Review of Psychology, 36, 613-648.

Fisher, W. P., Jr. (1992). Objectivity in measurement: A philosophical history of Rasch's scalability theorem. In M. Wilson (Ed.), Objective measurement: Theory into practice (pp. 29-60). Norwood, NJ: Ablex.

Fisher, W. P., Jr. (1994). The Rasch debate: Validity and revolution in educational measurement. In M. Wilson (Ed.), Objective measurement (pp. 36-72). Norwood, NJ: Ablex Publishing.

Fisher, W. P. Jr. (1998, June). What a mathematical language of development should look like. Paper presented at the Annual Symposium of the Jean Piaget Society, Chicago.

Flavell, J. H. (1971). Stage-related properties of cognitive development. Cognitive Psychology, 2, 421-453.

Flavell, J. H. (1992). Cognitive development: Past, present, and future. Developmental Psychology, 28, 998-1005.

Fowler, J. (1981). Stages of faith: The psychology of human development and the quest for meaning. San Francisco: Harper and Row.

Frankena, W. K. (1973). The concept of education today. In J. F. Doyle (Ed.), Educational judgments, (pp. 19-32). Boston, MA: Routledge and Paul.

Gilligan, C. (1982). In a different voice: Psychological theory and women's development. Cambridge, MA: Harvard University Press.

Gilligan, C., Ward, J. V., Taylor, J. M., & Bardige, B. (Eds.). (1988). Mapping the moral domain: A contribution of women's thinking to psychological theory and education. Cambridge, MA: Harvard University Press.

Giorgi, A. (1986). A phenomenological analysis of descriptions of concepts of learning obtained from a phenomenographic perspective. Publikationer fran institutionen för pedagogik, Göteborgs universitet, nr 18.

Goldberger, N. R., Clinchy, B. M., Belenky, M. F., & Tarule, J. M. (1987). Women's ways of knowing: On gaining a voice. In C. H. Phillip Shaver (Ed.), Sex and gender. Review of personality and social psychology, Vol. 7. (pp. 201-228). Newbury Park, CA: Sage Publications.

Goodheart, E., Dawson, T., Commons, M. (1996a, June). Primary, concrete, abstract, formal, systematic, and metasystematic operations as observed in a Piagetian logical task series. Paper presented at the Annual Meeting of the Society for reseach in Adult development, Boston.

Goodheart, E., Dawson, T., Commons, M. (1996b, November). A Rasch

analysis of developmental data on relations between more and less powerful persons. Paper presented at the Annual meeting of the Association for Moral Education.

Gruber, H. E., & Vonèche, J. (1977). The essential Piaget. London: Routledge & Kegan Paul.

Harding, L. W. (1944). A value-type generalisations test. Journal of Social Psychology, 19, 53-79.

Haste, H., & Baddeley, J. (1991). Moral theory and culture: The case of gender. In W. M. Kurtines & J. L. Gewirtz (Eds.), Handbook of moral behavior and development, Vol. 1: Theory, (pp. 223-249). Hillsdale, NJ: Lawrence Erlbaum.

Hautamäki, J. (1989). The application of a Rasch model on Piagetian measures of stages of thinking. In P. Adley (Ed.), Adolescent Development and School Science (pp. 342-349). London: Falmer.

Holstein, C. B. (1976). Irreversible, stepwise sequence in the development of moral judgment: A longitudinal study of males and females. Child Development, 47, 51-61.

Houle, C. (1961). The inquiring mind. Madison: University of Wisconsin Press.

Inhelder, B. (1982). Outlook. In S. Modgil & C. Modgil (Eds.), Jean Piaget: Consensus and controversy (pp. 411-417). New York: Praeger.

Inhelder, B., & Piaget, J. (1979). Procédures et structures. Archives de Psychologie, 47, 165-176.

Jakubowski, E., & Tobin, K. (1991). Teachers' personal epistemologies and classroom learning environments. In B. J. Fraser & H. J. Walberg (Eds.), Educational environments: Evaluation, antecedents and consequences (pp. 201-214). Oxford, England: Pergamon Press.

Jesness, C. F. (1972). Report on the youth center research project . Sacramento, CA: American Justice Institute.

Johnston, M. (1989). Moral reasoning and teachers' understanding of individualized instruction. Journal of Moral Education, 18, 45-59.

Johnston, M., & Lubomudrov, C. (1982). Teachers level of moral reasoning and their understanding of classroom rules and roles. Elementary School Journal, 88, 65-78.

Keasey, C. B. (1975). Implicators of cognitive development for moral reasoning. In D. Palma & Foley (Eds.), Moral development: Current theory and research. Mahwah, NJ: Lawrence Earlbaum.

Kegan, R. (1982). The evolving self: Problem and process in human development. Cambridge, Massachusetts: Harvard University Press.

Kegan, R. (1994). In over our heads: The mental demands of modern life. Cambridge, MA: Harvard University Press.

Keller, M., & Reuss, S. (1984). An action-theoretical reconstruction of the development of social-cognitive competence. Human Development, 27, 211-220.

Keller, M., & Wood, P. (1989). Development of friendship reasoning: A study of interindividual differences in intraindividual change. Developmental Psychology, 25, 820-826.

Kelley, P. R., & Schumacher, C. F. (1984). The Rasch model: Its use by the National Board of Medical Examiners. Evaluation and the Health Professions, 7, 443-454.

Kerlinger, F. (1956). The attitude structure of the individual: A Q study of the educational attitudes of professors and layment. Genetic Psychology Monographs, 53, 283-239.

Kerlinger, F. (1958). Progressivism and traditionalism: Basic factors of educational attitudes. Journal of Social Psychology, 48, 111-135.

Kerlinger, F. (1966). Attitudes toward education and perception of teacher characteristics: A Q study. American Educational Research Journal, 3, 159-168.

Kerlinger, F. (1967). The first- and second-order factor structures of attitudes toward education. American Educational Research Journal, 4, 191-205.

King, P., Kithchener, K., & Wood, P. (1992, Nov.). Moral and intellectual development beyond the college years: A ten-year study. Paper presented at the The Annual Meeting of the Association for Moral Eduacation, Athens, Georgia.

King, P. M., & Kitchener, K. S. (1994). Developing reflective judgment. San Francisco: Jossey Bass.

King, P. M., Kitchener, K. S., Wood, P. K., & Davison, M. L. (1989). Relationships across developmental domains: A longitudinal study of intellectual, moral, and ego development. In M. L. Commons, J. D. Sinnot, F. A. Richards, & C. Armon (Eds.), Adult development. Volume 1: Comparisons and applications of developmental models (pp. 57-71). New York: Praeger.

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., & Fischer, K. W. (1990). A skill approach to the development of reflective thinking. Contributions to Human Development, 21, 48-62.

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., Wood, P. K., & Davison, M. L. (1989). Sequentiality and consistency in the development of reflective judgment: a six-year longitudinal study. Journal of applied Developmental Psychology, 10, 73-95.

Knefelkamp, L. L., & Slepitza, R. (1976). A cognitive-developmental model

of career development: An adaptation of the Perry scheme. Counseling Psychologist, 6, 53-58.

Koch, H. L., Kentler, M., Dysart, B., & Streit, H. (1934). A scale for measuring attitude toward the question of children's freedom. Child Development, 5, 253-266.

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. (1971). From is to ought: How to commit the naturalistic fallacy and get away with it in the study of moral development. In T. Mischel (Ed.), Psychology and genetic epistemology (pp. 151-235). New York: Academic Press.

Kroll, B. M. (1992). Reflective inquiry in a college english class. Liberal Education, 78, 10-13.

Kronholm, M. M. (1994). The impact of a developmental instruction approach to environmental education at the undergraduate level on development of reflective judgment. Dissertation Abstracts International: Section B: The Sciences & Engineering, 54, 4417.

Kurfiss, J. (1976). Late adolescent development: A structural-epistemological perspective. Dissertation Abstracts International, 37, 954.

Kurfiss, J. (1977). Sequentiality and structure in a cognitive model of college student development. Developmental Psychology, 13, 565-571.

Labaree, D. F. (1997). Public goods, private goods: The American struggle over educational goals. American Educational Research Journal, 34, 39-82.

Lakoff, G. (1994). What is a conceptual system? In W. F. Overton & D. S. Palermo (Eds.), The nature and ontogenesis of meaning (pp. 41-90). Hillsdale, NJ: Lawrence Erlbaum.

Lasker, H., & Moore, J. (1980). Current studies in adult development: Implications for education, Adult development and approaches to learning (pp. 1-42). Washington, DC: U. S. Government Printing Office.

Levin, B. B., & Ammon, P. (1992). Toward an understanding of teacher development: A longitudinal case study of four teachers. Teacher Education Quarterly, 19, 19-37.

Levine, C. G. (1979). The form-content distinction in moral developmental research. Human Development, 22, 225-234.

Linacre, J. M., & Wright, B. D. (1997). A user's guide to Bigsteps. Chicago: Mesa.

Loevinger, J. (1987). Ego development. San Francisco: Jossey Bass.

Lowe, D. W., & Brock, J. (1994). Characteristics of effective graduate psychology courses: Student and faculty perspectives. Teaching of Psychology, 21, 82-85.

Lubomudrov, C. (1982). Teachers' moral development, attitudes toward

educational issues, and teaching practices. Unpublished doctoral dissertation, University of Utah.

Ludlow, L. H. (1985). A strategy for the graphical representation of Rasch model residuals. Educational & Psychological Measurement, 45, 851-859.

Magolda, M. B. (1989). Gender differences in cognitive development: An analysis of cognitive complexity and learning styles. Journal of College Student Development, 30, 213-220.

Magolda, M. B. (1990). Gender differences in epistemological development. Journal of College Student Development, 31, 555-561.

Magolda, M. B. (1992). Knowing and reasoning in college: Gender-related patterns in students' intellectual development: San Francisco: Jossey-Bass.

Marton, F. (1981). Phenomenography—describing conceptions of the world around us. Instructional Science, 10, 177-200.

Marton, F. (1994a). Phenomenography as a research approach. In T. Husén & T. N. Postlewaite (Eds.), The international encyclopedia of education (2 ed., pp. 4424-4429). New York: Pergamon Press.

Marton, F., Dall'Alba, G., & Beaty, E. (1993). Conceptions of learning. International Journal of Educational Research, 19, 277-300.

Masters, G. (1997). Fundamental measurement for outcome evaluation. In R. M. Smith (Ed.), Outcome measurement (Vol. 11, pp. 261-288). Philadelphia: Hanley & Belfus.

Masters, G. N. (1982). A Rasch model for partial credit scoring. Psychometrika, 47, 149-174.

Masters, G. N. (1988). Measurement models for ordered response categories. In R. Langeheine & J. Rost (Eds.), Latent trait and latent class models, (pp. 11-29). New York: Plenum Press.

Masters, G. N. (1994). Partial credit model. In T. Husén & T. N. Postlethwaite (Eds.), The international encyclopedia of education (pp. 4302-4307). London: Pergammon.

McIntyre, D. (1992). Theory, theorizing, and reflection in initial teacher education. In J. Calderhead (Ed.), Conceptualizing reflection in teacher development. London: Falmer Press.

Müller, U., Reene, K., & Overton, W. F. (1994, June). Rasch scaling of a deductive reasoning task. Paper presented at the Annual Symposium of the Jean Piaget Society, Philadelphia, PA.

Müller, U., Winn, M., & Overton, W. (1995, June). Rasch analysis of two recursive thinking tasks. Paper presented at the Annual Symposium of the Jean Piaget Society, Berkeley, CA.

Nevins, K. J. (1992). Understanding students as learners using student development models. Liberal Education, 1992, 24-27.

Nicholls, J. G. (1992). Students as educational theorists. In D. E. Schunk & J. L. Meece (Eds.), Student perceptions in the classroom (pp. 267-286). Hillsdale,

NJ: Lawrence Earlbaum.

Nicholls, J. G., Patashnick, M., & Nolen, S. B. (1985). Adolescents' theories of education. Journal of Educational Psychology, 77, 683-692.

Nicholls, J. G., & Thorkildsen, T. A. (1988). Children's distinctions among matters of intellectual convention, logic, fact, and personal preference. Child Development, 59, 939-949.

Nicholls, J. G., & Thorkildsen, T. A. (1989). Intellectual conventions versus matters of substance: Elementary school students as curriculum theorists. American Educational Research Journal, 26, 533-544.

Nicholls, J. G., & Thorkildsen, T. A. (Eds.). (1995). Reasons for learning: Expanding the conversation on student-teacher collaboration. New York: Teachers College Press.

Nisan, M., & Kohlberg, L. (1982). Universality and variation in moral judgment: A longitudinal and cross-sectional study in Turkey. Child Development, 53, 865-876.

Noam, G. G. (1988). A constructivist approach to developmental psychopathology. In E. D. Nannis & P. A. Cowan (Eds.), Developmental psychopathology and its treatment. New directions for child development, No. 39., (pp. 91-121). San Francisco, CA: Jossey-Bass.

Noelting, G., Coudé, G., & Rousseau, J. P. (1995, June). Rasch analysis applied to multi-domain tasks. Paper presented at the Twenty-Fifth Annual

Symposium of the Jean Piaget Society, Berkeley, CA.

Noelting, G., Rousseau, J. P., & Coudé, G. (1994, June). Rasch analysis applied to Piagetian-type problems. Paper presented at the Annual Symposium of the Jean Piaget Society, Chicago, IL.

Oja, S. N. (1990). Developmental theories and the professional development of teachers. San Francisco: Jossey Bass.

Oja, S. N. (1991). Teacher and supervisor development: A report of four studies from 1978 to 1988, Psychological intervention and human development., (pp. 111-121). Porto, Portugal: Instituto de Consulta Psicologica, Formacao e Desenvolvimento.

Oja, S. N., & Ham, M. C. (1984). A cognitive-developmental approach to collaborative action research with teachers. Teachers College Record, 86, 171-192.

Okin, S. M. (1990). Thinking like a woman. In D. L. Rhode (Ed.), Theoretical perspectives on sexual difference, (pp. 145-159). New Haven, CT: Yale University Press.

Oliver, R. A. C. (1962). Teacher's attitudes to education: The structure of educational attitudes. British Journal of Social and Clinical Psychology, 1, 56-69.

Oliver, R. A. C., & Butcher, J. H. (1968). Teacher's attitudes to education. British Journal of Educational Psychology, 38, 38-44.

Oser, F. (1994). The development of religious judgment. In P. Bill (Ed.),

Fundamental research in moral development. Moral development: A compendium, Vol. 2., (pp. 375-395). New York: Garland Publishing.

Overton, W. (in press). Developmental psychology: Philosophy, concepts, and methodology. In W. Damon (Ed.), Handbook of Child Psychology, (5 ed.,). New York: Wiley.

Pascual-Leone, J., & Goodman, D. (1979). Intelligence and experience: A neoPiagetian approach. Instructional Science, 8, 301-367.

Pepper. Multiple Intelligences.

Perry, W. G. (1970). Forms of intellectual and ethical development in the college years. New York: Holt, Rinehart, & Winston.

Peterson, F. E. (1933). Philosophies of education current in the preparation of teachers in the United States.

Piaget, J. (1970). Structuralism (Chaninah Maschler, Trans.). New York: Harper & Row.

Piaget, J. (1977a). Problems of equilibration. In H. E. Gruber & J. J. Vonéche (Eds.), The essential Piaget, (pp. 838-841). New York: Basic Books.

Piaget, J. (1985). The equilibration of cognitive structures: The central problem of intellectual development. Chicago: The University of Chicago Press.

Piaget, J., Grize, J.-B., Szeminska, A., & Vinh, B. (1977). Epistemology and psychology of functions. Boston: D. Reidel.

Powell, R. R. (1992). The influences of prior experiences on pedagogical constructs of traditional and non-traditional preservice teachers. Teaching and Teacher Education, 8, 225-238.

Pratt, M. W., Golding, G., & Hunter, W. J. (1983). Aging as ripening: Character and consistency of moral judgment in young, mature, and older adults. Human Development, 26, 277-288.

Pratte, R. (1971). Contemporary theories of education. Toronto: Intext Educational Publishers.

Rose, J. S., & Medway, F. J. (1981). Measurement of teachers' beliefs in their control over student outcome. Journal of Educational Research, 74, 185-190.

Rosenberg, S. (1988). Reason, ideology, and politics. Princeton, New Jersey: Princeton University Press.

Ryans, D. G. (1953). A statistical analysis of certain educational viewpoints held by teachers. Journal of Experiential Education, 22, 119-132.

Sälgö, R. (1979). Learning in the learner's perspective. I. Some common sense conceptions. Reports from the Department of Education, University of Göteborg, No. 76.

Sau-Ching Lam, M. (1995). Women and men scientists' notions of the good life: A developmental approach. Unpublished Doctoral dissertation, University of Massachusetts.

Schommer, M. (1993). Epistemological development and academic performance among secondary students. Journal of Educational Psychology, 85, 406-411.

Schommer, M. (1994). An emerging conceptualization of epistemological beliefs and their role in learning. In R. Garner, P. Alexander (Eds.), Beliefs about text and instruction with text (pp. 25-40): Lawrence Erlbaum Associates, Inc, Hillsdale, NJ.

Schommer, M., Calvert, C., Gariglietti, G., & Bajaj, A. (1997). The development of epistemological beliefs among secondary students: A longitudinal study. Journal of Educational Psychology, 89, 37-40.

Schommer, M., & Dunnell, P. A. (1994). A comparison of epistemological beliefs between gifted and non-gifted high school students. Roeper Review, 16, 207-210.

Schweder, R. A. (1987). Culture and moral development. In J. Kegan & S. Lamb (Eds.), The emergence of morality in young children, (pp. 1-90). Chicago: The University of Chicago Press.

Selman, R., & Damon, W. (1975). The neccessity (but insufficiency) of social perspective taking for conceptions of justice at three early levels. In D. Palma & Foley (Eds.), Moral development: Current theory and research. Mahwah, NJ: Lawrence Erlbaum.

Selman, R. L. (1971a). Conceptual role-taking development in early

childhood. Proceedings of the Annual Convention of the American Psychological Association, 6, 155-156.

Selman, R. L. (1971b). The relation of role taking to the development of moral judgment in children. Child Development, 42, 79-91.

Selman, R. L. (1977). A structural-developmental model of social cognition: Implications for intervention research. Counseling Psychologist, 6, 3-6.

Selman, R. L., & Schultz, L. H. (1990). Making a friend in youth: Developmental theory and pair therapy. Chicago, IL: University of Chicago Press.

Siegler, R. S. (1981). Developmental sequences within and between concepts. Monographs of the Society for Research in Child Development, 46, 84.

Simpson, E. L. (1976). A holistic approach to moral development and behavior. In T. Lickona (Ed.), Moral development and behavior: Theory, research, and social issues. New York: Holt, Rinehart, & Winston.

Sinclair, H., & von Glasersfeld, E. (1990). Epistemological chapters, Transforming children's mathematics education: International perspectives., (pp. 19-38): Lawrence Erlbaum Associates, Inc, Hillsdale, NJ, US.

Sinnott, J. D. (1989). A model for solution of ill-structured problems: Implications for everyday and abstract problem solving, Everyday problem solving: Theory and applications (pp. 72-99). New York: Praeger Publishers.

Sinnott, J. D. (1994a). New science models for teaching adults: Teaching as

a dialogue with reality. In J. Sinnott (Ed.), Interdisciplinary handbook of adult lifespan learning (pp. 90-104). Westport, CT: Greenwood Press.

Sinnott, J. D. (1994b). The relationship of postformal thought, adult learning, and lifespan development. In J. Sinnott (Ed.), Interdisciplinary handbook of adult lifespan learning (pp. 105-119). Westport, CT: Greenwood Press.

Smiley, P. A., & Dweck, C. S. (1994). Individual differences in achievement goals among young children. Child Development, 65, 1723-1743.

Snarey, J. R., Reimer, J., & Kohlberg, L. (1985). Development of social-moral reasoning among Kibbutz adolescents: A longitudinal cross-cultural study. Developmental Psychology, 21, 3-17.

Sontag, M., & Pedhazur, E. (1972). Dimensions of educational attitudes: Factorial congruence of two scales. Journal of Educational Measurement, 9, 189-198.

Souvaine, E., Lahey, L. L., & Kegan, R. (1990). Life after formal operations: Implications for a psychology of the self. In C. N. Alexander & E. J. Langer (Eds.), Higher stages of human development: Perspectives on adult growth (pp. 229-257). New York: Oxford University Press.

Sternberg, R. (1983). Higher order reasoning in postformal operational thought. In M. L. Commons, F. A. Richards, & C. Armon (Eds.), Beyond formal operations: Late adolescent and adult cognitive development. New York: Praeger.

Stinson, C., Milbrath, C., & Reidbord, S. (1993). Segmantation of

spontaneous speech in therapy. Psychotherapy Research, 31, 21-33.

Stuart, R. B. (1967). Decentration in the development of children's concepts of moral and causal judgments. Journal of Genetic Psychology, 3, 59-68.

Tarule, J. M. (1978). Patterns of transition in adulthood: An examination of the relationship between ego development stage variation and Gendlin's experiencing levels. Unpublished doctoral dissertation, Harvard, Cambridge, MA.

Tarule, J. M. (1980). The process of transformation: Steps toward change. In E. Greenberg, K. M. O'Donnel, & W. Berquist (Eds.), New directions for higher education: Educating learners of all ages. San Francisco: Jossey Bass.

Taylor, M. (1986). Learning for self-direction in the classroom: The pattern of a transition process. Studies in Higher Education, 11, 55-72.

Taylor, M. (1987). Self-directed learning: More than meets the observers eye. In d. Boud & V. Griffin (Eds.), Appreciating adults learning: From the learners' perspective (pp. 179-196). London: Kogan Page.

Thorkildsen, T. A. (1988). Theories of education among academically able adolescents. Contemporary Educational Psychology, 13, 323-330.

Turiel, E. (1983). The development of social knowledge: Morality and convention. Cambridge, England: Cambridge University Press.

Turiel, E., Edwards, C. P., & Kohlberg, L. (1978). Moral development in Turkish children, adolescents, and young adults. Journal of Cross-Cultural

Psychology, 9, 75-86.

van Geert, P. (1993). A dynamic systems model of cognitive growth: Competition and support under limited resource conditions. In L. B. Smith & e. Thelen (Eds.), A dynamic systems approach to development (pp. 265-331). Cambridge, MA: MIT Press.

van Rossom, E. J., Deijkers, R., & Hamer, R. (1985). Students' learning conceptions and their interpretation of significant educational concepts. Higher Education, 14, 617-641.

Vasudev, J., & Hummel, R. (1987). Moral stage sequence and principled reasoning in an Indian sample. Human Development, 30, 105-118.

Vuyk, R. (1981a). Critique and overview of Piaget's genetic epistemology, 1965-1980. (Vol. 1). New York: Academic Press.

Vuyk, R. (1981b). Critique and overview of Piaget's genetic epistemology, 1965-1980. (Vol. 2). New York: Academic Press.

Walker, L. (1984a). The validity of stages of moral development. In L. Kohlberg, D. Candee, & A. Colby (Eds.), Recent research in moral development, . Cambridge, MA: Harvard University Press.

Walker, L. J. (1982). The sequentiality of Kohlberg's stages of moral development. Child Development, 53, 1330-1336.

Walker, L. J. (1984b). Sex differences in the development of moral

reasoning: A critical review. Child Development, 55, 677-691.

Walker, L. J. (1986). Experiential and cognitive sources of moral development in adulthood. Human Development, 29, 113-124.

Walker, L. J. (1989). A longitudinal study of moral reasoning. Child Development, 60, 157-166.

Walker, L. J., & Taylor, J. H. (1991). Stage transitions in moral reasoning: A longitudinal study of developmental processes. Developmental Psychology, 27, 330-337.

Weathersby, R. (1977). A developmental perspective on adults' uses of formal education. Unpublished doctoral dissertation, Harvard, Cambridge, MA.

Weathersby, R. (1980). Education for adult development: The components of qualitative change. In E. Greenberg, K. M. O'Donnel, & W. Berquist (Eds.), New directions for higher education: Educating learners of all ages (Vol. 29, pp. 9-22). San Francisco: Jossey Bass.

Wehling, L. J., & Charters, W. W., Jr. (1969). Dimensions of teacher beliefs about the teaching process. American Educational Research Journal, 6, 7-30.

Welfel, E. R. (1982). The development of reflective judgment: Implications for career counseling of college students. Personnel & Guidance Journal, 61, 17-21.

Welfel, E. R., & Davison, M. L. (1986). The development of reflective

judgment during the college years: A 4-year longitudinal study. Journal of College Student Personnel, 27, 209-216.

Wheaton, W. M. (1984). An investigation of the relationship between cognitive developmental level, teachers' perception of role, and their understanding of issues related to the elementary curriculum. Unpublished doctoral dissertation, University of Utah.

Widick, C. (1977). The Perry Scheme: A foundation for developmental practice. Counseling Psychologist, 6, 35-38.

Willet, J. B. (1989). Some results on reliability for the longitudinal measurement of change: Implications for the design of studies of individual growth. Educational and Psychological Measurement, 49, 587-602.

Wilson, M. (1984). A psychometric model of hierarchical development. Unpublished doctoral dissertation.

Wilson, M. (1985). Measuring stages of growth: A psychometric model of hierarchical development. Occasional paper No. 29. Hawthorn, Victoria, Australia: Australian Council for Educational Research.

Wilson, M. (1989a). A comparison of deterministic and probablistic approaches to measuring learning structures. Australian Journal of Education, 33, 127-140.

Wilson, M. (1989b). Empirical examination of a learning hierarchy using an Item Response Theory model. Journal of Experimental Education, 57, 357-371.

Wilson, M. (1989c). Saltus: A psychometric model of discontinuity in cognitive development. Psychological Bulletin, 105, 276-289.

Wilson, M., & Draney, K. (1997). Partial credit in a developmental context: The case for adopting a mixture model approach. In M. Wilson, G. Englehard, & K. Draney (Eds.), Objective measurement IV: Theory into practice. Norwood, NJ: Ablex.

Witherell, C. (1978). A structural-developmental analysis of teacher conceptions of teaching and human development in relation to patterns of teaching behavior. Unpublished doctoral dissertation, University of Minnesotta.

Wright, B. D., & Linacre, J. M. (1989). Observations are always ordinal; Measurements, however, must be interval. Archives of Physical Medicine and Rehabilitation, 70, 857-860.

Wright, B. D., & Masters, G. N. (1982). Rating scale analysis. Chicago: Mesa Press.

Wright, B. D., & Stone, M. H. (1979). Best test design: Rasch measurement. Chicago, IL: MESA Press.

Wubbels, T. (1993). Taking account of student teachers' preconceptions. Teaching and Teacher Education, 8, 137-149.

Zeichner, K. M., & Liston, D. (1987). Teaching student teachers to reflect. Harvard Educational Review, 57, 23-48.

Endnotes

[1]It is interesting that the word, play, is associated predominantly with constructions at the lowest and highest orders of hierarchical complexity. Shifts in meaning lead to changes in word preference, some of which can loop back upon themselves (i.e., occur at different, non-contiguous, stages).

[2]Notice that, as hierarchical complexity increases, concepts that are not related to *education as play* at lower stages—in the present example, a concept of learning—become part of the construction of *education as play*. This means that the higher-order constructions belong simultaneously to more than one strand of meaning. At the highest orders, so many concepts are integrated into a notion of education as play, that a single proposition can be a member of numerous strands of meaning. In defining a domain, in this case, evaluative reasoning about education, this property requires attention. The next step in my research agenda is to examine relationships across strands of meaning, in order to learn more about within-stage requirements of meaning construction.

[3]My six-year-old step daughter, whose thinking is classically concrete, understands that reading and writing are used at work. "Daddy needs to know how to write because he writes e-mail."