Spring 1992

A comparison of methods for analyzing intraindividual change in student epistemological orientation during the transition to college

Trey Michael Buchanan
University of New Hampshire, Durham

Order Number 9225251

A comparison of methods for analyzing intraindividual change in student epistemological orientation during the transition to college

Buchanan, Trey Michael, Ph.D.

University of New Hampshire, 1992

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

A COMPARISON OF METHODS FOR ANALYZING INTRAINDIVIDUAL CHANGE IN STUDENT EPISTEMOLOGICAL ORIENTATION DURING THE TRANSITION TO COLLEGE

BY

TREY MICHAEL BUCHANAN
B. A., Baylor University, 1986
M. A., University of New Hampshire, 1989

DISSERTATION

Submitted to the University of New Hampshire
in Partial Fulfillment of
the Requirements for the Degree of

Doctor of Philosophy

in

Psychology

May, 1992

This dissertation has been examined and approved.

Dissertation director, Dr. Carolyn J. Mebert, Associate Professor of Psychology

Dr. Victor A. Benassi, Associate Professor of Psychology

Dr. Kathleen McCartney, Associate Professor of Psychology

Dr. John D. Mayer, Assistant Professor of Psychology

Dr. Pearl M. Rosenberg, Assistant Professor of Education

Dr. Mark B. Tappan, Assistant Professor of Education, Colby College

Date

DEDICATION

This dissertation is dedicated Adrian Arnspiger Buchanan, who spent her transition to adulthood as an sojourner in a foreign land, braving five New England winters, juggling numerous jobs, and thoroughly supporting me and my dreams through it all. Not only does her persistent affirmation and continual sacrifice give me the opportunity to ask the questions I want to, but—living Mark Twain's advice—she reminds me that scholarship shouldn't get in the way of one's education. It is also dedicated to either Mary Stuart or Matthew McFall Buchanan, whose rapidly approaching birth has given me the most important deadline I have ever had to meet.

ACKNOWLEDGEMENTS

The research reported in this dissertation was supported by numerous individuals. Primarily, I would like to thank Dr. Carolyn J. Mebert for her guidance and support for three years as my advisor, employer, and confidant. Perhaps unbeknownst to her, she was able to make my transformation from a historian of psychology to a developmental psychologist seem almost easy. Each of my additional committee members—Dr. Victor A. Benassi, Dr. Kathleen McCartney, Dr. John D. Mayer, Dr. Pearl M. Rosenberg, and Dr. Mark B. Tappan—was essential in shaping the work presented here. I owe a great deal of the success of this project to their constant support and collective affirmation that what I was doing was something valuable. In addition, I would like to thank Elizabeth A. Jordan for support and friendship thrcughout five years of graduate school; I feel fortunate to be able to count her as my first and favorite academic colleague.

Within the Department of Psychology at the University of New Hampshire, I would like to acknowledge the help of Michelle Goddard, who assisted in the data collection for the Fall 1990 cohort. A tremendous thanks also goes to the 235 undergraduates who voluntarily participated in the three studies presented here through either the kindness of their hearts or their desire to complete three-quarters of their Introductory Psychology Laboratory Experience requirement.

This research could not have been conducted without the financial aid of several entities at the University of New Hampshire. The Graduate School provided support for Study 1 through a 1990 Summer Teaching Fellowship and for the collection of data from the Fall 1991 cohort through a 1991-1992 Dissertation Fellowship. In addition, the Dissertation Fellowship provided the freedom to focus exclusively on completing the research and writing presented here. The Sophomore year follow-up was funded by the Institute for Policy and Social Science Research, which also furnished the computer resources used in collecting the longitudinal data described

here. I would especially like to acknowledge the assistance of R. Kelly Myers in developing and running the computer generated questionnaires used with those samples.

In addition, I would like to thank Judy Spiller, director of the Freshman Orientation Office, for providing me with the opportunity to obtain data from incoming students in Study 1, and the Office of Sponsored Research and Institutional Review Board for providing approval of all the research described here.

Outside of my sponsoring institution, I would like to acknowledge the influence and support of attendees of the conference "Best Methods for the Analysis of Change" held October 29-13, 1989 at the University of Southern California and sponsored by the Science Directorate of the American Psychological Association and the 1991 Meeting of the Association for Moral Education held November 7-9 at the University of Georgia, at which I originally presented portions of the research described here. I would also like to thank Michael Rovine and John B. Willett for their willingness to share with me various aspects of their work on the measurement of change. Their computer programs for analyzing longitudinal data made my job not only much easier than it would have been without them, but also made these quantitative investigations possible in the first place.

Last, I would like to acknowledge the support of my parents, John and Glenda Buchanan. Throughout my own development, their love and affirmation instilled in me a deep belief that with enough desire and hard work I could accomplish just about anything. To them I owe the persistence that carried me through five years of graduate work and enabled me to realize my dream of a doctorate in psychology.

TABLE OF CONTENTS

DEDICATION iv

ACKNOWLEDGEMENTS v

LIST OF TABLES ix

LIST OF FIGURES xi

ABSTRACT xii

SECTION PAGE

INTRODUCTION 1

I. Overview of Epistemological Development theory 5

Making Meaning in the Context of the Modern American University: An Overview of Perry's Scheme of Epistemological Development 7

Applications and Extensions of Perry's Scheme 10

Affect and Personality in the Breakdown of Dualistic Epistemological Orientations 14

Conceptualizing the Process of Epistemological Development: Change Versus Growth 16

II. Measuring Psychological Change: Methodological Questions and Answers . . . 19

Traditional Problems in Analyzing Change in Psychological Constructs 19

Conventional Methodological Approaches to Analyzing Change 30

Individual Growth Modeling: Describing and Predicting Individual Growth in Multiwave Data 35

III. Study 1: Assessment of Freshman Orientation Participants (June 1990) 43

IV. Study 2: Longitudinal Assessment of First-Semester Students (Fall 1990 and Fall 1991). 67

V. Conclusion 111

LIST OF REFERENCES 114

APPENDICES 123

LIST OF TABLES

Table 1. Comparison of advantages and disadvantages of five methods for analyzing change 34

Table 2. Descriptive statistics of dependent variables measured in Study 1 50

Table 3. Rotated and unrotated factor loadings from principle factor analysis of the BMIS (N =119) 53

Table 4. Comparison of Perry (1968) and Study 1 single factor loadings from principle factor analysis of the CLEV (N = 116) 53

Table 5. Rotated factor loadings from principle factor analysis of LOC (N = 121) 55

Table 6. Rotated factor loadings from principle factor analysis of STAI-STATE (N = 116) 57

Table 7. Rotated factor loadings from principle factor analysis of STAI-TRAIT (N = 116) 58

Table 8. Intercorrelations among dependent variables measured in Study 1 (N = 118) . . 60

Table 9. ANOVA results for gender differences in dependent variables measured in Study 1 62

Table 10. ANOVA results for differences in dependent variables as a function of home residency status 63

Table 11. Factor loadings from principle factor analysis of AT-20 (N = 105) 73

Table 12. Rotated factor loadings from principle factor analysis of MEQS (N = 121) . . 74

Table 13. Correlations between variables at Time 1 and change and residual change in dualism scores from Time 1 to Time 4 80

Table 14. Intercorrelations among dualism scores and ambiguity tolerance scores at four points (N = 105) 82

Table 15. Results of repeated measures ANOVA testing mean differences in dualism as a function of initial ambiguity tolerance group, time, and time x group interaction (N = 90) 84

Table 16. Correlations among variables obtained at Time 1 and regression coefficients in cubic growth model of dualism during the first semester in college 95

Table 17. Results of ANOVAs of demographic differences in linear-cubic change in dualism group (N = 53) 102

Table 18. Results of ANOVAs of differences in mean anxiety and and mean mood measures as a function of linear-cubic dualism change group (N = 53) 103

Table 19. Results of ANOVAs of differences in mean locus of control, mean ambiguity tolerance, and mean multiplicity as a function of linear-cubic dualism change group (N = 53) 104

Table 20. Comparison of the relative distribution of growth pattern groups obtained during the first semester in college (N = 180) and in the Spring 1992 follow-up (N = 28) 107

LIST OF FIGURES

Figure 1. Comparison of two hypothetical studies of change over two time points 23

Figure 2. Change in dualism scores of 10 randomly chosen individuals across four time points 78

Figure 3. Chart of mean dualism scores by initial ambiguity tolerance groupings over four time points (N = 90) 85

Figure 4. Comparison of ten linear, quadratic and cubic growth models of dualism over time (n = 10) 88

Figure 5. Histogram and summary statistics of slope estimates of true linear growth in students' epistemological dualism during the first semester in college (N = 209) 99

Figure 6. Histogram and summary statistics of slope estimates of true quadratic growth in students' epistemological dualism during the first semester in college (N = 209) 90

Figure 7. Histogram and summary statistics of slope estimates of true cubic growth in students' epistemological dualism during the first semester in college (N = 209) 91

Figure 8. Matrix crossing linear and quadratic types of change in students' dualism scores during the first semester in college (N = 180) 97

Figure 9. Classification of patterns of linear and cubic change in students' epistemological orientation during the initial transition to college (N = 180) 99

Figure 10. Matrix crossing linear and cubic types for change in students' dualism scores from the first semester in college to second semester sophomore year N = 28) 106

ABSTRACT

A COMPARISON OF METHODS FOR MEASURING INTRAINDIVIDUAL CHANGE IN STUDENT EPISTEMOLOGICAL ORIENTATION DURING THE TRANSITION TO COLLEGE

by

Trey Michael Buchanan
University of New Hampshire, May, 1992

The ability to measure the development of epistemological beliefs of 235 college students during their first-semester in college was investigated by comparing the results obtained using five different methodological approaches to measuring change. These approaches included the use of the two-wave difference score, the residual change score, the cross-time correlation matrix, repeated measures analysis of variance, and individual growth modeling. A cubic individual growth model was found to be superior to other methods in describing intraindividual differences in dualistic epistemological orientation during the transition to college. An investigation of the existence of systematic interindividual differences in growth as a function of anxiety, mood, tolerance of ambiguity, and various demographic characteristics failed to find any significant differences among four identified patterns of linear and cubic change in dualism: no change, fluctuating increase, straight increase, fluctuating decrease, and straight decrease. Suggestions are given for both future empirical investigations of epistemological development in college students and the measurement of longitudinal change in psychological constructs.

INTRODUCTION

The research reported in this dissertation reflects the confluence of two broad trends in contemporary developmental psychology. The first of these is the increased interest in issues of adult development, and particularly adult cognitive development. The second trend involves the application of new approaches for analyzing longitudinal data developed by methodologists in in educational and psychological research. I have been able to focus attention on specific aspects of both issues by choosing to examine the intellectual impact of the initial transition to college using several different methodological approaches.

College Student Epistemological Development

William G. Perry, Jr.'s (1968, 1970, 1977, 1981) work on college students' intellectual and ethical development provided the theoretical framework from which I began this project. Perry's view of the emerging adult as a natural epistemologist has been recently described as one of a "diversity of postformal soft-stage proposals in conceptualizing adult development" (Richards & Commons, 1990, p. 159) that attempt to extend Piaget's structural, cognitive developmental theory beyond adolescence. Soft-stage theories describe development that is "optional rather than necessary" (Kohlberg & Ryncarz, 1990, p. 204) and therefore differ from the traditional notion of universal development described by "hard-stage" theories of cognitive development. Perry stresses the culturally-specific nature of his scheme by stating that his ideas are most relevant for understanding "the intellectual and ethical development of late adolescence in a pluralistic culture" (1970, p. 3).

As a senior in college, I originally learned of Perry's scheme as part of a course on group processes. My best guess as to why my professor presented this theory is that he thought he might be able to enhance his students' development with a bit of an exposure to Perry's ideas. Since that time I have mentioned, or have heard others mention, Perry's scheme to numerous university instructors and administrators with it being met with almost the same reaction

every time. To these educators Perry's conception of college student development seemed so intuitive that few often inquire about the empirical evidence supporting the scheme. My curiosity in examining Perry's scheme empirically was initially sparked because of this lack of curiosity in others. In a sense, Perry's scheme seemed to good to be true, or was it?

In order to explore the empirical nature of Perry's scheme, this dissertation research examines the impact of this initial transition in the epistemological development of older adolescents. Even though this is rather circumscribed topic given the totality of Perry's scheme, I feel that much of the actual developmental processes that result in epistemological growth have gone largely unexamined. Although empirical research has focused on the social and emotional aspects of this phenomenon (Baker & Siryk, 1984; Constantinople, 1969; Cutrona, 1982; Feldman & Newcomb, 1969; Jay & D'Augelli, 1991; Waterman & Waterman, 1971), few studies have examined the intellectual changes students experience as they become part of the academic community. A primary goal of this project is to begin to fill the empirical gaps in developmental psychology's understanding of intellectual development during college, for many individuals a central event of their transition to adulthood.

In a concluding comment about my interest in the empirical validity of Perry's scheme, I am quite sure that the choice of this topic reflects my own struggle with issues of psychological knowledge. My initial background in the history and theory of psychology has made me keenly aware—much of the time unconsciously—of both the incredible potential and the critical limits of human knowledge. Therefore, I feel that my interest in Perry's work reflects not only a research interest, but at a personal level it reflects my own struggle with epistemological questions.

Methodological Issues in Developmental Psychology

In addition to my interest in psychological development during the transition to adulthood, I also chose this dissertation research to address my interest in the methodological problems associated with studying human development. My specific interest in methodology involves investigating the ability of different analytical approaches to both measure intraindividual

change and to account for interindividual differences in change. As Nesselroade (1991) has pointed out, this distinction—that developmental psychologists are inherently always dealing with two levels of change—is by no means a novel idea; but it is one that seems to have escaped the attention of many researchers who attempt to address developmental questions.

In 1989, through good fortune I was able to attend an APA sponsored conference entitled "Best Methods for the Analysis of Change" where I was first exposed to many of the critical issues and exciting possibilities that are currently being discussed among methodologists interested in change. From the proceedings of that conference I was inspired to try to directly address some of these methodological issues in my dissertation research. After completing this project, I have begun to realize that, not unlike most areas of intellectual inquiry, the more I think I "know" about analyzing change the more I'm sure that I've only been able to scratch the surface of the issue. By attempting to contribute to this discussion, however, I hope that the research presented here, if nothing else, supports the growing uneasiness in scientific psychology with questions that fail to take into account that human experience is constantly changing. That is to agree with James's contention that psychological events are never "got by us twice" (1890/1983, p. 225), but rather that "experience is remoulding us every moment" (1890/1983, p. 228).

Overview

Because the work presented in this dissertation involves the integration of substantive questions about late-adolescent development with methodological questions about measuring change, the first two sections provide separate discussions of these topics. In Section I, Perry's scheme of epistemological development is outlined, relevant research is reviewed, and a series of unanswered questions about the initial transition of the scheme are presented. Section II provides an overview of the major issues in the methodological literature on analyzing change, followed by a discussion of several conventional techniques used to measure change and the presentation of a more recent approach, the technique of individual growth modeling.

A report describing a pilot study conducted in the summer of 1990 is presented in Section III. This initial study was conducted in order to assess the psychometric properties of several scales designed to measure not only epistemological development, but several other variables, including anxiety, mood, and locus of control. Section IV presents the method and results of the longitudinal analysis of two cohorts of students during their first semester in college. These analyses compare the results obtained from several approaches to measuring change, with a focus on determining which approach or approaches best describes epistemological development during the transition to college. Last, Section V provides an overall discussion of the role of affect and personality in early epistemological development, the methodological lessons learned from comparing approaches to measuring change, and some suggestions for future research on both topics.

I. OVERVIEW OF EPISTEMOLOGICAL DEVELOPMENT THEORY

Within twentieth-century developmental psychology, the cognitive-structural approach has provided one of the most important influences on the development of theory, research, and practice. This approach views psychological development as the result of the dynamic interaction between the individual and the environment, principally through the successive adaptation of patterns of thought and behavior to the demands of a changing world. Beginning with the work of Baldwin (1902, 1968), the cognitive-structural approach has achieved impressive articulations in both Piaget's genetic epistemology (Flavell, 1963; Piaget, 1952), and Kohlberg's work on the development of moral reasoning (1981, 1984).

This theoretical approach, far from being a recent phenomenon, has its roots in several nineteenth-century intellectual, scientific, and political movements. Several of the most important of movements include the intellectual impact of Darwinian evolutionary theory (Charlesworth, 1992; Vidal, Buscaglia, & Voneçhe, 1983), the rise of the "new" scientific psychology in Europe and the United States (Albrecht, 1960; O'Donnell, 1985), and the beginnings of progressive American meliorism at the turn of the century (Cremin, 1980). The influence of each of these movements converged with a growing acceptance of the romantic conception of the child proposed by Rousseau two centuries earlier (1672/1979) and provided twentieth-century psychology with a broad framework within which many basic questions about the nature of the course of human life would be successfully addressed.

Despite numerous contributors to this enterprise, Piaget has been by far the most influential of the cognitive-structural development theorists in American psychology. Through the translations of his various writings—many of which were rediscovered in the midst of psychology's "cognitive revolution" of the 1960's—Piagetian thought has provided the starting point for much of the research on language development (1926), moral development (1932),

and cognitive development (1928/1969, 1929/1983, 1952). Although many of Piaget's ideas have been criticized (Case, 1985) and empirically challenged (Baillargeon, 1987, for example), his metaphor of the continual adaptation of internal psychological schemas to external physical reality remains central in developmental psychology.

While Piaget's own work focused exclusively on psychological development through adolescence,[1] many developmental psychologists have attempted to explore what has recently become known as postformal thought (Commons, Richards, & Armon, 1984; Richards & Commons, 1990). One of the most widely cited, and I will also contend the most empirically neglected, examples of postformal cognitive theory was developed by Perry in his 1970 book, Forms of Intellectual and Ethical Development in the College Years: A Scheme. Commonly referred to as "Perry's scheme," this theory is an attempt to describe the way in which certain individuals, principally college students, "make meaning" (Perry, 1981) out of their social and intellectual experiences as they complete their transition from adolescence to adulthood. Not only did Perry view his scheme as partly a legacy to Piaget's theory (Perry, 1970), but his scheme of "intellectual and ethical development" has come be to described as a theory of epistemological development, stressing the central theme of individuals' constructing more complex assumptions about the nature of reality and knowledge common to both psychologists' work.

In order to better understand Perry's contribution to the cognitive-structuralist enterprise, an overview of his scheme of development will be presented in this section, followed by a further discussion of its philosophical and theoretical roots. In addition, I will outline the existing research on Perry's scheme, including certain important applications and extensions of his work. Last, I will offer two aspects of Perry's scheme that will be

[1] One notable exception to this focus on nonadult development is Piaget's (1972) brief treatment of the cognitive demands faced by the adolescent during the transition to adulthood.

empirically addressed in this dissertation research: the conception of change and the role of other noncognitive domains in students' epistemological development.

Making Meaning in the Context of the Modern American University: An Overview of Perry's Scheme of Epistemological Development

In the late 1950's, the Bureau of Study Council at Harvard College, headed by William G. Perry, Jr. (1968), began a series of studies of the collegiate experience of undergraduate students at Harvard and Radcliffe Colleges. These studies focused primarily on the ways in which students constructed meaning out of their academic and social experiences. Perry's interest in examining students' "making meaning" out of their college experience was driven by the rise in the intellectual and moral pluralism that characterized the modern American liberal arts college (Adams, 1931, cited by Perry, 1970). In Perry's own words, the goal of the Bureau's investigations was "to illustrate the variety in students' responses to the impact of intellectual and moral relativism" on their college experience (1970, p. 7).

Although Perry and his colleagues fully expected to address the issue of college student's experience from an individual differences perspective—reflecting the tenor of American personality psychology—they soon began to realize the developmental nature of differences in students' assumptions about knowledge (Perry, 1981). Shifting the focus from interindividual differences to intraindividual differences reoriented the task of the Bureau of Study Council towards mapping the developmental process of the intellectual impact of the college experience.

During two decades, research conducted at Harvard and Radcliffe Colleges found that the intellectual and ethical development of students followed a common path, one that leads from an initial belief in a relatively rigid "dualistic" view of knowledge to a point of personal commitment to a specific set of epistemological beliefs within a relativistic world (Perry, 1968, 1970, 1981). Perry (1981) likened this process of development to Bunyan's allegory Pilgrim's Progress (1728/1939), by describing his scheme as one in which the seeker of knowledge faces various epistemological obstacles and temptations. The resulting developmental scheme described by Perry contained nine different "positions," each with its

own set of beliefs and assumptions about the nature of knowledge, the goals of education, and the place of the self in a world of ideas. It has been suggested that the first half of Perry's scheme (Positions 1-5) deals more directly with issues of intellectual development, and that its primary focus shifts to ethical and moral issues during the second half (King, 1978); the term epistemological development will be used here as a convenient term referring to both aspects of development in the college years.

Four Epistemological Orientations of Perry's Scheme

Perry proposed four broad phases of college student development, each characterized by a different epistemological stance: simple duality, multiplicity, relativism, and commitment in relativism (Perry, 1970). Put briefly, Perry's scheme traces students' epistemologies from the initial belief that knowledge is a matter of obtaining the "right facts" about the world (simple dualism), through the suspicion that knowledge is just a matter of opinion (multiplicity), through the realization that some opinions or theories are more satisfying or "right" than others (relativism), to the realization that continued maturity requires tentative yet strong decisions be made regarding the nature of knowledge in the face of uncertainty (commitment in relativism).

Simple Dualism (Positions 1-2). Perry found that the beliefs of the entering college student have a tendency to stress the certainty of knowledge, especially the "expert" knowledge found in the college classroom. As a result, students holding dualistic beliefs view their role in the educative process as largely a passive one in which one must "work hard, read every word, and learn the Right answers" (Perry, 1981, p. 79) that certified authorities provide. The general characteristic of this stance is the belief that knowledge is divided "into two realms—Good versus Bad, Right versus Wrong, We versus They" (Perry, 1981, p. 79). Perry views this epistemological orientation as one that enables students to temporarily create order out of the diversity of the educational atmosphere of a modern liberal arts college.

Multiplicity or Complex Dualism (Positions 3-5). Students who initially hold dualistic beliefs, according to Perry, soon modify them as the result of exposure to the pluralistic

intellectual and social contexts of the university. The challenges to dualism from the classroom and the dormitory leads dualistic students to question the certainty of at least some areas of knowledge, and therefore question the validity of authorities' expert knowledge. Perry found that this complexity of belief can be seen in students' skeptical view of truth in certain areas of inquiry, such as the humanities and social sciences, but not others, such as the physical and natural sciences. However, as students continue to reject their dualistic assumptions, multiplistic epistemological orientations appear as students begin to embrace the existence of a plurality of valid knowledge claims, characterized by the belief that "every one has a right to his own opinion; none can be called wrong" (Perry, 1981, p. 80).

Relativism (Positions 6-7). The third epistemological orientation in Perry's scheme appears as students' find that a multiplistic "anything goes" stance fails to solve the problem of knowing: they begin to realize that, although certainty is unattainable, useful judgments can be made concerning the relative accuracy of various competitors. By referring to this period as relativistic, Perry eschews the common negative connotations of the term, seeing them as more appropriate for describing the structure of multiplistic thought. His definition of the third developmental position describes relativism as the belief in "a plurality of points of view, interpretations, frames of references, value systems and contingencies in which the structural properties of contexts and forms allow for various sorts of analysis, comparison and evaluation" (Perry, 1970, Glossary, emphasis added). A central aspect of this relativistic epistemological orientation can be seen in a student's appreciation of the contextual nature of knowledge, enabling the evaluation of competing knowledge claims in light of their specific context.

Commitment in Relativism (Positions 8-9). The primary developmental result of epistemological relativism lies in its ability to allow students to exercise personal choice and commitment in the face of uncertainty. Perry's describes this as the "process of orienting the self in a relative world" that involves "a conscious . . . realization of identity and responsibility" (Perry, 1970, Glossary). Finding this epistemological stance in students near

the end of their college careers, Perry (1970) views such personal commitment as the appropriate response to societal demands of the post-college world, including the demands of career, lifestyle, and even religious, philosophical and political belief. One critical aspect of commitment in relativism, however, is its tentativeness. One of Perry's students voices this critical aspect by saying that "you constantly have times of doubt and tension . . . [which is] a natural thing in existing and being open, trying to understand the world . . . [and] the people around you" (Perry, 1981, p. 95).

With a concluding irony, Perry describes the "final" position in his scheme as one in which past epistemological issues are revisited: personal commitment may result in the resurrection of certain beliefs found in earlier positions. In profound sense, Perry views his developmental scheme as a recursive one, and in his later work chose to describe it in geometric terms of a "helix, perhaps with an expanding radius to show that when we face the 'same' old issues we do so from a different and broader perspective" than before (Perry, 1981, p. 97)

Applications and Extensions of Perry's Scheme

As one of the more influential investigations of cognitive development in late adolescence, Perry's work on the development of individual epistemologies has found the greatest influence in the areas of general practice in higher education (Kitchener, 1982; Widick, 1977), curriculum design (Heffernan, 1975; Stephenson & Hunt, 1977), and career development (Knefelkamp & Slepitza, 1976; Touchton, Wertheimer, Cornfield, & Harrison, 1977). Along with these applications, a great deal of effort has been put to the task of developing measures that assess intellectual growth along Perry's scheme.

Instrumentation

Even with just a cursory review of research devoted to Perry's scheme (Moore, 1991), it is not difficult to see that a huge emphasis has been placed upon developing reliable and valid measures of Perry's positions. Perry originally constructed a 90-item, self report scale for measuring epistemological orientation called the Checklist of Educational Views or CLEV (Perry, 1968), reported initial success with it, but later abandoned it in order to use a more qualitative

and flexible interview protocol. Because of a great deal of resource demands (i.e., time and financial) incurred by training interviewers and conducting the interviews themselves, numerous attempts have been made to develop written forms of the Perry interview, including Knefelkamp (1974) and Widick's (1975) KneWi, Moore's Measure of Intellectual Development (1982a), and most recently Baxter Magolda and Porterfield's Measure of Epistemological Reflection (1985, 1988). Each of these has achieved various levels of success, and they currently remain the instruments of choice for measuring epistemological orientation throughout the course of students' college careers. Although Perry and many others have moved away from the use of standard objective measures of epistemological orientation, several researchers have found some success in using shortened versions of the CLEV to measure levels of dualism in college student samples (Ryan, 1984a; 1984b; Wilkinson & Schwartz, 1987).

Theoretical Extensions

In addition to inspiring work on assessment, Perry's conceptions of epistemological development have been theoretically refined in the work of several developmental and educational psychologists by emphasizing either the intellectual or ethical dimensions of late-adolescent and adult development. The work of Kitchener, King, and their colleagues (King, Kitchener, Davidson, Parker, & Wood, 1983; King, Kitchener, & Wood, 1991; Kitchener, 1983; Kitchener, 1986; Mines, King, Hood, & Wood, 1990) on the Reflective Judgment model narrows the focus of Perry's model from general intellectual issues to the specific "way a person reasons to a conclusion about problems that do not have verifiable right and wrong answers" (Mines, King, Hood, & Wood, 1990, p. 539). Such "ill-structured" problems (Churchman, cited in Kitchener, 1983) are typically ones in which "there is not a single, unequivocal solution," but rather "a solution must be constructed by integrating and synthesizing diverse data and opinion" and choosing "which set of theoretical assumptions best fit the problem and evidence at hand" (Kitchener, 1983, p. 224). Similar to Perry's conclusions, proponents of the Reflective Judgment model describe a developmental sequence of

assumptions and strategies for dealing with ill-structured problems that culminates in the active "process of critical inquiry and synthesis" of facts and assumptions (Kitchener, 1986, pp 78-79).

Following Kohlberg's work in moral development theory (1981, 1984) recent attempts to reexamine the nature of late-adolescent and adult moral experience have resulted in a reassessment of the ethical dimensions of Perry's scheme. Using Perry's scheme, the theoretical collaboration of Gilligan, Murphy, and Tappan on moral development during this time of life (Gilligan & Murphy, 1979; Gilligan, Murphy, & Tappan, 1990; Murphy & Gilligan, 1980) provides a point of convergence on the issues of justice (Kohlberg, 1981, 1984) and care (Gilligan, 1977). Perry's work is cited by Gilligan, Murphy, and Tappan (1990) as evidence that, in adulthood, moral "restructuring is based upon the epistemological recognition that life choices are made in a world of relationships . . . [and] forms the basis for an expanding adult ethic that encompasses empathy . . . and understanding, thereby joining respect with response in bringing morality into the context of human relationships" (p. 224).

In addition to using Perry's scheme to attempt to resolve the issue of gender differences in moral development, Tappan (1989) has combined aspects of literary criticism, hermenuetical interpretation, and constructivist accounts of personal narratives to address the developmental nature of moral thought during the transition to adulthood. Through the examination of students "stories" of their first year in college, Tappan concluded that late adolescent moral development is better described by the "ironic" narrative structure of Perry's scheme (i.e., one that emphasizes that the eventual uncertainty of intellectual and moral claims) than the "romantic" narrative structure embodied in Kholberg's theory (i.e., one that emphasizes the attainment of an advanced moral perspective). This use of Perry's scheme in the realm of late adolescent moral reasoning may provide an avenue for improving the understanding of moral development in later life, an area of weakness acknowledged by Kohlberg himself in his later writings (e.g., Kohlberg & Ryncarz, 1990).

A Remaining Question about Perry's Scheme: The Transition from Dualism to Multiplicity

At least one commentator has described "the most fundamental transition in [Perry's] scheme" as the one involved in "moving from a primitive conception of knowledge as an unorganized set of discrete and absolute truths to a more mature conception of knowledge as an array of interpreted and integrated propositions" (Ryan, 1984b, p. 1227). Despite its critical importance in Perry's scheme, the process of the breakdown of dualistic thought has not been systematically examined. This may stem from the difficulty of identifying pure "dualists" among first-year college students (P. M. King, personal communication, November 10, 1991) and the finding, after a decade of research, that the average first-year student appeared to be developmentally ahead of Perry's original cohorts (Perry, 1981). Although some have had little success finding such pure dualists, recent research has shown that after more than 20 years of research on Perry's scheme, many first-year students continue to subscribe to certain aspects of epistemological dualism (Baxter Magolda, 1990).

Given that the seeds of later development begin to take root in the initial transition from dualism to multiplicity, what is empirically known about this transition? Although Perry sketched out the typical ways in which the breakdown of dualism is accelerated—through the influence of diverse opinions of peers and uncertain claims made by intellectual authorities—little is known about the actual psychological experience of students in their "fall" from epistemological "grace." [2] For example, what is the emotional impact of this transition? Are students happier, disappointed, more confused, or even relieved by the implications of questioning of intellectual uncertainty? In broader terms, how might the affective dimension of

---

[2] In one of Perry's favorite metaphors, he equates the breakdown of dualism with the Biblical tale of Adam and Eve's temptation and sin:

> Indeed there is so little that is novel about it that one finds almost a full expression [of this transition] in the Book of Genesis. . . . It was, after all, the serpent who pointed out that the Absolute (the truth about good and evil) was distinct from the Deity and might therefore be known independently—without His mediation. The Fall consisted of man's taking upon himself, at the serpent's suggestion, the knowledge of values and therefore the potential of judgment. (Perry, 1970, p. 60)

students transition to college be influenced by their epistemological development? Answers to this question and similar ones will begin to provide a more detailed view of the personal dimensions of epistemological development among college students.

Affect and Personality in the Breakdown of Dualistic Epistemological Orientations

Because Perry and his colleagues have left a major issue unaddressed in their research—the impact of the transition from duality to multiplicity—the extent to which intellectual development is associated with other areas of development in unknown. In their review of past research on Perry's scheme, Baxter Magolda and Porterfield stress that

> "exploring relationships between intellectual development and other types of development is . . . essential. Most developmental theorists would agree that strands of development are intertwined and that clarifying these connections would enhance our understanding of individual strands as well as overall development. Psychosocial, affective, moral, and personality development are a few logical areas to be pursued in relation to intellectual development. (1988, p. 62)

In addition, Perry's own theoretical statements about his scheme (1970, chap. 4, 1977, 1981) raise a number of questions about the role of other psychological domains in late adolescent intellectual development. This research will address Baxter Magolda and Porterfield's concern by examining the strands of development that link affect, personality, and epistemology in first-year college students' experience.

Epistemological Development and Affect

I will examine the role of affect in the process of epistemological development by measuring the quality of students' mood states as they accommodate their theories of knowledge. By this, I will address the following: Is the cognitive restructuring acquired in moving from dualistic to multiplistic epistemologies accompanied by certain emotional experiences?

Cognitive balance or consistency theory, which came to prominence in social psychology through the work of Festinger (1957) and Heider (1958), predicts a positive relation between a need for cognitive restructuring and negative affect. In its various forms, this theory centers on the psychic dissonance created by holding incompatible cognitions. For example, if a student

believes that hard work is the basic requirement for success in college, she may experience anxiety when a roommate receives a better grade on a similar assignment that took her roommate half as long to complete. If the belief that effort leads to success (or other dualistic attitudes such as the assurance of right answers and trust in expert authorities) is disconfirmed by the pluralistic intellectual and social milieus of the university, students exhibiting greater duality should also report greater anxiety and more negative mood states.

Epistemological Development and Personality

The second psychological domain I will examine in relation to epistemological development will be broadly defined as personality, and will include two specific dimensions of student's personality: tolerance of ambiguity and locus of control.

As a personality construct, tolerance of ambiguity can be defined as the tendency to perceive ambiguous situations as unthreatening (Budner, 1962, cited in Kirton, 1981). In a review of the development and use of tolerance of ambiguity scales, it has been found to be negatively associated with a number of personality measures, including dogmatism, rigidity, and hind-sight bias (Buchanan, 1986). This finding suggests that tolerance of ambiguity may be part of a more global personality factor, one resembling an authoritarian-democratic dimension (J. D. Mayer, personal communication, August 13, 1991).

Perry's description of the earliest stages of intellectual development in college students, and primarily his notion of epistemological dualism, is theoretically similar to much of the original work on authoritarian personality (Adorno, Frenkel-Brunswick, Levinson, & Stanford, 1950) and intolerance of ambiguity (Frenkel-Brunswik, 1949). Therefore, I will examine the relationship between epistemological orientation and tolerance of ambiguity as one way of linking Perry's work to broader notions of personality development. Among students who exhibit a change in epistemological orientation, I expect to find a consistent negative correlation between dualism and tolerance of ambiguity over time. In other words, as students begin to take on more multiplistic beliefs about the nature of knowledge, I expect them to become more tolerant of ambiguous situations and statements.

To address these issues developmentally, this research will examine the changing relationships among affective experience, ambiguity tolerance and individual epistemology during the initial transition to college. However, because of the complexity of examining psychological change at several levels, different methodological approaches need to be combined in order to investigate the nature of epistemological development.

Conceptualizing the Process of Epistemological Development: Change Versus Growth

One of the most difficult aspects of exploring the details of Perry's scheme empirically centers around his conceptualization of development. Although he contends that "the sequence of structures we observe in our data qualifies as a 'developmental' pattern in the special sense . . . of an orderly progress in which more complex forms are created by the differentiation and reintegration of earlier, simple forms" (Perry, 1970, p. 44), he is quick to observe that his notion of epistemological development does not necessarily imply a linear succession of "stages" in the sense of Piagetian and Kohlbergian models of development. Rather, Perry argues that intellectual and ethical development is better described with an "Emeshing Metaphor" (1977, p. 51) of recursive change:

> The structure of a metaphor of [epistemological] growth must . . . contain at least the complexity of a helix in which the same ground may be traversed at higher levels—and hopefully from broader perspectives.
>
> I do not know the geometrical name for a helix with an increasing radius as for example a tornado. Let us suppose, however, that we take the present ["linear"] scheme and twist it round into such an expanding spiral in which it would be "more advanced" to be wrestling with Position 1 the second or third time around than with Position 9 the first time around. Then add to this the necessary "horizontal decalage." I'm sure that the metaphor will give a better fit and that this is the kind of theory we must learn to use. (Perry, 1977, pp. 51-52)

In his 1981 review of his scheme, Perry again expressed his concern about the

> limits of the linear structure embedded in the metaphor of our journey. We have followed our students in their cumulative expansion of the meanings of their worlds. Our map of their adventure has required only two dimensions, for in the time at the students' disposal they could traverse his "Pilgrim's Progress" only once. But any adults who have perused the diaries of their teens know well that growth and discoveries are recursive. We are shocked at finding we "knew" at sixteen what we discovered yesterday. Have we just been going around in circles? Yet the "same" issues, faced over and over again, may not really be the same.

> Perhaps the best model for growth is neither the straight line nor the circle, but a helix, perhaps with an expanding radius to show that when we face the "same" issues we do from a different and broader perspective. (Perry, 1981, p. 97)

One of the central results of this conceptualization of development involves the definition of what is meant in Perry's scheme by the terms development, growth, and change. As seen in the quote above, Perry views his scheme as developmental and involving "growth" in the very broad sense of increasing complexity. However, the actual process of change that underlies this development is not as easily described. In reading Perry's account of the typical college student's intellectual development, what does appear is the idea that epistemological change is characterized by a type of "bouncing around" among inconsistent claims about the nature of knowledge. Drawing from the work of Riegel (1973), for example, Perry acknowledges the dialectical nature of adult thought. Does this conception of development, however, result in a model that describes development that is more like change and less like growth?

Perhaps one way to understand Perry's scheme is to emphasize the importance of viewing epistemological development as a process that occurs at two levels, one being the transitions from one position to another and one being the overall developmental progress as various epistemological orientations are attained. In such a view, an indivdual's epistemological journey follows a path defined by both microgenetic change and macrogenetic growth (M. B. Tappan, personal communication, April 2, 1992). In other words, the idea of overall development as a function of microgenetic change and macrogenetic growth is similar to the notion of a planet moving along its orbit by completing a series of smaller elliptical orbits. Such a complex metaphor acoomodates both the necessary recursive nature of epistemological development described by Perry, while at the same time retaining its the general notion of developmental growth as a the result students "making meaning" within a pluralistic society.

Given the unconventional conception of development that arises out of Perry's scheme, empirical investigations of epistemological development must be able to accommodate such "twists" and "turns" in their choice of methodology, both at the level of research design and statistical analysis. Although Perry's scheme appears to present a unique set of methodological

issues, it will be shown in the following section that much, if not all, of developmental research is faced with dealing with the problems of appropriately measuring and analyzing psychological change.

II. MEASURING PSYCHOLOGICAL DEVELOPMENT: METHODOLOGICAL QUESTIONS AND ANSWERS

One interesting aspect of the history of modern scientific psychology is that it can largely be told as a tale of the evolution of methodology. Wundt's pioneering use of controlled introspection (Hilgard, 1987), Freud's analysis of the unconscious using dream analysis and the clinical interview (Shakow & Rapaport, 1964), and Cattell giving American psychology its characteristic focus on individual differences by developing the psychological test (Sokal, 1987) are just a few examples of the way in which method has informed the practice of psychological science. The study of human development is no different: as McCall (1977) outlined more than a decade ago, the major challenges to developmental psychology continue to be essentially methodological in nature. A summary of these challenges is presented in the following question: how can psychologists interested in developmental processes best measure, describe, and account for change in individuals' abilities, attitudes, beliefs and behaviors?

Because the major focus of the research presented here deals directly with answering this question, this section will discuss several specific problems facing the measurement of change, address the limitations of conventional methodologies for analyzing change, and present a description of a recent methodological improvement in the measurement of change, individual growth modeling. It addition, I will argue that individual growth modeling currently is the best method for addressing the dynamic nature of psychological processes.

Traditional Problems in Analyzing Change in Psychological Constructs

In each of the past three decades, developmental methodologists and researchers have continued to discuss the problems inherent in investigating change in psychological constructs. From this vast literature, three representative authors provide distinct yet congruous statements regarding both these problems and possible avenues to their solution: Bereiter (1963), McCall (1977), and Rogosa (1988).

Bereiter's (1963) Three Persisting Dilemmas in Measuring Change

In one of the first attempts to bring methodologists together to solely discuss the problems associated with measuring psychological change,[3] Bereiter (1963) argued that many developmental processes go unstudied because of the lack of appropriate methods of analysis. Bereiter believed that whereas many difficult areas of research often inspire creative ways around methodological problems, the investigations of developmentally oriented researchers are frustrated by the persistence of three dilemmas: the over-correction—under-correction dilemma, the unreliabilty—invalidity dilemma, and the physicalist—subjectivist dilemma. Because each of these dilemmas involves choosing between "equally undesirable alternatives" (p. 3), a review of their implications for developmental research will provide a context for the methodological rationale taken in this work.

The Over-Correction—Under-Correction Dilemma. This dilemma centers around what steps should be taken to correct for the common finding of a negative correlation between raw change, measured by the difference score, and initial status. Bereiter traces this observation back to a 1924 study in which Thorndike determined that initial status and the difference score have the same measurement error, yet each has a different sign (i.e., + or -). The resulting negative correlation is, therefore not inherent in the data, but rather a computational artifact produced by calculating the difference score. Because of this statistical artifact, many methodologists have devised methods for attenuating this spurious negative correlation.

The most common procedure for attenuating the correlation between initial status and change is to compute partial correlations controlling for differences in initial status. Although partialing out the effects of initial status in most cases leads to an increase in the correlation between change and other independent measures associated with initial status, Bereiter argues

---

[3] Bereiter's chapter appeared in an edited book based upon the proceedings of a 1962 conference sponsored by the Committee on Personality Development in Youth of the Social Science Research Council which included, in addition to Bereiter, some of the more notable methodologists in psychology: R. Darrell Bock, Donald T. Campbell, Raymond B. Cattell, Chester W. Harris, Wayne H. Holtzman, and Frederick M. Lord.

that it still underestimates actual population correlations because "the zero-order correlations entering into the partial correlation must be corrected for attenuation beforehand" (Bereiter, 1963, p. 8). Without this correction for measurement error, the partial correlation approach falls short of producing unbiased estimators of correlations with true change.

The dilemma of over-correction versus under-correction, therefore, involves a choice between leaving change and initial status unattenuated which leads to an overcorrection for the effects of initial status, or performing a typical partial correlation that attenuates for a computational artifact but not for the effects of measurement error. Bereiter suggests avoiding a choice between these two unfavorable alternatives in favor of a third: partialing out the effects of true initial status (i.e., corrected for measurement error) from correlations involving the difference score.

The Unreliabilty-Invalidity Dilemma. Bereiter held that this second dilemma "stems from the fact that high reliability of change scores usually requires low test-retest correlations" (p. 20). If increasing the reliability of a difference score results in the two variables being uncorrelated, how valid is a difference score between two different things? Bereiter concluded that for many researchers "once it is allowed that the pretest and posttest measure different things, it becomes embarrassing to talk about change" (p. 11).

Although this may appear to be a serious dilemma, Bereiter concluded that it was, in fact, a false one. Invalidity (i.e., low correlation) is only a cause for concern if the analytic goal deals with final status or outcome; if analyzing change is in and of itself the analytic goal, invalidity ceases to be as salient a concern. Drawing from an example presented by Bereiter (pp. 12-13), consider the ability of two hypothetical studies of change to predict both individual differences in final status and individual differences in change. In Study A, subjects are given a spatial reasoning test as children and again as adolescents, and a high correlation, $r = .95$, between subjects' scores at Time 1 and Time 2 is computed. In Study B, a sample of factory workers is given a work satisfaction inventory at the beginning and end of a production year, and a low correlation, $r = .15$, is obtained for the two measures.

Figure 1 presents a comparison of the results of these hypothetical studies by plotting the change in each pair of dependent variables over time for ten subjects from each study. As the magnitude of their respective correlations would suggest, Figure 1 shows that the rank-order of subjects from Time 1 to Time 2 is maintained in Study A but not in Study B. How adequately does each study address the goals mentioned above? The results of Study A appear to be more than adequate for predicting final status from initial status, but the high Time 1-Time 2 correlation results in very little individual variation in change. Because the maintenance of rank-order from Time 1 to Time 2 causes the individual difference scores to be relatively homogeneous, it becomes difficult to predict variation that does not exist.[4]

The results of Study B, however, show a substantial amount of variation in change compared to Study A.[5] Allowing the Time 1-Time 2 correlation to be weak enables the difference scores to be extremely heterogeneous. If a large Time 1-Time 2 correlation is required for an investigation of change to be valid, then Study B would be rejected as methodologically flawed. In other words, rigidly fixing the criterion of Time 1-Time 2 reliability greatly decreases the interindividual variability of growth. Referring to results similar to those from Study B, Bereiter argued that

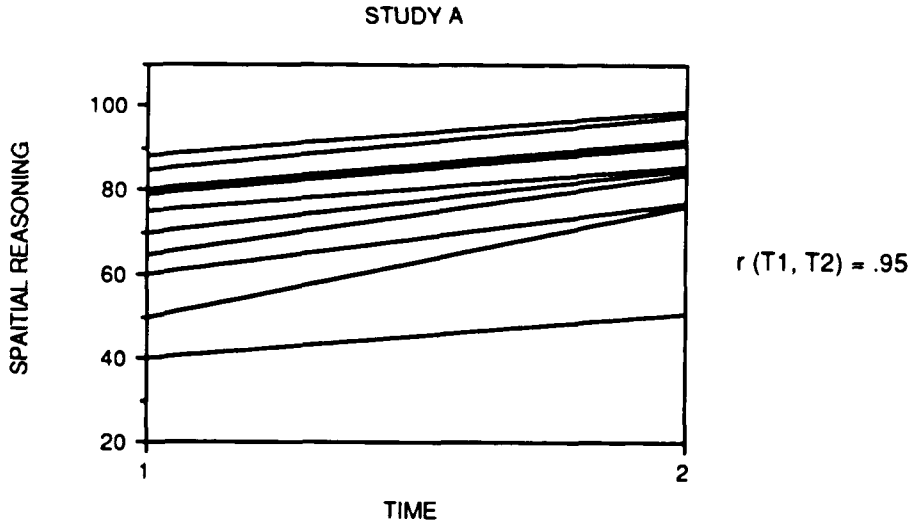
> it would appear that the critical question is not whether or not a test measures the same thing at time one and time two. The critical question is whether, taking into account whatever we know about the meanings of scores at time one and time two, individual differences in change on the test are of any interest to us. . . . Social scientists are interested in what causes attitudes to change; and if some people's attitudes change in one direction under similar conditions and others' change in the opposite direction under similar conditions, this make the subject more interesting—at the same time that it lowers the test-retest correlation. (p. 13)

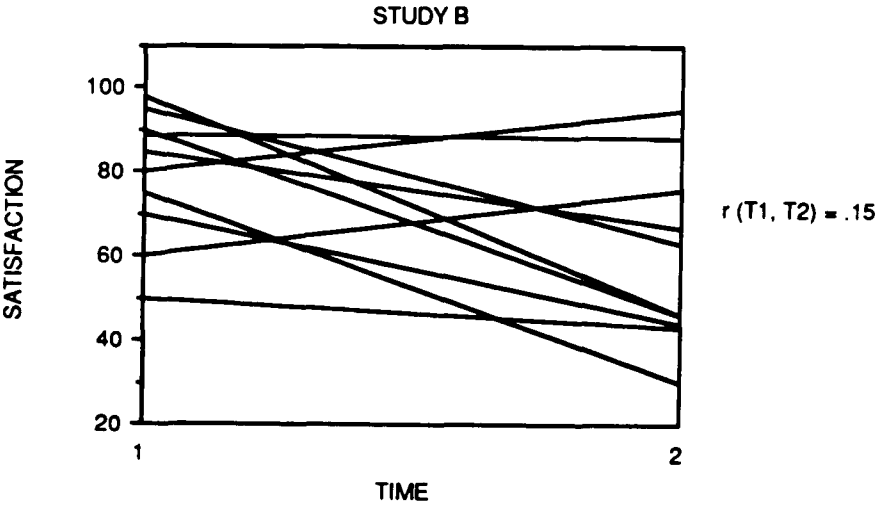
---

[4] Theoretically, a collection of difference scores could be heterogeneous yet still maintain their relative rank-order from initial to final status. However, this could occur only if the variability of the difference scores were substantially smaller than the variability in both initial and final status. This would enable individuals' difference scores to differ, but not "cross" when displayed visually (i.e., as in Figure 1).

[5] In fact the variance of the difference scores in Study B is almost 30 times greater than in Study A: σ^2 (Study A) = 23.04, σ^2 (Study A) = 605.16.

Figure 1: Comparison of two hypothetical studies of change over two time points.





Therefore, according to Bereiter, the unreliability-invalidity dilemma is often a false one because most researchers measuring change are interested in asking questions about change, not just final status, with these questions about change usually in the form of individual differences.

The Physicalist-Subjectivist Dilemma. According to Bereiter, the third dilemma facing the analysis of change is the one least likely to be addressed by developmentally oriented researchers because it strikes at the core of psychometric theory: do we assume, ignoring measurement error, that equal changes in an observed variable at various points on a measurement scale correspond with equal changes in the latent construct that is being measured? If we act is if there is a one-to-one correspondence between the observed and latent variable, then we assume a physicalist or objectivist interpretation. A move away from such a position results in holding to the assumption that measured change and latent change are, while not necessarily being mutually exclusive, to some degree independent entities. A common example of this dilemma occurs when one interprets Likert-type responses: can the dimension between "strongly agree" and "strongly disagree" be interpreted as representing some consistent (i.e., across individuals) interval scale of measurement, or should it be interpreted as achieving only an ordinal level of measurement? This interpretation becomes more complicated, however, when one is dealing with measuring change, because in addition to assuming interindividual consistency, one is faced with assuming consistency over time by taking a physicalist stance.

Far from being an esoteric and technical detail, the issue of making appropriate inferences from measurement is a central one in psychometrics. In a review of the long-lasting debate that centers on this issue, Stine (1989) concluded that "the measurement context . . . is of the utmost importance. Performing sophisticated analyses that are appropriate for one scale of measurement (e.g., interval) on data that reflect less structured scales (e.g., ordinal) yields nonsense" (p. 154). Researchers interested in examining change, therefore, are faced with the question of whether or not the technique they use to measure change requires a choice between

either alternative of the physicalist-subjectivist dilemma. Bereiter describes this as a situation in which "one has the unpleasant option of sticking with [the] particular scale units given . . . (physicalism), or else abandoning the given units in favor of others that seem to conform to some underlying psychological units (subjectivism)" (1963, p. 5).

Upon reflection it is likely that most developmentalists hold strongly to both views, each for its own purpose. The physicalist position seems indispensable for the sake of practice because it becomes exceeding difficult to efficiently measure change if developmentalists can not meaningfully compare change across time and across individuals. Likewise, the subjectivist position seems indispensable for the sake of theory because developmentalists imagine that latent psychological entities develop independently of the measurement of change. Because these two positions are important at different levels of the investigative process, the physicalist—subjectivist dilemma, in deference to Bereiter's conclusions to the opposite, is not unlike its predecessors. It too can be viewed as a false dilemma because of the utility of assuming both; in the classical pragmatic view (James, 1907/1987; McDermott, 1976), doing so allows us to continue asking productive questions about development.

Although I would argue that developmentalists who choose to hold both positions stand on firm philosophical ground, turning to the second representative author in the continuing discourse regarding the problems in measuring change, one finds that, until the mid-1970's, serious challenges still faced the investigative and methodological practice of developmental psychology.

McCall's (1977) Critique of Non-Developmental "Developmental" Methods

In an influential review paper, McCall (1977) proffered several serious methodological challenges to developmental psychology which he claimed resulted primarily from the discipline's full embrace of the experimental method. Although he conceded that research conducted in the vein of traditional experimental psychology may be appropriate for understanding some developmental issues, for the most part this approach kept developmental psychologists from adequately addressing questions about naturally occurring development.

According to McCall, the use of laboratory techniques, experimental control, and artificial situations had lead to a developmental psychology of what "can" happen in human development, which may be far from what "does" actually happen (p. 336).

Drawing on the work of other methodologists (Schaie, 1965; Wohlwill, 1973), McCall also chastised the developmental community for "not [being] sufficiently developmental," and therefore providing a "discipline [that] is developmental in name only" (p. 337). McCall felt that due to the fact that the primary focus of developmental research up to the mid-1970's had been "on specific behaviors displayed by immature organisms at a single age" (p. 337), developmental psychology's "developmental-ness" had been totally undercut. This type of psychology, perhaps better referred to as "child psychology," is only able to provide a taxonomy of the child's behavioral and mental abilities, and falls short of shedding light on the nature of many developmental processes.

When it came to describing what were appropriate methods for studying development, McCall was quick to attack the simplest developmental approach:

> if a primary mission [of developmental psychology] is to discern ontogenetic change within individuals, the sequence and timing of developmental transitions, and the changing social and environmental factors that permit development to occur, then we must use longitudinal, not cross-sectional, approaches to our subject matter (p. 337).

As conveyed in this quote, McCall's challenges to developmental psychology have two profound implications for the definition of developmental research. The first is that the success of developmental psychology depends largely on the effectiveness of its methods to measure change over time. The second implication is that developmental psychology is in essence an ideographic rather than a nomothetic science. Although developmental psychology may draw general conclusions about the nature of certain normative developmental process (e.g., the onset of puberty, the sequences of cognitive development, etc.), these conclusions can only be extrapolated from findings based on investigations of individual development. Perhaps developmental psychology should not be viewed as the extension of the field of child development to other periods of life commonly seen in "life-span" approaches. Rather, it may

be more productive to view the discipline as an extension of the psychology of individual differences that incorporates the ubiquitous dimension of continuous change.

Rogosa's (1988) Discussion of the Myths of Longitudinal Research

Reviving the discussion of the analysis of change within the educational measurement community in the 1980's, Rogosa, Willett, and their colleagues (Rogosa, 1980; Rogosa, 1988; Rogosa, Brandt, & Zimowski, 1982; Rogosa & Willett, 1985; Willett, 1988, 1989), have challenged the solutions provided by an earlier generation of methodologists including Bereiter (1963) Cronbach and Furby (1970) and Lord (1963). Rogosa (1988) described this challenge as a realization that "longitudinal research in the behavioral and social sciences has been dominated, for the past 50 years or more, by a collection of damaging myths and misunderstandings" that have impeded the development and application of useful methods (p. 171).[6]

According to this view, researchers analyzing longitudinal data have accepted without much question most methodological claims about analyzing change, including the unavoidable negative correlation between initial status, the unreliability of the difference score, and the adequacy of two-wave designs in measuring development. Rogosa's 1988 book chapter entitled "Myths About Longitudinal Research" presents a discussion of nine such myths and how to avoid them. Three of the myths most directly related to this discussion of the analysis of change include the acceptance of two-wave longitudinal designs, the inadequacy of the difference score, and the ability of correlations over time to inform about change.[7]

---

[6] It is interesting to note that Rogosa displays the influence of McCall (1977) and others (Schaie, 1965; Wohlwill, 1973) by focusing exclusively on longitudinal research, thus reframing the primary focus of research design in developmental psychology away from two-wave designs.

[7] The six additional myths described by Rogosa (1988) involve erroneous beliefs about the benefits of the residual change score, the correlation between change and initial status, the unavoidability of regression to the mean, the effectiveness of analyses of covariance in measuring change, the interpretation of stability coefficients over time, and the use of cross-lagged correlations as evidence of causal relationships.

"Myth #1: Two observations a longitudinal study make" (Rogosa, 1988, pp. 174-177). This myth—which primarily results from the popularity of the pretest-posttest design in psychological research—is true in a technical sense. However, according to Rogosa, "two observations do provide some information about change over time, but this design has many critical limitations" (p. 174). One of these limitations is that a two-wave study can provide some information regarding amount of change, but is unable to estimate the latent true growth rate, especially if that growth rate is nonlinear.

A second limitation offered by Rogosa is that "the amount of change [obtained from two observations] will often be deceptive" (p. 175). When a maximum of two values are used to measure change, the temporal choice of time points is, in Rogosa's words, "extremely consequential" (p. 175): any slight variation in when data is collected is likely to effect the amount of change measured. In addition, the two-wave change score is hopelessly contaminated by measurement error associated with each observation, because any amount of variability in either observation is passed directly and in full to the amount of change measured.

"Myth #2: The difference score is intrinsically unreliable and unfair" (Rogosa, 1988, pp. 177-180). Rogosa traces this myth to the "impressive amount of psychometric literature over the past 50 years [that] has sought to demonstrate deficiencies in the difference score" (p. 177). The most important finding made by Rogosa and his colleagues (Rogosa, Brandt, & Zimowski, 1982; Rogosa & Willett, 1985) regarding the difference score is that the wholesale acceptance of the unreliability of the difference score is based upon evidence obtained only under certain psychometric conditions. For example, most of the influential critiques of the difference score (Linn & Slide, 1977; Lord, 1963) assume homogeneity of change across individuals when they obtain weak reliability estimates for the difference score. Because estimates of reliability are principally a ratio of the variance of score and its error variance, it should not be surprising that "the difference score cannot be expected to detect" individual differences in change that does not exist (p. 178). In fact Rogosa demonstrates that when there

are individual differences in change—as there likely are in a large majority of developmental research—the difference score has adequate reliability.

"Myth #3: You can determine from the correlation matrix for the longitudinal data whether or not you are measuring the same thing over time" (Rogosa, 1988, pp. 180-182). Because of (often untrue) slander heaped upon the difference score, many researchers choose to analyze change over time by examining intercorrelations among measures obtained longitudinally, usually in the form of a variable-over-time by variable-over-time correlation matrix. The rationale undergirding the inspection of intercorrelation matricies appears to be that the correlations found between observations over time say something meaningful about change over time. Rogosa correctly points out, however, that a strong correlation "cannot 'prove' that the same thing is being measured at both ends of the observation interval, only that the ordering of individuals in the initial measure is similar to the ordering of individuals in the final measure" (p. 181-182).

Each of these myths reveals that serious limitations exist in ways in which methodologists have chosen to deal with the analysis of change. The question remains, however, how do these limitations manifest themselves in the research literature of developmental psychology? Next I will describe the advantages and disadvantages of four of the more prominent approaches used in studies of development to understand change: the difference score, the residual change score, the use of intercorrelations over time, and repeated measures analysis of variance. A comparison of these approaches will demonstrate that in order to talk meaningfully about change, developmentalists should employ a fifth approach—what Rogosa has referred to as the "hero" of the analysis of change—namely, individual growth modeling.

Conventional Methodological Approaches to Analyzing Change

Despite the efforts of Bereiter, McCall, Rogosa, and others,[8] developmental researchers have continued to use methods for analyzing change that neglect many of the central issues raised by these methodologists. When faced with analyzing longitudinal data, the developmental researcher has typically chosen from a group of methodological approaches that measure growth inefficiently, that look at group change rather than individual change, and that take segmented "snap-shot" views of change over time. These limitations are apparent in four of the most common techniques for measuring change: the difference score, the residual change score, the multiwave correlation matrix, and repeated measures analysis of variance.

Using the Difference Score as a Measure of Change

The difference or change score is the most widely used measure of change because it requires only two observations and is computed by simply subtracting initial status from final status. Not only is the pretest-post test difference score typically one of the first quasi-experimental designs that undergraduate psychology majors are exposed to, but it often appears to be nicely suited for the practical needs of intervention studies in educational and clinical research.

Despite its popularity, the difference scores is also the most widely maligned measure of change among developmental researchers. Within the methodological literature, two major drawbacks of the difference score are commonly discussed: its assumed unreliability and its negative correlation with initial status (Bereiter, 1963; Cronbach & Furby, 1970; Lord, 1963). In the review above of Rogosa's debunking of one myth surrounding the difference score, the argument was made that the difference score can be a unreliable measure of change, but only under largely artificial conditions (i.e., homogeneity of growth across individuals).

---

[8] Rogosa (1988) provides the most up-to-date survey of the various contributors to this discussion. A recent additional source of information regarding methodologies for measuring change can be found in a volume edited by Collins and Horn (1991).

Two additional limitations of the difference score lie in its utilization of only two pieces of information for measuring growth, initial and final status. One limitation outlined by both Rogosa (1988) and Willett (1988) is the inability for the difference score to measure nonlinear growth. In the same way that scientific psychologists are beginning to realize the multivariate "complexity of the 'real' world" (Tabachnick & Fidell, 1989, p. 7), it appears that they are beginning to realize its nonlinearity as well. Consequently, the difference score may be suited for certain questions about change, primarily as the result of laboratory manipulations over small periods of time. Therefore, it cannot successfully address many questions of human development. The second limitation associated with a measure based on two observations is that it tends to produce a less than robust measure of change: any slight "wobble" in either observation will effect the obtained difference score (John Willett, personal communication, October 15, 1991). What is needed, given the unavoidable measurement error associated with developmental research, is a measure of change that, in a sense, "increases the N" of observations.

The Residual Change Score: An Improvement on the Difference Score?

Despite the fact that unreliability of the difference score occurs only in certain circumstances, most researchers have been more concerned with the negative correlation between change and initial status than its reliability. Because of this concern, many researchers have embraced the use of the residual change score as a measure of change that ignores differences in initial status despite serious questions about its utility. Describing these questions, Willett (1988) contends that with the residual change score "there is disagreement as to exactly what is being estimated, how well it is being estimated, and how the outcomes of the estimation can be interpreted" (p. 380). Subsequently, Willett reports that "among methodologists . . . the residual change score has now been largely discredited as a measure of individual change" and that "in addition to the many technical and practical problems that arise in its application, there are also issues of logic and substance" that cloud this method (p. 380, emphasis in original).

For example, partialling out the effects of initial status results in treating all individuals as equivalent in initial status. In terms of most questions of development, what is gained by assuming initial equality? Unfortunately, not much. By controlling for initial status the residual change score ignores important interindividual differences that are highly dependent on when initial status is measured. Because initial status is typically obtained at some arbitrary point in time, it is unclear how consistent the residual change score would be if initial status was obtained at some other point in the developmental process.

In addition to the myth of the unreliability of the difference score and the deficiencies of the residual change score, a measure of change based on only two observations is restrictive because it only allows for the estimation of linear change. In light of this limitation, Rogosa offered the motto that that "two waves of data are better than one, but maybe not much better" (1982, p. 744). Nonetheless, because two-waves of data are relatively easy to collect, the wide acceptance of the pretest-posttest design, and the computational simplicity of the difference score, it is unlikely that the difference score will vanish from psychological research. However, it is the hope of methodologists interested in measuring change that researchers take the extra time and effort to obtain multiwave data in order to meaningfully address the nature of developmental processes.

Multiwave Intercorrelation Matrix

One common approach to measuring change over multiple waves is to compute intercorrelations among variables as a function of time. Although several of the limitations of this approach offered by Rogosa (1988) have been discussed above, two additional problems deserve to be mentioned.

The first of these is that in looking at a series of correlations over time, one only gains information about change at the level of the group, not the individual. As summaries of the maintenance of rank order within a group from observation to observation, correlations are more than adequate for examining the within-group stability of a variable at one observation and another. Although such summaries are helpful for addressing such research questions, they

are not sufficient for addressing questions about individual differences in development. This limitation is compounded by the second limitation that examining correlations over time provides only a segmented, "snap-shot" view of change over time, and thus is unable to address the continuous nature of underlying true growth posited by most developmental researchers.

Repeated Measures Analysis of Variance Procedures

Much of what limits the correlation matrix approach is also applicable to the use of repeated measures analysis of variance (ANOVA) procedures. The central problem with repeated measures ANOVA is that it a methodologically excellent approach for talking about some types of change (Hertzog & Rovine, 1985), but unfortunately these are not the questions that developmentalists tend to ask. Most commonly repeated measures ANOVA is used to address the existence of an overall main effect for some nominal grouping variable (i.e. gender, experimental condition, etc.), the overall main effect of differences in the dependent measures over time, and their interaction. If developmentalist psychologists take McCall's (1977) criticisms seriously, can the use of repeated measures analysis of variance be any help in addressing individual development? The use of broad grouping variables does not begin to address questions of individual development, and choosing to test for the main effect of the individual is assured to be rejected. Therefore, repeated measures ANOVA is only appropriate for addressing questions about differences in the average change among groups of individuals, and is ill suited for addressing questions about individual development.

Summary

Table 1 presents a summary of the advantages and disadvantages of five approaches to analyzing change typically used by developmental researchers. Although each of the initial three alternatives to the standard difference-score approach provide advantages for analyzing change, each is still hindered by several disadvantages. Aside from psychometric problems, these disadvantages center around two major limitations: the inability to measure interindividual differences in change using multiwave data and the inability to estimate

Table 1. Comparison of advantages and disadvantages of five methods for analyzing change.

Method	Advantages	Disadvantages
Difference score	1. Requires only two waves of data 2. Ease of computation	1. Often produces biased and unreliable estimates of true change 2. Sensitive to the effects of outliers 3. Restricted to linear modeling 4. Unable to measure change in multiwave data
Residual Change Score	1. Requires only two waves of data 2. Uncorrelated with initial status 3. Widely accepted as a solution to problems associated with the difference score	1. Often produces biased and unreliable estimates of true residual change 2. Sensitive to the effects of outliers 3. Restricted to linear modeling 4. Unable to measure change in multiwave data 5. Unclear what is being controlled for if Time 1 is arbitrarily chosen 6. Ignores meaningful interindividual differences in initial status
Multiwave Correlation Matrix	1. Can be used with multiwave data 2. Ease of computation and interpretation 3. Provides information regarding reliability of measurements over time	1. Provides no information regarding interindividual differences in change 2. Cannot be used to estimate underlying true growth rates due to segmentation of time variable
Repeated Measures Analysis of Variance	1. Can be used with multiwave data 2. Allows for the comparison of group difference in time main effects 3. Nonlinear trends can be analyzed	1. Provides no information regarding interindividual differences in change 2. Cannot be used to estimate underlying true growth rates due to segmentation of time variable
Individual Growth Modeling	1. Can be used with multiwave data 2. Can be used to model nonlinear growth 3. Allows for the analysis of intraindividual differences over time and the analysis of systematic interindividual differences in true growth 4. Estimates true growth from observed growth 5. Precision of estimates of true growth can be increased by weighted least squares analysis 6. Can accommodate variability in measurement times (in both number and spacing)	1. Limited by the specification of a single growth model

underlying true rates of growth. Although the fifth method for analyzing intraindividual change—individual growth modeling—is described in the following paragraphs, it is included in Table 1 in order to emphasize its superiority over traditional methods in situations where developmentalists are interesting in addressing questions about the nature of invidivual growth over time.

Individual Growth Modeling: Describing and Predicting Individual Growth in Multiwave Data

Several of the reoccurring problems in the analysis of change are addressed by the recent development of a family of quantitative techniques that will be referred to here as individual growth modeling or IGM. The power of these approaches lies in their ability to both reliably measure interindividual change and to predict interindividual differences in change. Although the basic computational procedures used in growth modeling can be found in many textbooks on introductory statistics (Glass & Hopkins, 1984; Hays, 1981; Kirk, 1990), this application of individual growth modeling is a recent development. As it will be explained below, the primary advantage of individual growth modeling is its ability to successfully avoid two limitations of conventional approaches: the inability to deal efficiently with multiwave data and the inability to measure change at the individual level at which it logically occurs.

Developmental psychology has seen a significant increase in the discussion of this approach in the last three years, as articles describing and using growth modeling have appeared in American Psychologist (Fletcher, Francis, Pequegnat, Raudenbush, Bornstein, Schmitt, Brouwers, & Stover, 1991), Child Development (Burchinal & Applebaum, 1991), Developmental Psychology (Huttenlocher, Haight, Bryk, Seltzer, & Lyons, 1991) Journal of Consulting and Clinical Psychology (Francis, Fletcher, Stuebing, Davidson, & Thompson, 1991; Willett, Ayoub, & Robinson, 1991), Psychological Bulletin (Singer & Willett, 1991), and Psychology and Aging (Aldwin, Spiro, Levenson, & Bossé, 1989). A general overview of growth modeling will be given next, with an examination of the limits of this method for analyzing change over time.

An General Overview of the Two-Step Individual Growth Model Approach

The two most prominent approaches to growth modeling—Willett's (1987-1991, 1988) weighted least squares approach and Bryk and Raudenbush's (1987, in press) hierarchical linear modeling—both use a two-stage technique to describe and predict individual growth. Because the primary difference between these two approaches is in their computational methods (Willett, 1991), the approach developed by Willett and his colleagues (Rogosa & Willett, 1985; Sayer & Willett, 1991; Singer & Willett, 1991; Willett, 1987-1991, 1988, 1989, 1991; Willett, Ayoub, & Robinson, 1991) will be dealt with here.

Step 1: Describing Intraindividual Differences in Growth

Although individual growth can be modeled on two waves of data, it is preferable to obtain data from more than two time points. Multiwave data allows the analyst to model growth using an infinite number of nonlinear functions, and as Willett (1989) has shown, increasing the frequency of observations dramatically increases the reliability of the measurement of change. Conceptually this is analogous to the argument for increasing sample size to ensure the reliability of a statistical test: as the size of a random sample approaches the size of the population, sampling error decreases. If we replace our usual interpretation of randomly sampling individuals from a population with the notion of randomly sampling time points from a population, then it becomes clear that as we increase the number of observations (time points), then measurements become more reliable.[9]

Given that we have obtained multiwave data, the first step in individual growth modeling involves performing a series of bivariate regression analyses for each individual regressing the dependent variables of interest on time. The simplest regression technique computes a linear model equation by fitting an ordinary least squares (OLS) regression line for each

[9] Currently it is unclear how important it is to truly sample time points randomly. Although most growth modeling approaches do not require sampling the same time points for each individual, it is assumed that these techniques are rather robust to violations of random sampling time points.

individual in the analysis. The conventional equation for a linear model is derived from traditional true score theory and is presented by Willett (1988, Equation 1) as:

$$X_{ip} = \xi_p (t_i) + \varepsilon_{ip} \quad (1)$$

where X denotes the dependent measure, i refers to the occasion of measurement, p refers to the individual, the expression symbol $\xi_p (t_i)$ represents the true status of the variable X at a given time, and ε represents measurement error unique to each individual at a given time point. The primary parameter of interest in this equation is $\xi_p (t_i)$, the slope of the regression line measuring the observed rate of change in a dependent variable for a given individual over time. This estimate will be referred to here as an individual's linear growth rate. Linear growth rates can be both positive and negative and are interpreted just as conventional slope coefficients are (i.e., the amount of change in one variable with one unit change in a second variable). Other useful values obtained in each individual regression analysis include the intercept, standard error of estimate, and proportion of variance accounted for by the regression equation (r^2).

Although the linear growth model is the simplest model and may be an appropriate model given the data at hand, it does not take full advantage of one of the advantages of collecting multiwave data: the investigation of nonlinearity. With the popularization of nonlinear statistical techniques in psychology, researchers have become sensitized to the usefulness of nonlinear models. This is equally true when analyzing change over time: the assumption of a continuous rate of change for growth can be a limiting one, especially when one is modeling developmental processes that are theoretically discontinuous.

In order to chose an appropriate growth model, a useful strategy described by Willett (1988) is to compare the overall fit of various nonlinear growth models with that of a simple linear model. An examination of the difference in r^2 values for each individual regression, or a comparison of the mean r^2 for each model, will suggest which model best describes the multiwave data. It is statistically true that the more nonlinear a model is, the greater the goodness of fit will be for any type of multiwave data. It is important to keep in mind,

however, that the greater the degree of nonlinearity, the more complex the interpretation of the growth models becomes. For example, by fitting a linear model, one can interpret the individual differences in growth among individuals in terms of a constant rate of change (i.e., the slope coefficient). Differences among individual growth based on a quadratic model, on the other hand, have to be interpreted in terms of the rate of change as a function of time; in other words, at any given point in time growth may be accelerating or decelerating at any given rate. Obviously, this difference becomes increasingly more complex as one fits higher-order polynomials such as cubic or quartic trends. Because of the competing benefits of goodness of fit and interpretability, one must try to maximize both when choosing among growth models.

To summarize, the major goal of the first step in individual growth modeling is to obtain a measure of true growth rate for each individual in the form of a slope coefficient. Inspection of individual scatterplots of the variables over time and results from individual regression analyses (i.e, r^2) can provide the information necessary to evaluate the appropriateness of the growth model chosen.

Step 2: Accounting for Interindividual Differences in Growth

The second step in growth modeling involves examining for the existence of any systematic differences among the individual growth rate obtained in the previous step. By treating the estimated growth rates as a dependent measure, this is accomplished by attempting to account for variance among the growth rates through any number of standard inferential techniques. Analysis of variance can be used to determine if modeled rates of change differ significantly according to some nominal variable, or any number of correlation-based techniques, including multiple regression (Buchanan, 1991) or structural equation modeling (Sayer & Willett, 1991), if one is interested in predicting growth rates from interval-level data.

In Willett's approach, one of the central benefits of analyzing multiwave data is that the measurement error associated with fitting individual growth models can be assessed and corrected for. In order to correct for such error, Willett suggests a Weighted Least Squares (WLS) regression approach that provides a best linear unbiased estimator of growth, or BLUE

(SAS Institute, 1985), by weighting each individual growth rate by the precision of fit of the growth model. Individual weights are found by using a modified formula first presented by Hanushek (1974, cited in Willett, 1988) which computes values based upon the size of the standard error of estimating the growth rate. This results in larger weights (approaching 1.0) for individual that are best described by the specified growth model, and smaller weights (approaching 0.0) for individuals who deviate from the specified growth model.

In conducting the second step of individual growth modeling, the search for covariates of change may not necessarily allow developmental researchers to directly address the question "What causes differences in growth?" It does allow for the prediction of individual differences in change. Answering McCall's (1977) challenge to the science of developmental psychology, this ability to account for individual differences in growth holds great promise to a discipline that is interested in understanding the impact of biological, environmental, and social forces on individual change.

Limitations of Individual Growth Modeling

One important limitation of individual growth modeling is the forced specification of a single model type. Although there is a theoretically infinite number of growth models possible (i.e., linear, quadratic, cubic, etc.), current methods require the specification of a single model for use in the first step. The result of this requirement is that each individual's growth record is assessed according to some a priori model, and any individual whose growth does not fit the model is treated as deviating from that a priori model. Burchinal and Applebaum (1991) have made the argument that this is a hidden assumption of the individual growth modeling approach, and one that can have serious limitations if there is substantial heterogeneity in the overall shape of individuals' developmental functions.

Although correcting for error as the result of measurement can be achieved by using a weighted least squares regression model, it is unclear how to correct for error that results from a misspecification of the true growth model. In some empirical situations, the assumption of homogeneity of the shape of developmental functions may be warranted. One example of the

tenability of this assumption can be found in the study of individual differences in the rate of early language acquisition. McCartney, Buchanan, and Jordan (in preparation) have found an average r^2 of .84 using linear model to describe individual growth in children's mean length of utterance from 21 to 28 to 35 months of age. Other empirical situations do not appear to be as well described by the use of a single model. For example, in a previous study of maternal depression, I (Buchanan, 1991) found that the use of a single linear model produced an average r^2 of .33. Given what is known about the developmental paths of language acquisition and depression, it should not be surprising that the former closely fits a linear model and the latter does not.

The issue that lies at the heart of Burchinal and Applebaum's (1991) criticism of the use of a single growth model is that if we suspect that individual differences in the rate of change exist, why not suspect that there may also be individual differences in the overall shape of change? In certain empirical situations the application of a single growth model may cloud important differences in the nature of true growth.

Patterns of Individual Growth: Accommodating Interindividual Differences in Growth Models

One solution to the limitation of specifying a single growth model has been developed by Rovine (M. Rovine, personal communication, February 10, 1992) and utilized in a study of marital relations during the transition to parenthood (Belsky & Rovine, 1990). This technique Combines Willett's IGM approach with the work of other methodologists (von Eye & Nesselroade, 1988 cited in Belsky & Rovine, 1990; Bishop, Feinberg, & Holland, cited in Belsky & Rovine, 1990), in order to allow each individual to be fitted with any number of growth models (i.e., linear, quadratic, cubic, etc.), which are then evaluated for their appropriateness using a series of contrast coding coefficients. This enables each individual to be characterized not just as more or less linear for example, but in terms of some combination of growth models. For example a positively linear-negatively quadratic growth model is one in which there is a general upward decelerating trend. Once such individual patterns of change are obtained, they are then categorized along some coding scheme. Systematic differences in group

membership in each pattern of change can then be examined using such multivariate techniques as cluster analysis and discriminant function analysis. Answering Burchinal and Applebaum's (1991) criticism of individual growth modeling, Rovine's extension makes it possible for the analysis of individual development in domains that tend not be best described using a single growth model.

Summary

This discussion of the individual growth modeling approach can be summarized by stating three general methodological claims made by its proponents (Bryk & Raudenbush, 1987; Burchinal & Applebaum, 1991; Rogosa, Brandt, & Zimowski, 1982; Willett, 1988):

1. By using intraindividual growth as the level of analysis, IGM allows researchers to address change at the natural and logical level at which it occurs. This ideographic approach avoids the use of gross grouping variables such as gender, experimental condition, or socioeconomic status to describe both intraindividual and interindividual differences in growth.

2. Observed growth rates derived from multiwave data provide the most stable and valid estimates of true growth. A corollary to this is that reliability and validity increase dramatically with each additional wave of data collected (Willett, 1989). This can be illustrated by the fact that individual growth rates are less influenced by single data points as the number of total data points increases. This is one of the major advantages of IGM over the two-wave difference score.

3. Care should be taken in the choice of individual growth models. As argued by Burchinal and Applebaum (1991) the choice of a single growth model (i.e., linear, quadratic, etc.) allows for the measurement of individual differences in growth rate, but is unable to measure individual differences in the overall developmental shape of growth. The choice of growth models should always be informed by theory and prior empirical evidence, with an

eye to maximizing both the amount of variance accounted for by the model and its interpretability. When faced with a wide range of variability in the shape of observed growth, an alternative grouping procedure (Belsky & Rovine, 1990) may be useful.

III. STUDY 1: ASSESSMENT OF FRESHMAN ORIENTATION PARTICIPANTS (JUNE 1990)

As part of a university orientation program in June 1990, an initial study was conducted to assess the psychometric properties of several instruments designed to measure affect, mood, and epistemological orientation. The goal of this study was to provide an empirically sound collection of instruments for use in a longitudinal study of epistemological development during the transition to college. In addition to psychometric assessment, this study provided a unique pre-college cohort for cross-sectional comparisons with first-semester students participating in later longitudinal studies.

The primary objective of Study 1 was to evaluate four psychological scales, Perry's Checklist of Educational Views (1968), the State-Trait Anxiety Inventory (Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983), Mayer and Gaschke's Brief Mood Introspection Scale (1988), and a modified version of Levenson's (1973) Multidimensional Locus of Control Scale developed by Roberts and Nesselroade (1986). This involved investigating the distribution of scores, internal reliability, and factor analytic structure of each scale, comparing their results to previously published findings, and evaluating the construct validity of each scale. An additional concern was the possibility of any demographic differences in this sample, primarily as a function of gender and home residency status (i.e., in-state or out-of state). These two demographic variables were viewed as salient because of recent discussion regarding possible gender differences in epistemological development (Baxter Magolda, 1990) and the large percentage of out-of state residents in this population (approximately 40%).

Measuring Epistemological Orientation

In his 1970 book, Perry argued that the initial epistemological stance of many students entering college is one in which knowledge is viewed as a set a concrete facts dictated by Authority-oriented experts and teachers. This initial attitude, however, is transformed as the result of pluralistic academic and social experiences and is replaced by more flexible, relativistic assumptions about knowledge that accommodate diversity and uncertainty. In order

to measure the degree to which an individual holds certain epistemological assumptions, a shortened version of Perry's original Checklist of Educational Views (1968, CLEV) was constructed as a measure of adherence to dualism.

Development of the Checklist of Educational Views

Perry and his colleagues originally administered a 90-item version of the CLEV to students entering Harvard University in Fall 1953, and later constructed a shortened version containing 46 items in the Fall 1954. Factor analyses of the responses of 1954 sample yielded a three-factor solution accounting for one-third of the total item variance, with internal reliability estimates between .66 to .73 (Perry, 1968, p. 103). In addition, an outcome study conducted by Perry and his colleagues supported the construct validity of the shortened version of the CLEV (Perry, 1968). From the Fall 1954 sample, students scoring high on the scale were found to take course loads heavier in the natural and military sciences, and were more likely to withdraw from college citing academic difficulty. Conversely, students scoring low on the scale exhibited higher academic achievement, obtained higher SAT scores, and performed better on a test of reading ability. From the perspective of Perry's scheme, these differences in academic performance reflect the extent to which students are able to accommodate their assumptions about the nature of knowledge to the epistemological goals of liberal education curricula.

In a series of follow-up studies conducted in the Fall of 1958 and 1959, however, the CLEV failed to distinguish among students in terms of academic performance, aptitude, and course choice. Perry interpreted this failure to replicate the 1954 findings as a product of a "marked rise and compression of scholastic aptitude" in the later students. Because of this disconfirmation and subsequent criticisms by faculty and administrators overseeing the Harvard studies that many of the CLEV items were too ambiguous, Perry and his colleagues dropped the this scale as a measure of intellectual development and replaced it with a less structured, semi-clinical interview.

Perry did concede, however, that the CLEV might be applicable to more heterogeneous populations, and indeed a modified version of the CLEV has been successfully used to account for differences in public college student verbal ability (1984a; Ryan, 1984b). Using a seven-item version of the CLEV, Ryan found that dualistic and relativistic students used different criteria for evaluating both text comprehension (1984b) and the coherence of prose (1984a). Ryan's version of the CLEV had a strong test-retest reliability, $r(34) = .84$, $p < .001$, and was found to be positively correlated with students' responses to a series of dualistic claims (1984b).

Because of past success with the CLEV as a measure of epistemological dualism, another version was created for repeated use in a longitudinal study of epistemological development. This version was constructed by choosing 20 items with the highest single-factor loadings reported by Perry (1968, Tables 16-17). Each of the 20 items had factor loading scores that ranged from .70 to .38; Perry interpreted this factor to be a measure of the latent construct of adherence to dualistic beliefs. Because these factor loadings were computed across different samples obtained in 1954 and 1959, the factor structure of the twenty-item version will be assessed in Study 1. Each item was scored on a six-point scale ranging from strongly agree to strongly disagree, and a composite "dualism" score for each respondent was computed by summing responses, resulting in a range of possible scores from 20 to 120.

Demographic Differences in Epistemological Development

Gender and home residency status were two demographic variables chosen as possible covariates of epistemological dualism. Although numerous studies of Perry's scheme have failed to find differences between male and female college student epistemological development (Baxter Magolda, 1990), persistent theoretical questions regarding the gender specificity of "ways of knowing" have been sparked by the work of Gilligan (1977, 1982, 1988) and others (Belenky, Clinchy, Goldberg, & Tarule, 1986). Therefore, dualism scores obtained in Study 1 will be analyzed for gender differences.

The dualism scores of in-state students will be compared with those obtained from out-of-state students to investigate the degree to which students' residency status may be related to

epistemological orientation. Because 36% of the student population retains primary residency outside of New Hampshire (Factbook FY 91, 1991), this variable provides a convenient blocking variable. In addition, numerous differences commonly found in this population related to residency status—including academic achievement, socioeconomic status, and high school preparedness—have logical connections to possible interindividual differences in epistemological orientation. It is hypothesized, however, that no differences in epistemological orientation will be found as a function of residency status.

Epistemological Dualism and Locus of Control

In an attempt to further explore the construct validity of the 20-item CLEV, a scale measuring locus of control, the extent to which individuals attribute control of events to internal or external forces, was also chosen for use in Study 1. In a study of career development, Knefelkamp and Slepitza (1976) have suggested a link between locus of control and epistemological dualism: "Lower stage dualistic thinking students are almost exclusively controlled by externals in their environment. Adhering to the belief that there is only one right career for them, they tend to turn to parents, teachers, counselors, [and] interest inventories . . . to define both the self and the right career decision" (p. 54). In an empirical investigation of 35 undergraduates, Knefelkamp and Slepitza (1976) reported that only 32% of dualists were "motivated to deal with career planning from an 'internal need'" whereas 67% of students identified as making the transition from dualism to multiplicity expressed the same motivation (p. 57).

Locus of control was measured using Roberts and Nesselroade's (1986) modified version of Levenson's (1981) Multidimensional Locus of Control Scale (LOC). Levenson's original scale was constructed to measure one dimension of internal locus of control and two external dimensions, one viewing events as being controlled by external "chance" forces and the other by external forces attributed to "powerful others." For a study examining the factor structure of the LOC scale over repeated administrations, Roberts and Nesselroade (1986) created a modified version of Levenson's scale consisting of 17 items measuring internal, external-

chance, and external-powerful others locus of control. Due to the emphasis placed by Perry (1968; 1970; 1981) on the central role of external authority in the acquisition of knowledge by students learning within a dualistic framework, a positive correlation was hypothesized between the external-powerful others dimension of locus of control and dualism. This result would provide evidence of the construct validity of the twenty-item version of the CLEV as a measure of one important dimension of Perry's conception of epistemological dualism.

Measuring Affect and Mood

Affect and mood were measured according to the generic-specific distinction made by Mayer and Salovey (1988). In this view, affect is related to general "generic arousal" and mood is related to more specific "feeling states" (Mayer & Salovey, 1988, p. 88). Although both reflect transient states experienced by individuals, the important difference lies in their specificity; accordingly, affect will be defined in this investigation as a non-specific feeling of arousal or non-arousal, and mood will be defined as related to specific feeling states (i.e., happy, sad, content, calm, etc.).

To obtain a measure of student affect, the state-anxiety subscale of the State-Trait Anxiety Inventory (STAI-Y) was used to measure general anxiety or the presence of "transitory feelings of fear or worry" (Chaplin, 1984, p. 626). The STAI-Y also produces a measure of trait anxiety, the "relatively stable tendencies [of individuals] to respond anxiously" to stressful situations (Chaplin, 1984, p. 626). Both subscales contain 20 items and have been shown to possess both adequate internal reliability estimates (ranging from .65 to .86.) and significant construct validity (Chaplin, 1984).

The 17-item Brief Mood Introspection Scale (Mayer & Gaschke, 1988, BMIS) was used to measure four specific dimensions of mood: Pleasant-Unpleasant, Arousal-Calm, Positive-Tired, Negative-Relaxed. In addition, the BMIS provides an overall measure of mood scored on a scale from -20 (very unpleasant) to +20 (very pleasant). The BMIS was chosen because of its strong psychometric properties (Mayer & Gaschke, 1988) and its ease of administration. For the four subscales of the BMIS, Mayer and Gaschke report strong internal reliability

estimates, with Cronbach's alpha coefficients ranging from .76 to .83; one exception, however, is a weaker internal reliability estimate for the Arousal-Calm dimension, with a Cronbach's alpha coefficient of .58 (1988, p. 104).

The psychometric performance of the STAI-Y and BMIS with this sample will be investigated by computing internal reliabilities estimates and assessing the factor structure of each scale. As evidence of construct validity, significant correlations between anxious affect and each the four mood states are expected, with positive correlations predicted for state anxiety and negative and aroused moods, and negative correlations for state anxiety and pleasant and positive moods. Similar to epistemological orientation, scores for each subscale will also be examined for group differences related to gender and residency status. No differences in mood and affect are expected between these groups.

Method

Participants

Study 1 was conducted during the summer of 1990 using 108 incoming college students recruited from the University of New Hampshire's June Freshman Orientation program. The mean age of the 60% female sample was 17.7 (SD = .61). The primary function of the two-day orientation program was to introduce incoming students to university life, both social and academic, as well as advise and pre-register them for the Fall 1990 academic semester. Each incoming student attended one of six orientation sessions held at the University of New Hampshire between June 1-29, 1990.

Procedure

As part of the pre-registration process, incoming students were asked to volunteer for a study of student attitudes towards college. After reading and signing an informed consent form, volunteers were asked to complete a questionnaire packet including the following instruments: the Checklist of Educational Views (Perry, 1968), the State-Trait Anxiety Inventory (Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983), the Brief Mood Introspection Scale (Mayer & Gaschke, 1988), and Roberts and Nesselroade's (1986) version

of Levenson's Multidimensional Locus of Control Scale. In addition, several demographic variables were obtained from each subject, including age, gender, residency status (i.e., in-state versus out-of-state), and academic major.

The following dependent variables and their abbreviated names were computed from the four scales administered in Study 1: dualism (DUALISM), internal locus of control (INTERNAL), external-chance locus of control (CHANCE), external-powerful others locus of control (OTHERS), pleasant-unpleasant mood (PLEASANT), aroused-calm mood (AROUSAL), positive-tired mood (POSITIVE), negative-relaxed mood (NEGATIVE), overall mood (MOOD), state anxiety (STATE), and trait anxiety (TRAIT).

Results

Preliminary Analyses

Descriptive statistics for 11 dependent measures are presented Table 2, including sample size, means, standard deviations, distribution characteristics, and internal reliability estimates. Univariate analyses revealed means and standard deviations similar to previously published results for the locus of control subscales (Levenson, 1973), the anxiety measures (Nezu, Nezu, & Nezu, 1987), and the mood scale (Mayer & Gaschke, 1988). Exceptions to this were found the mean scores on two of the BMIS subscales and one of the LOC subscales. Students in this sample reported significantly higher mean scores on the BMIS-AROUSAL (Mean = 19.34, SD = 3.85) than the undergraduates used in Mayer and Gaschke's (1988) study (Mean = 17.50, SD = 4.39), t (571) = 4.31, p <.001, effect size (r) = .17. Likewise, students in Study 1 reported higher BMIS-POSITIVE scores (Mean = 9.32, SD = 3.95) than Mayer and Gaschke's subjects (Mean = 7.92, SD = 3.98), t (572) = 3.40, p <.001, effect size (r) = .14, measures. Students participating in Study 1 also scored significantly lower on internal locus of control (Mean = 28.27, SD = 4.90) compared to a norms reported in Levenson's 1973 study (Mean = 35.5, SD = 6.30), t (574) = - 3.52, p < .001, effect size (r) = .24.

Table 2. Descriptive statistics of dependent variables measured in Study 1.

Dependent variable	N	Mean	SD	W[a]	p	Internal Reliability[b]
DUALISM	108	64.55	13.32	.984	.737	.83
INTERNAL	121	28.27	4.90	.933	< .001	.71
CHANCE	121	21.95	4.40	.976	.342	.37
OTHERS	121	17.07	4.46	.967	.048	.48
PLEASANT	114	8.84	6.67	.966	.047	.83
AROUSAL	116	19.34	3.85	.977	.320	.59
POSITIVE	117	9.32	3.95	.967	.055	.77
NEGATIVE	118	7.33	2.51	.932	< .001	.76
MOOD	119	6.32	2.79	.880	< .001	. [c]
STATE	114	38.66	10.63	.958	.008	.92
TRAIT	117	39.96	9.10	.961	.014	.90

[a] Shapiro-Wilk's W statistic (Shapiro & Wilk, 1965) and associated two-tailed probability.

[b] Cronbach's alpha coefficient

[c] Value could not be computed from available data

Based on a test for departure from normality (Shapiro & Wilk, 1965), the distributions of four variables were found not to be normally distributed. Significant Shapiro-Wilk's W statistics were computed for measures of internal locus of control, overall mood, negative mood, and state anxiety. Visual inspection of these distributions revealed that internal locus of control and overall mood scores were negatively skewed, and that negative mood and state anxiety scores were positively skewed.

Psychometric Analyses

Seven of the ten composite measures revealed adequate internal reliabilities, with Cronbach's alpha coefficients greater than .70. Weak internal reliabilities were obtained for three measures, the Arousal-Calm dimension of the BMIS and the two external locus of control dimensions. These lower reliabilities appear to be consistent with past psychometric assessments of the BMIS (Mayer & Gaschke, 1988), but are much lower than the internal reliabilities reported by Roberts and Nesselroade for the LOC subscales (1986).

The results of principle component factor analysis for the BMIS are presented in Table 3, and the factor structure of the BMIS in Study 1 was found to be, for practical purposes, identical to Mayer and Gaschke' s (1988) original results. A two-factor solution accounted for more than one-third of the total item variance for both unrotated and varimax rotated solutions.

Table 4 compares the factor structure of Perry's original samples and the responses obtained in Study 1. If two items with factor scores below the .35 cut off used to construct the 20-item version of the CLEV are ignored (i.e., #16 and #18), there is a strong positive correlation between these two solutions, r (20) = .69, p = .001.

Results of a principle components factor analysis of the locus of control measure constructed by Roberts and Nesselroade (1986) are presented in Table 5. Although a forced three-factor varimax rotated solution accounted for 43% of the total item variance of LOC items, only nine items are characterized by single-factor or simple structure loadings, whereas eight produced complex structures by loading strongly on more than one factor.

Table 3. Rotated and unrotated factor loadings from principle factor analysis of the BMIS (N=119).

	Unrotated		Rotated	
ITEM	PLEASANT (I)	AROUSAL (II)	POSITIVE (I')	NEGATIVE (II')
Lively	.61	.48	.77	- .05
Peppy	.54	.40	.67	- .06
Active	.61	.42	.74	- .07
Happy	.78	.06	.61	- .48
Loving	.50	.48	.69	.02
Caring	.48	.40	.62	- .03
Drowsy	- .47	- .28	- .54	.10
Tired	- .41	- .06	- .34	.22
Nervous	- .45	.49	- .01	.66
Calm	.51	- .29	.18	- .56
Gloomy	- .64	.43	- .19	.55
Fed up	- .61	.33	- .23	.65
Sad	- .37	.64	.15	.72
Jittery	- .42	.50	.03	.65
Grouchy	- .60	.20	- .31	.56
Content	.70	.01	.53	-.46
Variance explained	31%	15%	26%	11%

Note: BMIS = Brief Mood Introspection Scale (Mayer & Gaschke, 1988)

Table 4. Comparison of Perry (1968) and Study 1 single factor loadings from principle factor analysis of the CLEV (N = 116).

	Perry (1968) Factor Loading[a]	Study 1 Factor Loading
1954 CLEV Items (N = 457)		
If professors would stick more to the facts and do less theorizing one would get more out of college.	.70	.61
College professors should remember more often that people of action are more important in a society than intellectuals and artists	.59	.59
Educators should know by now which is the best method of teaching, lectures or small discussion groups.	.58	.58
Students sometimes get rebellious ideas, but as they get older they ought to get over them and settle down.	.57	.59
Putting a non-conformist in a position where he/she can influence students isn't a good idea.	.57	.57
There is nothing more annoying than a question that may have two answers.	.57	.63
It is a waste of time to work on a problem that has no possibility of coming out with a clear-cut and unambiguous answer.	.56	.61
It is a pretty callous student who feels anything but love and gratitude to his or her parents.	.54	.46
There is no point having professors from foreign countries teach if they won't learn to speak English well.	.51	.54
The best thing about science courses is that most problems have only one right answer.	.47	.50
The worst thing about a lazy student is that he/she is letting his/her parents down.	.45	.41
It is annoying to listen to a lecturer who cannot seem to make up his/her mind as to what he/she really believes.	.45	.54

Table 4. (Continued)

	Perry (1968) Factor Loading[a]	Study 1 Factor Loading
It helps a child in the long run if he/she is made to conform to his/her parents ideas.	.43	.48
Any student who needs psychological counseling should not come to college.	.43	.39
It is only right to think that one's own college is the best.	.38	.54
1959 CLEV Items (N = 60)		
In the final analysis, the student who skips class is throwing away good money.	.61	.31
The inspiring teacher puts across to students things as they really are.	.59	.52
We all have the tendency to make judgments which are too simple and final: it is the goal of education to make judgments more complex and tentative.[b]	.49	.13
Students must first master what is already known before they are told to exercise their own judgment.	.47	.51
A good teacher's job is to keep his/her students from wandering from the right track.	.46	.54
Variance explained	30%	27%

[a] Source: Perry (1968), Tables 16-17.
[b] Reverse-scored item
Note: CLEV = Checklist of Educational Views (Perry, 1968)

Table 5. Rotated factor loadings from principle factor analysis of LOC (N = 121).

Item	Hypothesized Dimension	Factor 1	Factor 2	Factor 3
I feel fortunate	Chance	.64	.02	.01
I feel plans are being completed	Internal	.55	- .08	.04
I feel in control	Internal	.38	- .52	.01
I feel like a follower	Powerful Others	- .38	.65	.06
I feel destined	Chance	.28	.12	.40
I feel I can get what I want	Internal	.62	.00	.08
I feel a need to please	Powerful Others	.18	.35	.55
I feel a need to take care of others	Powerful Others	.22	- .28	.63
I feel lucky	Chance	.50	- .07	.12
I feel decisive	Internal	.45	- .40	.03
I feel accident prone	Chance	- .20	- .04	.69
I feel a sense of accomplishment	Internal	.65	- .11	.00
I feel fatalistic	Chance	- .28	.28	.45
I feel a need to fit in with others	Powerful Others	.27	.66	.04
I feel controlled by others	Powerful Others	.05	.75	.07
I feel driven	Chance	.64	.14	- .03
I feel competent	Internal	.58	- .35	- .19
Variance explained		20%	14%	9%

Note: LOC = Multidimensional Locus of Control Scale (Roberts & Nesselroade, 1986, adapted from Levenson, 1973)

Compared to the three dimensions hypothesized by Roberts and Nesselroade, three of six internal locus of control items loaded primarily on Factor 1, two of five powerful others items loaded primarily on Factor 2, and three of six chance items loaded primarily on Factor 3.[10]

Factor analytic results for STAI-STATE and STAI-TRAIT scales are presented in Tables 6 and 7, respectively. Three factor solutions accounted for a total of 53% of the total item variance in state anxiety items and 51% in trait anxiety items; in addition, fifteen of 20 state anxiety items and 17 of 20 trait anxiety items had simple structure loadings.

Correlational Analyses

Intercorrelations among all eleven dependent measures are presented in Table 8.[11] Five of six predicted associations were found to be significant. These included positive correlations predicted between dualism and external-powerful others locus of control and state anxiety and negative mood, and negative correlations between state anxiety and both pleasant and positive moods. One hypothesis not confirmed was the predicted negative correlation between state anxiety and the arousal-calm dimension of the BMIS; this correlation was not significantly different from $r = .00$. In addition, significant negative correlations found between internal locus of control and both state and trait anxiety, a finding common in the locus of control literature (Phares, 1976).

In light of previously published findings, surprising intercorrelations were found among subscales of the LOC scale. Although Roberts and Nesselroade (1986) report that their modified version of Levenson's Multidimensional Locus of Control Scale "exhibited a sufficient match to the Levenson dimensions" (p. 534), the intercorrelations between internal and external dimensions obtained in Study 1 failed to agree with Levenson's (1973) results. As

---

[10] Following guidelines described by Tabachnick & Fidell (1989), item-to-factor correlations (i.e., factor loadings) greater than $\pm .35$ were considered meaningful.

[11] Due to the non-normal distribution of several variables, Spearman rank correlations were also computed. However, because the average difference between Pearson and Spearman r's was negligible (Mean difference = .003), only the former coefficients are reported.

Table 6. Rotated factor loadings from principle factor analysis of STATE (N = 116).

Item	Factor 1	Factor 2	Factor 3
I feel calm[a]	.48	.20	.45
I feel secure[a]	.59	.41	.32
I am tense	.28	.28	.70
I feel strained	.20	.05	.76
I feel at ease[a]	.57	.08	.43
I feel upset	.09	.16	.54
I am presently worrying about possible misfortunes	.24	.33	.18
I feel satisfied[a]	.65	.21	- .02
I feel frightened	.12	.74	.25
I feel comfortable[a]	.67	.12	.19
I feel self confident[a]	.59	.32	- .11
I feel nervous	.23	.52	.56
I am jittery	.24	.78	.03
I feel indecisive	.10	.67	.09
I am relaxed[a]	.72	.26	.26
I feel content[a]	.78	.18	.14
I am worried	.21	.46	.45
I feel confused	.23	.69	.21
I feel steady[a]	.64	.16	.21
I feel pleasant[a]	.72	.03	.21
Variance Explained	23%	16%	14%

[a] Reversed-scored item

Note: STATE = State-Trait Anxiety Inventory, Form Y-1 (Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983)

Table 7. Rotated factor loadings from principle factor analysis of TRAIT (N = 116).

Item	Factor 1	Factor 2	Factor 3
I feel pleasant[a]	.62	.16	.12
I feel nervous and restless	.14	.69	.15
I feel satisfied with myself[a]	.65	.25	.13
I wish I could be a happy as others seem to be	.42	.65	.17
I feel like a failure	.20	.69	.31
I feel rested[a]	.66	- .03	.23
I feel "calm, cool, and collected"[a]	.52	.33	- .08
I feel that difficulties are piling up so that I cannot overcome them	.19	.57	.41
I worry too much over something that really doesn't matter	- .05	.27	.59
I am happy[a]	.73	.33	.19
I have disturbing thoughts	.19	.18	.67
I lack self confidence	.16	.66	.14
I feel secure[a]	.53	.57	.03
I make decisions easily[a]	.20	.09	.13
I feel inadequate	.24	.51	.27
I am content[a]	.73	.28	.18
Some unimportant thought runs through my mind and bothers me	.23	.06	.69
I take disappointments so keenly that I can't put them out of my mind	.24	.38	.60
I am a steady person[a]	.72	.16	.11

Table 7. (Continued).

Item	Factor 1	Factor 2	Factor 3
I get in a state of tension and turmoil as I think over my recent concerns and interests	.08	.12	.77
Variance explained	20%	17%	14%

[a] Reversed-scored item

Note: STAI-TRAIT = STAI Form Y-2 (Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983)

Table 8. Intercorrelations among dependent variables measured in Study 1 (N = 118).

	2	3	4	5	6	7	8	9	10	11
1. DUALISM	- .10	.02	.37	- .05	.19	.09	.18	.01	.07	.04
(p)	(.30)	(.86)	(<.001)	(.63)	(.05)	(.37)	(.07)	(.90)	(.47)	(.71)
2. INTERNAL		.32	- .10	.21	.02	.19	- .25	.23	- .41	- .50
(p)		(<.001)	(.27)	(.03)	(.80)	(.04)	(.007)	(.01)	(<.001)	(<.001)
3. CHANCE			.14	.23	.17	.23	- .09	.25	- .11	- .11
(p)			(.12)	(.01)	(.06)	(.01)	(.33)	(.005)	(.25)	(.24)
4. OTHERS				.01	.15	.13	.04	.01	.08	.26
(p)				(.92)	(.10)	(.15)	(.65)	(.90)	(.39)	(.004)
5. PLEASANT					.24	.79	- .64	.59	- .74	- .49
(p)					(.01)	(<.001)	(<.001)	(<.001)	(<.001)	(<.001)
6. AROUSAL						.74	.51	.09	.04	.01
(p)						(<.001)	(<.001)	(.35)	(.66)	(.92)
7. POSITIVE							- .17	.44	- .46	- .31
(p)							(.07)	(<.001)	(<.001)	(<.001)
8. NEGATIVE								- .46	.70	.42
(p)								(<.001)	(<.001)	(<.001)
9. MOOD									- .57	- .42
									(<.001)	(<.001)
10. STATE										.61
(p)										(<.001)
11. TRAIT										

expected by the traditional locus of control theory (Phares, 1976), Levenson found no correlation between internal and external locus of control scales and a significant correlation between the two external scales, $r(164) = .54$, $p < .01$ (Levenson, 1973, p. 400). In the present study, a significant positive correlation was found between internal and external-chance scores and no correlation was found between the two external subscales. Both results raise questions concerning the adequacy of this scale to measure locus of control in this sample. One additional unexpected result was the weak positive correlation found between BMIS-PLEASANT and BMIS-AROUSAL; in their examination of the BMIS, Mayer and Gaschke (1988) reported no correlation between these two subscales.

Demographic differences

Results of analysis of variance (ANOVA) tests for gender and residency status differences on each dependent variable are presented in Table 9 and Table 10. As predicted, these analyses revealed no significant differences between males and females or in-state and out-of-state students in this sample. Tables 9-10 also present the results of Bartlett-Box F tests for homogeneity of variance, and suggest that the difference in sample size between groups in either comparison did not result in a violation of the ANOVA homogeneity of variance assumption. Given the violation of the normality assumption in several cases, post-hoc nonparametric analyses or "assumption-freer tests" (Kirk, 1990) for differences as a function of gender and residency status were also performed on the rank scores of following variables: LOC-INTERNAL, BMIS-NEGATIVE, BMIS-MOOD, STAI-STATE. Wilcoxon rank-sum tests, analogous to the Mann-Whitney U test, revealed no differences as a function of gender or residency status, agreeing with the results from ANOVA procedures presented in Table 9 and Table 10.

Discussion

Based on the results of Study 1, the Checklist of Educational Views, State-Trait Anxiety Inventory, Form Y, and Brief Mood Introspection Scale provide sufficient measurement of the three constructs of primary interest. Only the locus of control scale appeared to have poor

Table 9. ANOVA results for gender differences in dependent variables measured in Study 1

		ANOVA					Homogeneity of Variance	
Dependent Variable	n	Mean	SD	F	p	r[a]	Bartlett-Box F	p
CLEV-DUALISM								
Female	60	64.6	12.9	0.34	.56	.06	0.06	.81
Male	38	66.3	13.6					
LOC-INTERNAL								
Female	64	27.8	5.0	1.16	.28	.10	0.22	.64
Male	44	28.8	4.7					
LOC-CHANCE								
Female	66	21.5	4.2	0.11	.74	.03	0.69	.41
Male	44	21.8	4.7					
LOC-OTHERS								
Female	66	17.3	4.3	0.74	.39	.08	0.64	.43
Male	44	16.6	4.4					
BMIS-PLEASANT								
Female	63	9.0	6.6	0.03	.87	.02	< 0.01	.96
Male	41	8.8	7.0					
BMIS-AROUSAL								
Female	65	19.1	3.8	0.51	.48	.07	0.06	.81
Male	41	19.6	3.9					
BMIS-POSITIVE								
Female	65	9.0	4.0	1.02	.32	.10	0.20	.66
Male	42	9.8	4.2					
BMIS-NEGATIVE								
Female	65	7.3	2.6	0.06	.81	.02	0.51	.47
Male	43	7.3	2.3					
BMIS-MOOD								
Female	65	6.3	3.1	0.08	.78	.03	1.80	.18
Male	44	6.4	2.4					
STAI-STATE								
Female	65	40.0	11.7	0.67	.42	.08	1.79	.18
Male	43	38.3	9.6					
STAI-TRAIT								
Female	66	39.4	9.3	0.29	.59	.05	< 0.01	.98
Male	44	40.3	9.4					

[a] Effect size estimate

Table 10. ANOVA results for differences in dependent variables as a function of home residency status.

		ANOVA					Homogeneity of Variance	
Dependent Variable	n	Mean	SD	F	p	r[a]	Bartlett-Box F	p
CLEV-DUALISM								
In-state	52	65.5	11.2	0.02	.90	.01	4.48	.04
Out-of-state	45	64.4	15.2					
LOC-INTERNAL								
In-state	58	27.6	5.0	1.88	.17	.13	0.71	.40
Out-of-state	50	28.9	4.5					
LOC-CHANCE								
In-state	58	21.3	4.9	0.83	.37	.09	5.61	.02
Out-of-state	50	22.1	3.7					
LOC-OTHERS								
In-state	58	17.4	4.4	1.44	.23	.12	0.20	.88
Out-of-state	50	16.4	4.3					
BMIS-PLEASANT								
In-state	56	8.9	7.3	0.01	.94	.01	1.02	.31
Out-of-state	47	9.0	6.1					
BMIS-AROUSAL								
In-state	57	19.1	3.6	1.18	.28	.11	0.21	.64
Out-of-state	47	19.9	3.9					

Table 10. (Continued)

		ANOVA				Homogeneity of Variance	
Dependent Variable	n	Mean	SD	F	p	r^a	Bartlett-Box F	p
BMIS-POSITIVE							
In-state	58	9.4	4.0	0.08	.77	.03	0 24	63
Out-of-state	47	9.6	3.8					
BMIS-NEGATIVE							
In-state	57	6.9	2.3	2.04	.16	.14	0.05	83
Out-of-state	49	7.6	2.5					
BMIS-MOOD							
In-state	58	6.0	3.0	3.27	.07	.17	4.95	.03
Out-of-state	49	6.9	2.1					
STAI-STATE							
In-state	57	38.6	10.4	0.01	.93	.01	0.79	.38
Out-of-state	49	38.8	11.3					
STAI-TRAIT							
In-state	58	41.3	9.9	3.53	.06	.18	4.02	.05
Out-of-state	47	38.0	7.9					

[a] Effect size estimate

psychometric qualities, with low internal reliability estimates for both LOC-CHANCE and LOC-OTHERS subscales, and subscale intercorrelations inconsistent with previously published results. The evidence that the factor structures of the CLEV, BMIS, and STAI-Y are consistent with the results of others suggests that these measures are suitable for a more in-depth, longitudinal investigation of epistemological development, affect, and mood. Conversely, the results of psychometric analysis of the LOC scale suggests that the factor structure of the responses of this sample do not adequately match previous results (Levenson, 1973; Roberts, 1979, cited in Roberts & Nesselroade, 1986). Therefore, only tentative interpretations were made from results based upon composite scores from this instrument.

The lack of any demographic differences reveal that each instrument in this population is largely free of bias in terms of gender and residency status differences. The lack of any gender difference in CLEV scores confirms previous findings reported by others (Baxter Magolda, 1990; Baxter Magolda & Porterfield, 1988), and suggests that incoming male and female students share the same degree of dualism in their beliefs about the nature of knowledge. Because of this finding, no steps will be taken to ensure the same number of male and female participants in further longitudinal studies. In addition, that lack of differences between in-state and out-of-state students will allow future analyses to ignore residency status as a possible covariate of the dependent measures examined in Study 1.

Students participating in Study 1, however, generally reported higher scores for both positive and aroused moods, which is likely due to the general atmosphere of the orientation program. Although students may experience an enhanced mood during their participation in an orientation program, the causal nature of this relation is unclear without a suitable control group. Responses obtained on these BMIS scales in later studies conducted during academic semesters will be analyzed for similar differences.

Summary

These results support the view that the instruments chosen for Study 1 generally provide adequate measures of the psychological constructs of interest in an investigation of

epistemological development during the transition to college. In order to begin to understand how this developmental process occurs, and the role of affective experience in it, a second longitudinal study was designed and conducted during the Fall 1990 academic semester. This study will provide the initial investigation of this process by following students as they begin their college careers.

IV. STUDY 2: LONGITUDINAL ASSESSMENT OF FIRST-SEMESTER COLLEGE STUDENTS (FALL 1990 AND FALL 1991)

Based upon the psychometric results of Study 1, a longitudinal study of epistemological orientation during the transition to college was conducted with two separate cohorts during the Fall 1990 and Fall 1991 academic semesters. A four-wave longitudinal design was chosen in order to enhance the ability of the overall study to describe the developmental processes underlying the earliest stages of Perry's scheme of epistemological development. Although researchers have examined epistemological development during the first year of college (Baxter Magolda, 1990), there has yet to be a detailed examination of the impact of the first semester on the way in which students view the nature and purpose of knowledge.

Of special interest in this longitudinal study was the nature of the relationship among measures of emotion, including affect and mood, and changes in students' epistemological orientation across their first semester in college. It was hypothesized that students who accommodate their epistemological orientations would express 1) decreasing levels of negative affect and mood and 2) increasing levels of positive affect and mood across the four time points, with accommodation of epistemological orientation operationalized as a general decrease in dualism scores over time. In order to determine the most appropriate method for measuring this initial change in epistemological orientation, the results of four different methodological approaches were investigated: the difference score/residual change score approach, the use of cross-time correlations, repeated measures analysis of variance, and individual growth modeling.

In order to examine these issues, the epistemological development of an initial cohort of students was examined during the Fall 1990 academic semester. In order to address the issue of the relation of epistemological development to other psychological domains, students' emotional states and were measured concurrently with their epistemological orientation.

Following the participation of the Fall 1990 cohort, a second cohort was recruited and followed throughout their first-semester at college during the Fall of 1991. This second cohort was used in order to increase the overall sample size, to investigate any cohort-dependent findings, and to obtain measures of two additional psychological constructs not obtained with the first cohort, tolerance of ambiguity and multiplistic epistemological orientation.

Measures of tolerance for ambiguity were obtained from students in the Fall 1991 cohort in order to provide information of the construct validity of the CLEV (Perry, 1968) as a measure of epistemological dualism. Following Frenkel-Brunswik and her colleagues' (Adorno, Frenkel-Brunswik, Levinson, & Stanford, 1950; Frenkel-Brunswick, 1949) original conception of the tolerance for ambiguity as part of the authoritarian personality, Bochner (1965) defines tolerance for ambiguity in the negative, listing several primary characteristics of a person intolerant of ambiguity as including a

> (a) rigid dichotomizing into fixed categories—"a need for categorization"; (b) seeking for certainty and avoiding ambiguity—"need for certainty"; (c) inability to allow for the coexistence of positive and negative features in the same object, e.g., "good" and "bad" traits in the same person; (d) acceptance of attitude statements representing a rigid white-black view of life; (e) a preference for the familiar over the unfamiliar; (f) a positive rejecting of the difference or unusual. (p. 394)

Evaluated on their face validity, certain CLEV items appear to reflect similar rejections of cognitive ambiguity, including "There is nothing more annoying than a question that may have two answers," "It is a waste of time to work on a problem that has no possibility of coming out with a clear-cut and unambiguous answer," and "It is annoying to listen to a lecturer who cannot seem to make up his/her mind as to what he/she really believes" (Kirton, 1981). It was hypothesized that a strong consistent negative correlation would be found between students' dualism and ambiguity tolerance at each all time points, supporting the construct validity of the CLEV.

To enhance the measurement of epistemological orientation, a measure of multiplicity was constructed for use with the Fall 1991 cohort. As previously cited, Ryan (1984a) has stressed the importance of multiplistic thought in epistemological development, arguing that it embodies

the scheme's "most fundamental transition" (p. 1227). Although such a transition is achieved fully by reaching the period described by Perry as commitment to relativism, the seeds of its mature conception of knowledge can be found in the assumptions and attitudes of multiplistic thought. It was hypothesized that not only would an inverse relation between students' dualistic and multiplistic orientations be found at each time, but that decreases in dualism would be accompanied by increases in multiplicity over time.

Method

In order to investigate the relation between affect and change in epistemological orientation during the transition to college, two separate longitudinal studies involving a total of 235 subjects were conducted during the Fall 1990 and Fall 1991 academic semesters. In each study, responses to a series of questionnaires were obtained from participants at four different times, roughly once every three weeks. The dates of the four observation sessions for the Fall 1990 cohort were October 1 and 3, October 29 and 31, November 14 and 19, and December 3 and 4. The dates of the four observation sessions for the Fall 1991 cohort were October 7 and 8, October 28 and 29, November 18 and 19, and December 9 and 10.

Participants

Students in both longitudinal cohorts volunteered as part of an introductory psychology laboratory requirement, and received course credit for their participation. Because only the epistemological development of first-semester students were of interest, students beyond their first semester, including transfer students, were excluded from participation. No restrictions were placed upon the number of males or females accepted into either cohort because of the lack of gender differences in the results of Study 1; subsequently females comprised 80% of the total sample. Of the 250 total participants recruited, 235 completed at least the first observation, and 209 completed at least three of the four observations. Differences in sample sizes reported below reflect either instances of missing data or the inclusion of certain measurements with only one of the cohorts. Upon completing the final

observation session, participants were fully debriefed according to guidelines established by the University of New Hampshire's Institutional Review Board (IRB).

Instrumentation

The instrumentation used with both longitudinal cohorts included a demographic questionnaire and a survey packet made up of the following questionnaires evaluated in Study 1: a shortened version of Perry's (1968) Checklist of Educational Views, the State-Trait Anxiety Inventory-Form Y (Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983), the Brief Mood Introspection Scale (Mayer & Gaschke, 1988), and Roberts and Nesselroade's (1986) version of Levenson's Multidimensional Locus of Control Scale. In addition, participants in the Fall 1991 cohort received an expanded version of the original demographic questionnaire as well as two additional psychometric questionnaires, a measure of ambiguity tolerance developed by Kirton (1981), and a scale designed specifically for this study to measure students' multiplistic epistemological orientation.

Based upon a review of scales for measuring tolerance of ambiguity (Buchanan, 1986), a shortened version of MacDonald's (1970) AT-20 developed by Kirton (1981) was chosen as a measure of ambiguity tolerance. This 11-item version of the original AT-20 resulted from Kirton's (1981) finding that removing nine of the original 20 items increased its internal reliability from .61 to .72. To develop a measure of multiplicity, twenty statements regarding various aspects of a postdualistic orientation were constructed from a collection of statements from students identified as multiplistic by Baxter Magolda and Porterfield (1988) using a written measure of epistemological development. The constructed scale, the Multiplistic Epistemological Orientation Scale (MEOS), was designed to measure three features of multiplistic thought, including the uncertainty of knowledge in some domains, the central role of the learner in creating knowledge, and the influence of personal opinion in the determination of knowledge. Psychometric analyses, similar to those performed in Study 1, were conducted to assess the measurement properties of both the AT-20 and MEOS.

Procedure

An efficient procedure for longitudinal data collection was made possible through the use of a computer-administered survey program that allowed participants to enter responses to questionnaire items directly to a computer data file. After being briefly instructed on the use of the computer terminal and survey program, up to 10 students at a time were able to complete the full survey usually in less than 20 minutes each. During the initial session, students were given a study-specific identification number used throughout their participation that allowed them to access the computer-aided survey program as they returned for additional sessions. In addition to the computer-aided surveys, all participants completed a written demographic questionnaire at Time 1, which was later coded and added to the existing survey data file.

In the Spring of 1992, a follow-up assessment was made with participants in the Fall 1990 cohort. Of the original 125, 89 were contacted twice through the mail and asked to volunteer for a follow-up session related to their original participation. Twenty-nine sophomore students responded and were administered a written version of the computer-aided survey used during the four sessions of the Fall 1990 study. Each volunteer was paid $5.00 for his or her participation in the follow-up session.

Results

Addressing the methodological issues raised in Section II, the results of four different approaches to analyzing change in epistemological orientation are presented below: the difference score/residual change score, cross-time correlations, repeated measures ANOVA, and individual growth modeling. Results are also presented from preliminary analyses for both cohorts and psychometric analyses of the two additional instruments used with the Fall 1991 cohort.

Preliminary Analyses

After data collection was complete, descriptive statistics for all variables were computed and examined for errors. Similarly, the distribution of each composite measure (i.e., multi-

item scores) was examined for any possible errors or unusual values. None was found in either instance. Due to the automation of the data collection process, errors made by participants were highly unlikely; however, steps were take to audit the computer data files for the signs of any errors. Again, none was found.

To determine if the Fall 1990 and Fall 1991 cohorts could be treated as equivalent, group means as a function of cohort for each of eleven dependent variables were compared at each of four time points. A series of ANOVA tests found no significant differences between the cohorts on any measures at any time point. Because this required 44 ANOVA tests of significance to be run, a family-wise alpha level was set at .05; as such, the probability of each test was set to $\alpha = .001$ (i.e., .05 ÷ 44). The only result that approached significance was a slight difference in powerful others—external locus of control. Although on average the Fall 1990 cohort scored slightly higher (Mean = 17.15, SD = 4.39) than the Fall 1991 cohort (Mean = 15.56, SD = 3.88), this difference was not significant at the predetermined alpha level, $F(1, 203) = 7.54$, $p = .007$. In addition, the effect size of this result was found to be weak, $r = .17$. Because of the lack of group differences among the dependent variables, the two cohorts were combined in order to increase statistical power and reliability. This resulted in a total sample size of 235 first-semester students participating at up to four time points.

Psychometric Analyses

Using data obtained at Time 1, both the measure of ambiguity tolerance and the measure of multiplistic epistemological orientation were found to have inadequate internal reliabilities for this sample, AT-20 Chronbach's $\alpha = .35$, MEOS Chronbach's $\alpha = .31$. In order to investigate their underlying factor structure, principle components factor analysis of both instruments were conducted using responses obtained at Time 1. Rotated factor loadings for the shortened version of the AT-20 and the 20-item MEOS are presented in Table 11 and Table 12,

Table 11. Factor loadings from principle factor analysis of AT-20 (N = 105).

Item	Factor 1	2	3	4
There's a right way and a wrong way to do just about everything.	.40	.22	- .18	.48
Practically every problem has a solution.	.47	.44	- .26	- .11
I have always felt that there is a clear difference between right and wrong.	.44	.42	.13	- .23
Nothing gets accomplished in this world unless you sick to some basic rules.	.40	.16	.48	.07
If I were a doctor, I would prefer the uncertainties of a psychiatrist to the clear and definite work of someone like a surgeon or X-ray specialist.[a]	- .33	.07	.27	.02
Vague and impressionistic paintings really have little appeal to me.	.42	- .28	.06	.25
Before an examination, I feel much less anxious if I know how many questions there will be.	- .00	.18	- .04	- .11
The best part of working on a jigsaw puzzle is putting in the last piece.	.34	- .28	- .28	- .22
I don't like to work on a problem unless there is a possibility of coming up with a clear-cut and unambiguous answer.	.45	- .03	.00	- .01
I like to fool around with new ideas, even if they turn out later to be a total waste of time.[a]	- .43	.37	.02	.04
Perfect balance is the essence of all good composition.	.48	- .28	.23	- .14
Variance explained	48.2%	23.7%	15.5%	12.5%

Note: AT-20 = McDonald's Ambiguity Tolerance 20, Shortened Version (Kirton, 1981)
[a] Reversed-scored item

Table 12. Rotated factor loadings from principle factor analysis of MEQS (N = 121)

Item	Factor 1	2	3	4	5	6	7
College grades should be based upon the time and effort a student puts into a course, not on actual performance on tests or assignments.	.61	.25	.10	.17	.28	.04	- .01
When I make a decision, I often worry whether or not I made the right choice.[a]	- .69	- .08	.01	.07	.01	.03	- .30
What is important about someone's personal opinion is not how many facts they have to back it up, but rather how strongly they believe it is true.	.52	.25	.28	- .02	.24	- .02	.27
I like being in a class where I can express my opinion, because so much of what you learn is just the professor's opinion anyway.	.11	.72	- .19	.02	.06	.18	.25
Often I feel like it's difficult to know if my ideas are right or wrong.	.74	.01	- .36	-.18	- .01	.03	- .17
There are so many things about the world that we don't know the answer to yet, and we probably will never know the answers to many of them.	.15	.79	.05	- .09	- .10	- .02	.08
In college, I've learned that the important thing isn't whether or not you get an answer right, but rather how well you can support your answers with evidence and reason.	.05	.41	.02	.05	.58	.10	- .12
Where authorities and experts do not know the answer, any opinion is as good as another.	.20	.26	.24	.01	.11	.01	.57
I think the most important goal of college should be to teach students to look at things from different perspectives.	.03	.49	- .03	.37	.29	- .05	- .21

Table 12 (Continued)

Item	Factor 1	2	3	4	5	6	7
I usually think more about short-term consequences than future ones when I try to decide what do to in a situation.[a]	- .21	.06	.04	.63	- .03	- 03	-.09
Too much time is often wasted on class discussions because some students just like to hear themselves talk. [a]	- .38	.07	.70	.21	- .09	< .01	01
When two people give different explanations for the same thing, their explanations are affected by their own personal beliefs, values, and biases.	.08	- .02	- .02	.76	.11	.01	.26
The best way to make the right decision in most situations is to get expert advice and follow it.[a]	.04	- .13	.16	.60	- .43	- .14	- .22
Group projects in college courses are a bad idea because there is always a "slacker" in every group who never does his/her part of the work.[a]	- .09	- .01	.10	- .03	- .28	.81	- .13
I would prefer to take a course in which students are required to work together to learn class material than having to do everything individually.	.18	.23	- .12	- .02	.28	.73	-.05
I would be against requiring all students to take courses that stress non-traditional points of view, like Women's Studies or African-American Literature.[a]	.14	- .16	.81	- .07	.04	- .01	.07
Some college courses are only able to present opinions and theories, but others are able to present facts and real answers.	.03	.02	- .31	.05	- .01	- .18	.68

Table 12 (Continued).

Item	Factor 1	2	3	4	5	6	7
The purpose of an college instructor is to provide their expert knowledge on a specific topic.[a]	- .21	.22	- .03	.04	- .80	.12	- .17
If I had the choice, I would rather take an exam that was multiple-choice than essay.[a]	- .02	.10	.02	.11	- .13	- .14	- .17
It seems to me that it is impossible to accurately judge what a student has learned in a college class.	.19	.34	- .10	.11	- .01	- .49	- .38
Variance explained	16.3%	10.2%	8.7%	7.7%	6.4%	6.2%	5.5%

[a] Reversed-scored item.
Note: MEQS = Multiplistic Epistemological Orientation Scale

respectively.[12] For the AT-20 these factor loadings reveal that six of the eleven items have a complex factor structure by loading strongly (i.e., > ± .35) on more than one factor. Similarly, six of 20 items of the MEQS were found to have a complex factor structure. Because of both the complex factor structure of many of their items and their poor internal reliability, only tentative interpretations were made from results based upon composite scores from each of these instruments.

Results of Different Approaches to Analyzing Change in Epistemological Orientation During the First Semester in College

Difference Scores

In order to evaluate the adequacy of the difference score approach to measure change in epistemological orientation, a raw difference score for dualism was obtained by subtracting students' composite CLEV scores at Time 4 from those at Time 1. The 200 CLEV difference scores measuring linear change ranged from - 41 to + 38, had a mean of - 1.76 and standard deviation of 8.70, and were found to be normally distributed, Shapiro-Wilk's $W = 0.96$, $p = .70$.

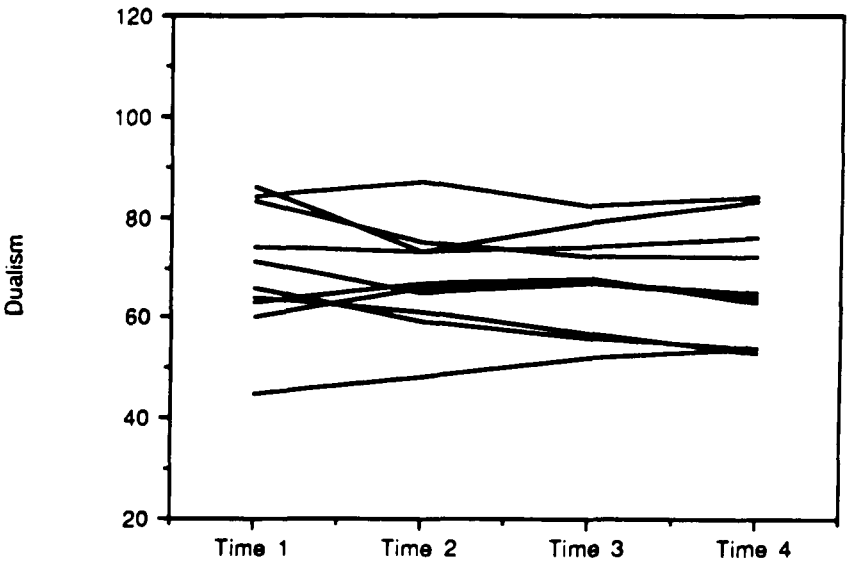
In terms of what these difference scores mean, it is worth noting that the average amount of linear change between CLEV scores at Time 1 and Time 4 was extremely small. In fact, to achieve even a liberal level of significance, $t(200) = 1.20$, $p = .10$, an individual's amount of change had to be at least ± 12.2 to be significant. Using this criterion to group students according to their degree of linear change, eight (3.4%) had increasing CLEV scores between Time 1 and Time 4, 50 students (21.4%) had declining CLEV scores between Time 1 and Time 4, and 176 (75.2%) had no change between Time 1 and Time 4. From this nominal grouping, it is apparent that most students report no linear change from Time 1 to Time 4.

The degree to which change may be nonlinear, however, cannot be addressed by the difference score. Figure 2 presents a random sample of CLEV scores of 10 students at four

---

[12] Only those extracted factors having eigen values greater than 1.0 are presented in Table 11 and Table 12.

Figure 2. Change in dualism scores of 10 randomly chosen individuals across four time points.



time points, and suggests that some individuals' change may not be measured adequately with a linear change model like the difference score. These patterns of change, given the nonlinear conception of development described by Perry (1977, 1981), indicate that the difference score, a measure of linearity, is not the best method for measuring change in epistemological orientation.

Residual Change Scores

Despite the problem of using a linear model to measure change in first-semester students' epistemological orientation, residual change scores in dualism scores between Time 1 and Time 4 were computed in order to address the adequacy of this typical attempt to improve the difference score as a measure of change. Table 13 presents both zero-order correlations between raw change and 17 variables obtained at Time 1 and partial correlations controlling for initial status between the same variables. These partial correlations provide a measure of residual change in CLEV dualism scores controlling for interindividual differences in initial status. The only significant relationship found was a negative correlation was between raw change and initial status,[13] an unsurprising finding given the discussion of the difference score presented in Section II. In addition, of the 16 partial correlations computed, significant negative correlations were found between residual change and four measures of mood: pleasant, aroused, positive, and overall mood.

Although some of these partial correlations are significant, their meaning is unclear because of the methodological implications of equalizing individual differences in initial status. Ignoring interindividual differences in initial status has been shown by Rogosa, Brandt, and Zimowski (1982) to have serious flaws in terms of both its computation—most formulae for computing residual change scores result in biased and imprecise measures of true change—and

---

[13] Although the negative correlation between raw change and the number of campus activities approached significance, after inspecting the scatterplot of these variables, a decision was made to delete one case because of its extreme value for number of campus activities (ACTIVE = 7, z = 4.04, $p < .01$). With this atypical case removed, the correlation between these variables was nonsignificant, $r = -.16$, $p = .10$.

Table 13. Correlations between variables at Time 1 and change and residual change in dualism scores from Time 1 to Time 4.

Time 1 Variable	Correlation with Change in Dualism N	r	p	Correlation with Residual Change in Dualism[a] N	r	p
Dualism	201	- 0.25	< 0.001	79	-	-
Internal LOC	200	0.01	0.93	79	- 0.14	0.23
Chance LOC	200	0.09	0.21	79	0.01	0.90
Powerful Others	200	0.03	0.72	79	0.01	0.94
State Anxiety	200	0.03	0.64	79	0.02	0.88
Trait Anxiety	201	- 0.01	0.94	79	0.18	0.11
Pleasant Mood	201	- 0.07	0.29	79	- 0.29	0.01
Aroused Mood	201	- 0.10	0.15	79	- 0.23	0.04
Positive Mood	201	- 0.09	0.21	79	- 0.30	0.01
Negative Mood	201	- 0.05	0.52	79	0.12	0.30
Overall Mood	201	- 0.08	0.26	79	- 0.26	0.02
Ambiguity Tolerance	100	- 0.06	0.57	79	- 0.30	0.01
SAT-Verbal	94	0.13	0.22	79	- 0.03	0.78
SAT-Quantitative	95	0.11	0.28	79	0.05	0.69
High School GPA	86	- 0.09	0.41	79	0.01	0.96
Campus Activity	102	- 0.18	0.07	79	- 0.20	0.07
Course Difficutly	102	- 0.06	0.58	79	- 0.20	0.08

a Partial correlation controlling for dualism score at Time 1

its logic—it replaces a rather straight-forward question about true intraindividual change with the complex question of how much individuals would have changed if they all began at the same point.

Cross-time Correlation Matrix

In order to examine the ability of a cross-time correlation matrices to measure change, intercorrelations between dualism and ambiguity tolerance scores were computed for each time point. These intercorrelations, presented in Table 14, reveal that tolerance for ambiguity and dualism are consistently negatively correlated at four different points in students' first semester in college. However, based solely on these results, it is difficult to generate an interpretation of how change in either variable is related to the change in the other. Because this analysis contains no information about estimates of underlying true change, these correlations present discrete and unrelated information (at least as a function of time) about these two constructs.

Repeated Measures ANOVA

The ability of a repeated measures analysis of variance to provide information about change over time was examined by investigating the relationship between mean dualism scores over each of the four points and initial levels of ambiguity tolerance. Initial level of ambiguity tolerance was chosen as a grouping variable given its its correlation with dualism at each time point. Each of 105 subjects was assigned to one of three groups by partitioning the distribution of initial ambiguity tolerance scores approximately into thirds. With this assignment strategy, 37 students were classified as high in ambiguity tolerance, 27 students were classified as low in ambiguity tolerance, with the remaining 26 students classified into a moderately ambiguity tolerant group. A 3 (group) X 4 (time) repeated measures ANOVA was used to test for any mean differences in dualism as a function of either ambiguity tolerance, time, or their interaction.

Given this classification, it was hypothesized that there would there be a significant main effect for both ambiguity tolerance group and time. In addition, a significant group X time

Table 14. Intercorrelations among dualism scores and ambiguity tolerance scores at four points (N= 105).

	Dualism at Time 1	Dualism at Time 2	Dualism at Time 3	Dualism at Time 4
Ambiguity Tolerance at Time 1	-0.57	-0.57	-0.66	-0.59
Ambiguity Tolerance at Time 2	-0.54	-0.59	-0.66	-0.66
Ambiguity Tolerance at Time 3	-0.51	-0.58	-0.67	-0.65
Ambiguity Tolerance at Time 4	-0.53	-0.41	-0.58	-0.53

Note: All correlations significant at $p < .001$ (two-tailed)

interaction was expected, because it was felt that students initially high in ambiguity tolerance would experience a greater decrease in dualism scores over time. This specific interaction hypothesis was based on the assumption that ambiguity tolerant students would be more open to modifying their epistemological beliefs that less ambiguity tolerant students.

The results of the repeated measures ANOVA are presented in Table 15, including summary table information; cell sizes, means, and standard deviations; and marginal means. As expected there was a highly significant main effect for initial ambiguity tolerance. Collapsing across time, almost half of the total variance in the dualism scores is accounted for by initial levels of ambiguity tolerance. Collapsing across groups, there was a significant main effect for differences in dualism scores over time, however this only accounted for 3% of the variance in dualism scores. Last, no significant interaction of differences in the amount of change as a function of ambiguity tolerance group, disconfirming the hypothesis that students who started the semester more tolerant of ambiguity would experience more change in their dualism scores than other students. A visual summary of these results is shown in Figure 3.

Although the repeated measure ANOVA approach is able to provide some information about differences in epistemological orientation during the transition to college—for example that CLEV scores as a function of grouped levels of initial ambiguity tolerance are different from one another at some of the four points—this information about differences cannot be interpreted as describing change in a continuous sense. Like the cross-time correlation matrix, no attempt is made to estimate differences as a function of continuous time, but rather relies on information about change in the form of discrete time points. Another limitation of this type of analysis is that much of the information regarding interindividual difference in change are lost when subjects are grouped into such ordinal categories as "high," "moderate," and "low."

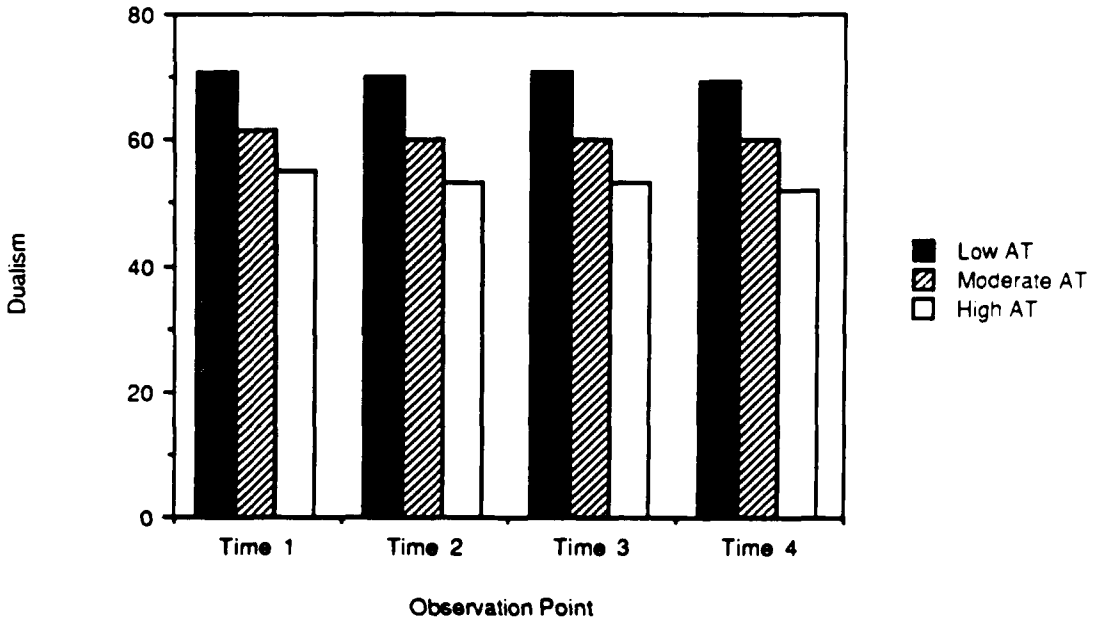
Relative to previously described approaches to analyzing change, repeated measures ANOVA does a fairly good job of providing information about change, even if that change is at the level of some broad grouping variable. This may be adequate for some if not many types of

Table 15. Results of repeated measures analysis of variance testing mean differences in dualism as a function of initial ambiguity tolerance group, time, and group by time interaction (N=90).

Source	n	Mean	SD	F	p	r^2
Ambiguity Tolerance Group	2	17705.53	8852.76	19.54	< 0.001	.45
Error	87	39409.47	452.98			
Time	3	191.22	63.74	2.82	0.04	.03
Time*Ambiguity Tolerance Group	6	40.20	6.70	0.30	0.94	.01
Error	261	5900.76	22.61			

Initial Ambiguity Tolerance	n	Time 1 Mean	Time 1 SD	Time 2 Mean	Time 2 SD	Time 3 Mean	Time 3 SD	Time 4 Mean	Time 4 SD	Marginal Mean
High	37	55.00	10.53	53.15	11.11	53.11	9.79	52.11	10.73	53.34
Mid	26	61.73	11.72	60.15	14.00	60.23	9.07	59.92	13.96	60.51
Low	27	70.74	11.49	70.22	11.00	70.75	9.07	69.22	10.30	70.23
Marginal Mean		62.49		61.17		61.36		60.42		

Figure 3. Chart of mean dualism scores by initial ambiguity tolerance groupings over four time points (N = 90).



questions developmentalists ask. However, one critical limitation of this approach is its inability to say anything about individual growth.

Growth curve modeling

In order to analyze change in individual students' epistemological orientation, a series of an individual growth modeling analyses was performed (Rogosa, Brandt, & Zimowski, 1982; Willett, 1988). As discussed in Section II, IGM involves performing individual regression analyses and obtaining estimates of true latent growth in the form of slope coefficients. These estimates of growth can then be used as dependent or criterion variables in order to address questions about the existence of any systematic interindividual differences in true growth.

Describing Individual Growth Using Three Models. Initially three different true growth models were fit to each individuals' four dualism scores: a linear model, a quadratic model, and a cubic model. Formula 1 (see Section II) provides the general "true score" mathematical model with which the linear model regression equation was constructed. Using similar notation, the quadratic model regression equation was:

$$X_{ip} = \xi_p (t_i) + \xi_p (t_i)^2 + \varepsilon_{ip} \quad (2)$$

With the quadratic model, a second term, $\xi_p (t_i)^2$, has been added to the linear model. This term, in essence, is an estimate of the degree to which the true growth model has a single curvilinear "bend" to it. Likewise the cubic model adds a third term:

$$X_{ip} = \xi_p (t_i) + \xi_p (t_i)^2 + \xi_p (t_i)^3 + \varepsilon_{ip} \quad (3)$$

The third term estimates the degree to which true growth has a second "bend" or change in direction.

Note that both Formula 2 and Formula 3 describe polynomial models of growth; that is each has more than one slope estimate to describe growth, and each includes the terms from previous "lower-order" models. Therefore, a linear IGM will provide a measure of the linearity of an individual's true growth; a quadratic IGM will provide a measure of the linearity and the "quadratic-ness" of an individual's true growth; and a linear IGM will provide a measure of the linearity, the "quadratic-ness" and the "cubic-ness" of an individual's true

growth. In addition, given that four measures of dualism data were obtained from each student, a cubic growth model will be able to account for all of the variance in the data.

Figure 4 presents a comparison of the increasing complexity of these three growth models in their estimation of true growth in a collection of 10 students' dualism scores obtained over the Fall 1990 academic semester. In the first of this series of plots, one student's true linear growth in dualism can be seen to begin at approximately 85 and grows to 100 by Time 4, suggesting strongly positive linear growth. However, comparing the following plots, it is clear that a simple linear model does not tell the whole developmental story. Using a quadratic model, this same student's true growth could be described a positively linear and negatively quadratic. Last, by moving to a cubic model, more detail can be seen in this student's true growth, which could be described as positively linear, negatively quadratic, and negatively cubic (i.e., its overall up-and-down oscillation has a downward trend).

Stem-and-leaf histograms displaying the results of each growth model are presented in Figure 5 (linear model), Figure 6 (quadratic model), and Figure 7 (cubic model).[14] These histograms display the frequency of the estimated true rates of growth (i.e., slopes) across each subject. In addition, descriptive statistics for the estimates of true growth rate are presented, along with mean goodness-of-fit of each particular model across all students.

These estimates provide meaningful information about individual growth in dualism scores across student's first semester, and although the general magnitude of change is not too remarkable, a great deal of variability in rates of growth—across all models—is notable. Information about the relative ability of each model to account for variance in students' scores can be seen in the values of the r^2 statistics. As would be expected, the increasing complexity of each growth model allows it to account for more variance than previous models: the linear model accounts on average for less than half of the variance, the quadratic model accounts for

[14] A version of the SAS® computer program for conducting a linear individual growth model analysis developed by Willett (1987/1991) appears in Appendix A.

Figure 4. Comparison linear, quadratic, and cubic growth models of dualism over time (n = 10).

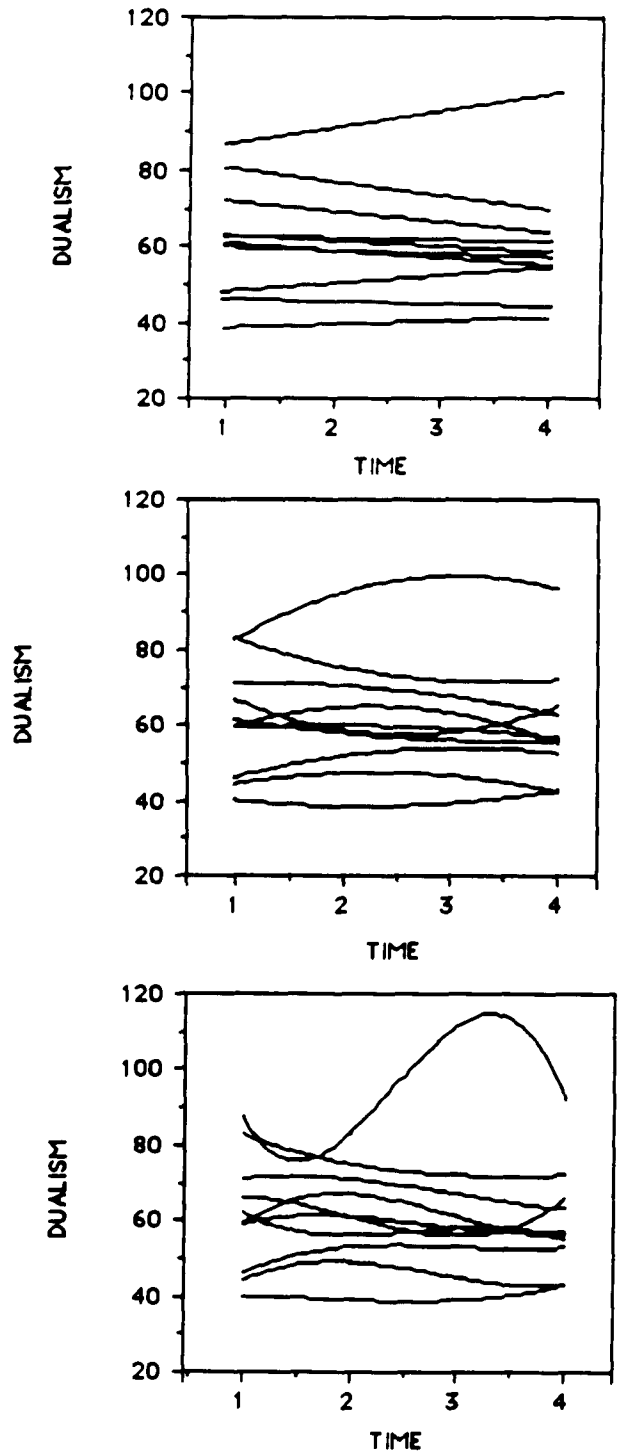


Figure 7. Histogram and summary statistics of slope estimates of true linear growth in students' epistemological dualism during the first semester in college (N = 209).

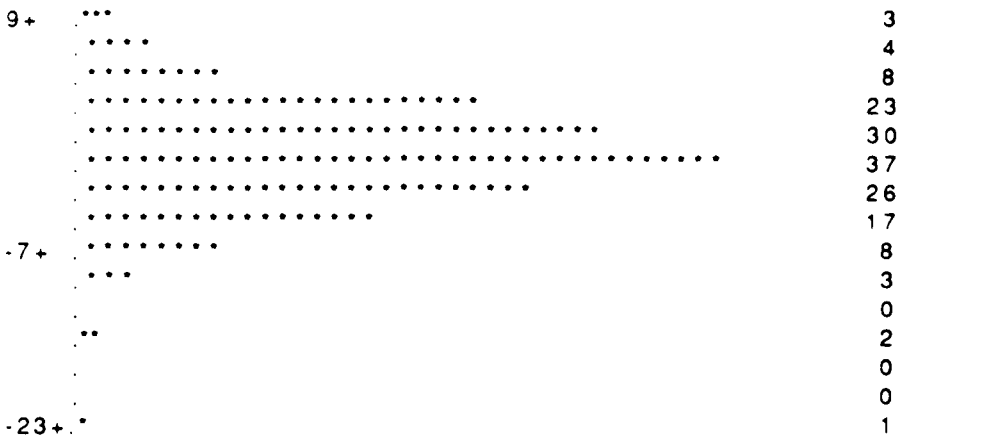
Stem	Leaf	Frequency
11	*	1
10		0
9		0
8		0
7	*	1
6		0
5	***	3
4	***	3
3	********	8
2	**********	19
1	**********	30
0	**********	24
-0	**********	27
-1	**********	30
-2	**********	18
-3	**********	27
-4	*******	7
-5	****	4
-6	***	3
-7	**	1
-8	*	1
-9	*	1
-10		0
-11		0
-12	*	1

Mean	- 0.65
Standard Deviation	2.85
Median	- 0.50
Mode	- 1.90
Range	23.40
Shapiro-Wilk's W Statistic	0.99 (p = .96)

Mean r^2 for linear growth model = .46

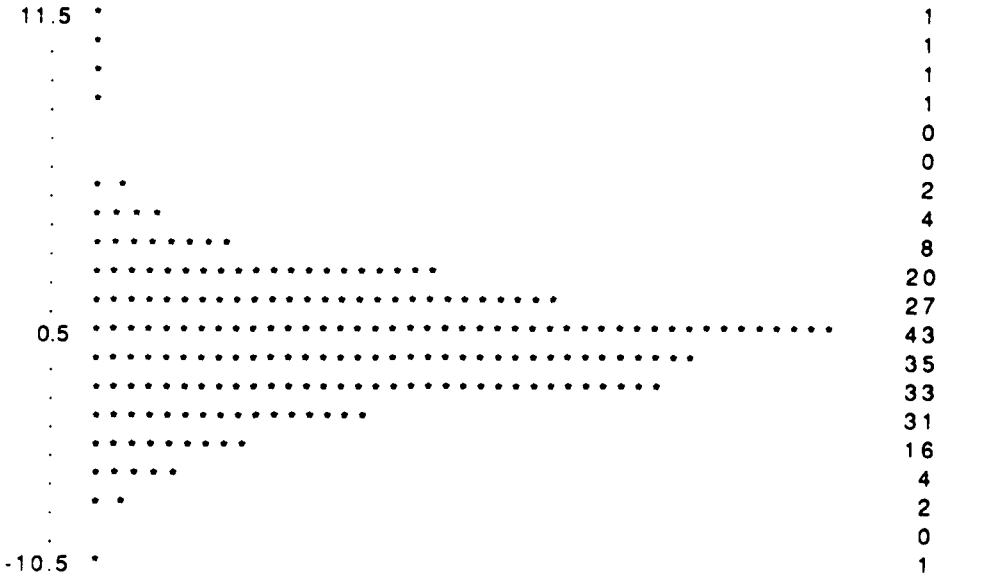
Figure 6. Histogram and summary statistics of slope estimates of true quadratic growth in students' epistemological dualism during the first semester in college (N = 209).

Linear Slope Estimates



Mean	- 0.80	Standard Deviation	4.12
Median	- 0.50	Mode	- 1.50
Range	32.90	Shapiro-Wilk's W Statistic	0.93 (p < .01)

Quadratic Slope Estimates



Mean	- 0.06	Standard Deviation	2.61
Median	0.00	Mode	0.50
Range	21.75	Shapiro-Wilk's W Statistic	0.94 (p < .01)

Mean r^2 for quadratic growth model = .52

Figure 7. Histogram and summary statistics of slope estimates of true cubic growth in students' epistemological dualism during the first semester in college (N = 209).

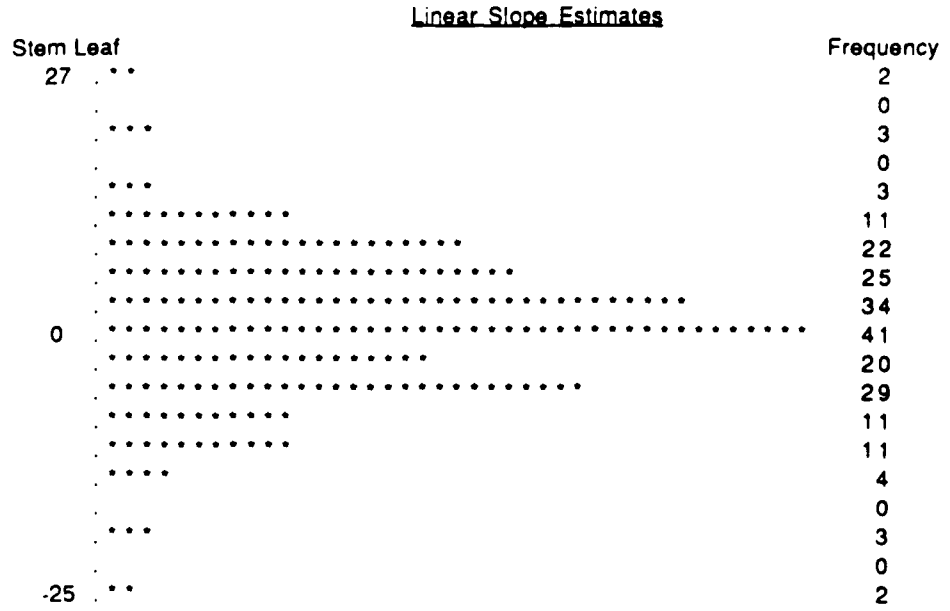


Figure 7. (Continued)

Quadratic Slope Estimates

Stem	Leaf	Frequency
21	. * * *	3
	.	0
	. * * *	3
	.	0
	. * * *	3
	. * * * * * * * * *	9
	. * * * * * * * * *	9
	. * * * * * * *	7
	. *	24
	. *	37
0	. *	44
	. *	41
	. *	33
	. * * * * * * * * * *	9
	. * * * * * * * * * * * * *	12
	. * * * * * * * * * * * * *	12
	. * * * * * *	6
	.	0
	. * *	2
	.	0
	. * *	2
	.	0
-25	. * *	2

Mean	- 0.06	Standard Deviation	- 6.11
Median	0.00	Mode	- 0.50
Range	46.5	Shapiro-Wilk's W Statistic	0.96 ($p < .01$)

Figure 7. (Continued)

Cubic Slope Estimates

Stem	Leaf	Frequency
17	*	1
	.	0
	.	0
11	.	0
	. **	2
	. **	2
5	. **********	18
	. **********	18
	. **********	40
-1	. **********	47
	. **********	18
	. **********	12
-7	. ***	3
	. *	1
	.	0
-13	. *	1

Mean	0.10	Standard Deviation	3.50
Median	- 0.17	Mode	- 1.67
Range	30.00	Shapiro-Wilk's W Statistic	0.96 ($p < .01$)

Mean r^2 for cubic growth model = 1.00

a little over half the variance, and the cubic model accounts for all of the total variance in the change in students' dualism scores.

Because of its enhanced precision estimating true growth from observed growth, the cubic model was chosen as the most appropriate model to describe epistemological growth during the initial transition to college. The cubic model was also seen as the most appropriate model for measuring the type of growth that Perry (1977, 1981) indicated typifies intellectual and ethical development. A cubic model allows for the measurement of both linear growth and recursive growth; both the general trend away from dualism and the fluctuations between dualistic and multiplistic beliefs can be accommodated with this model.

Investigating Systematic Interindividual Differences in True Growth. As described in Section II, the second step involved in IGM is designed to answer questions about why some individuals may be growing at different rates then others. When growth is measured using a linear model this is done by correlating linear slope estimates of true growth with some other variable, usually measured at Time 1 or before. However, when a cubic model is chosen, the investigator is faced with the problem of accounting for differences in three dimensions of growth. The most straight-forward procedure for to accounting for differences in cubic growth is to extend the approached used with a linear model by correlating individual predictor variables with each of the three slope coefficents obtained in the cubic growth modelling analysis (i.e., linear, quadratic, and cubic).

Table 16 presents correlations among 17 variables measured at Time 1 and each of the regression coefficients obtained by modelling the cubic growth in dualism scores over time. From this collection of intercorrelations, it appears that only in a few instances are initial affective, personality, or demographic variables associated with differences in each of the three compnents of true cubic growth. Aside from this lack of association, this correlational approach also presents some difficulty in interpretation due to the fact that each regression term is analyzed independently. In order to analyze the relation between predictor variables

Table 16. Correlations among variables obtained at Time 1 and regression coefficients in cubic growth model of dualism during the first semester in college.

		Linear Slope		Quadratic Slope		Cubic Slope	
Variable at Time 1	N	r	p	r	p	r	p
Dualism	180	.01	.85	.17	.02	-.16	.03
Internal LOC	179	-.02	.77	.04	.59	-.02	.75
Chance LOC	179	.07	.37	-.09	.22	.04	.63
Powerful Others LOC	179	.04	.62	-.02	.80	.01	.92
State Anxiety	179	.03	.73	-.07	.37	.04	.57
Trait Anxiety	180	-.05	.45	-.13	.09	.11	.14
Pleasant Mood	180	-.11	.14	.03	.68	.01	.86
Aroused Mood	180	-.10	.18	-.04	.59	.05	.51
Positive Mood	180	-.17	.02	-.03	.65	.08	.29
Negative Mood	180	.08	.27	-.03	.66	-.03	.73
Overall Mood	180	-.02	.83	.06	.42	-.04	.57
Ambiguity Tolerance	90	-.11	.31	.03	.76	.01	.96
SAT-Verbal	82	.08	.45	.03	.79	-.05	.67
SAT-Quantitative	83	.06	.61	.22	.05	-.16	.14
High School GPA	77	.03	.76	.11	.36	-.09	.43
Campus Activity	90	-.11	.30	.08	.44	-.06	.53
Course Difficutly	90	.09	.40	.11	.32	-.12	.24

and more than one dimension of true growth, an alternative strategy developed by Belsky and Rovine (1990) was also used with this data.

In a study of marital satisfaction, Belsky and Rovine (1990) suggested that when faced with variability in the type, as well as the amount, of true growth, a sensible strategy is to classify subjects into various patterns of change that correspond to their slope estimates. Once subjects are classified, various inferential techniques can be used to investigate the existence of systematic differences among true growth patterns. One drawback to this procedure is that in classifying subjects a certain amount of information about individual differences in growth is lost, as it is in repeated measures ANOVA. Thus the researcher interested in understanding differences among patterns of growth is faced with a choice between using a simpler model to describe growth (i.e., linear), or to loose some information regarding interindividual change. For this study, it was decided that any loss of information would be outweighed by the more accurate measure of growth presented by the cubic model.

Subjects' growth rates were classified by transforming their raw dualism scores into ordinal component scores, through a process described by Belsky and Rovine (1990) similar to orthoginal polynomial contrast codings.[15] Once each of the three slope estimates of each individual's growth in dualism was obtained, it was transformed into a standardized z-score.

For ease of interpretation, and because of the conceptualizations of change found in Perry's model, only the standardized linear and cubic slope estimates were chosen as adequate descriptions for classifying students' epistemological growth. Figure 8 presents a matrix crossing linear and cubic slope estimate z-scores for 180 students, with each student's change in dualism scores described in terms of its linear and cubic change. For example, a student exhibiting a strong overall decrease and an overall fluctuating change is located in this matrix by intersecting the linear (z = -3) and cubic (z = -2) dimensions.

---

[15] The formulae used in determining patterns of linear, quadratic, and cubic growth are presented in Appendix B.

Figure 8. Matrix crossing linear and cubic types for change in students' dualism scores during the first semester in college (N=180).

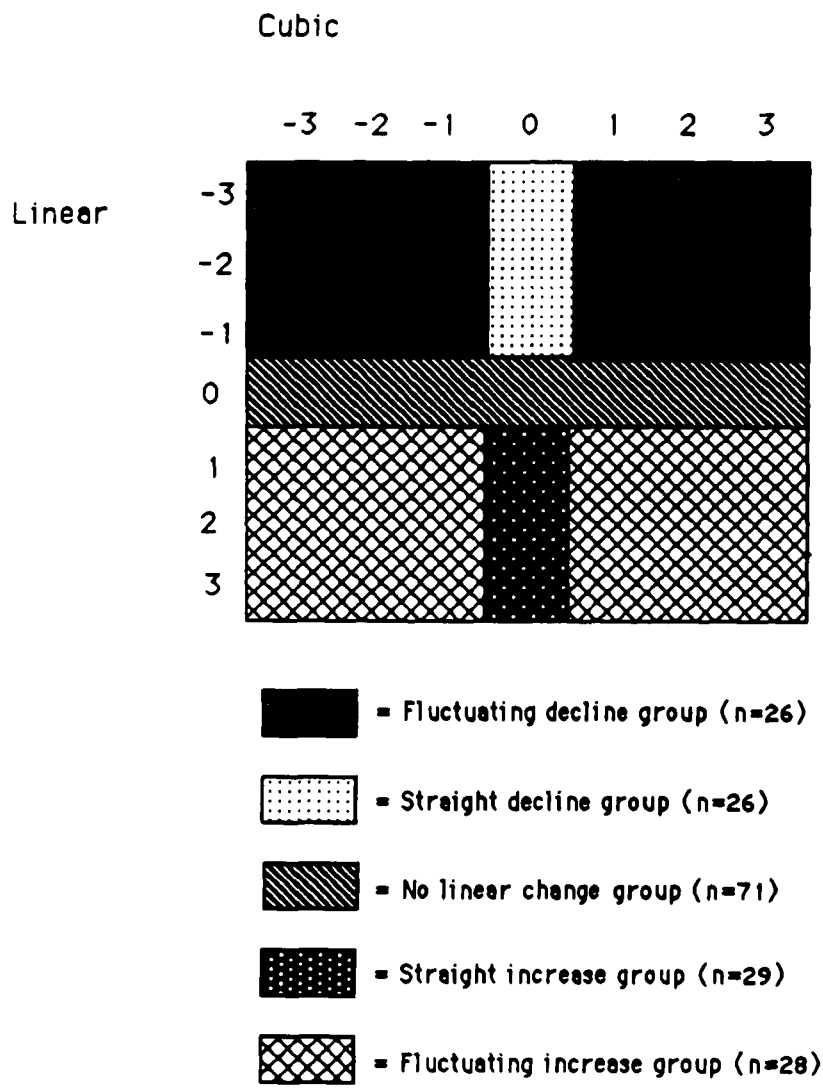
Linear \ Cubic	-3	-2	-1	0	1	2	3	
-3	0	1	0	0	0	0	0	1
-2	0	0	1	3	2	0	0	6
-1	0	3	8	23	6	3	2	45
0	0	2	22	30	14	3	0	71
1	1	1	9	28	10	1	0	50
2	1	0	1	1	0	2	0	5
3	0	0	1	0	1	0	0	2
Total	2	7	42	85	32	10	2	180

As expected given previous results, the large number of students had z-scores between -1 and +1 for the dimensions of linear and cubic change (n = 30). These values represent those students who reported relatively no change in their dualism scores across four time points. Moving away from the center of the matrix, some variation in both dimensions is found, with cell frequencies becoming smaller and smaller. Given this distribution of linear and cubic slope estimates, five patterns of change were designated: no linear change, straight decline, straight increase, fluctuating decline, and fluctuating increase. Figure 9 presents a summary of these patterns obtained by partitioning the cells of the linear-cubic matrix into five groups.

Once the grouping describing patterns of change was formed, a series of one-way ANOVAs was performed to investigate the ability of four of these patterns to account for variance in variables measured initially and throughout the Fall semesters. In order to increase the chance of finding differenced among these variables, students falling into the no change group were not included in these analyses. It was felt that by emphasizing the more extreme patterns of change, more could be learned about the process of the initial breakdown of epistemological dualism.

Five demographic variables measured at Time 1 were included in these analyses: students' self-report of their SAT-Verbal scores, SAT-Quantitative scores, and high school grade point averages; the total number of campus activities students anticipated they would be involved in during the semester; and students' average perceived difficulty of their semester course load. Seven indices of affect and mood over the four time points were also examined in relation to students' patterns of change in dualism. These indices included the within-individual means of the following: state and trait anxiety; pleasant mood; aroused mood; positive mood; negative mood; and overall mood. Last, five indices of personality were examined in relation to students' patterns of change in dualism, also computed by averaging scores across each of the four time points. These included three dimension of locus of control—internal, external-chance, and external-powerful others; tolerance of ambiguity; and multiplicity.

Figure 9. Classification of patterns of linear and cubic change in students' epistemological orientation during the initial transition to college (N = 180).



For the analyses of demographic variables, it was hypothesized that students in both increasing groups would report more difficult course loads, and students in the straight decreasing group would report the greatest levels of course difficulty. Because there is no clear link between epistemological orientation and performance on tests of scholastic aptitude (King, 1978), no significant differences were expected in SAT scores or high school GPA as a function of the pattern of change in dualism. In addition no difference was expected in students' level of campus activity as a function of dualism change pattern.

Regarding measures of affect and mood, it was hypothesized that students in the decreasing patterns would report lower levels among the measures of negative aspects of affect and mood, including state anxiety, trait anxiety, and negative mood due to the apparent accommodation of their epistemological beliefs. Following this hypothesis, students in the straight declining group would report the lowest levels of all groups. Similarly, students in the "accommodating" patterns would report higher levels of positive aspect of mood, including pleasant mood, positive mood, and the measure of overall mood.

Last, the measures of personality expected to be associated with greater decreases in epistemological dualism included internal locus of control, ambiguity tolerance, and agreement with multiplistic epistemological beliefs. Based upon Knefelkamp and Slepitza's (1976) findings, students in these groups were expected to report higher levels of internality than students with increasing patterns of dualistic change. Given their close theoretical relation and differences in dualism related to initial ambiguity tolerance found in the repeated measures ANOVA procedure reported above, it was hypothesized that students with declining patterns would be more tolerant of ambiguity and more multiplistic that students with increasing patterns of change.

The results of the ANOVAs testing for mean differences in demographic variables as a function of patterns of change are presented in Table 17. Because these and all others analyses required performing 16 one-way ANOVAs, an experiment-wise level of Type I error was kept at $\alpha = .05$. This resulted in requiring any single F statistic to have a probability of $p < .003$ to

be considered significant. Given these constraints, no differences among the patterns of change were found in any demographic variables.

The results of the ANOVAs testing for mean differences in measures of affect and mood as a function of patterns of change are presented in Table 18. Although no significant differences were found in any of these seven comparisons, mean values were in the predicted direction hypothesized above for state anxiety, pleasant mood, negative mood, and overall mood. However, the amount of variance accounted for in any one of these variables is negligible.

The results of the ANOVAs testing for mean differences in measures of personality as a function of patterns of change are presented in Table 19. As above, although no significant differences were found among patterns of change for any of the variables assessed, mean values were in the predicted direction for internal locus of control, ambiguity tolerance, and multiplicity.

Results of Sophomore Year Follow-Up Analyses

Because of the relatively small number of individuals' reporting significant linear change in their dualism scores during the first semester, the change in epistemological orientation of the 29 subjects from the Fall 1990 cohort who participated in the Spring 1992 follow-up was examined using the Belsky and Rovine's (1990) modification of IGM. It was hoped that this analysis, covering five time points and extending into the middle of students' sophomore year, would allow for more informative conclusions about epistemological development in the early college years.

A cubic model measuring change across five points was chosen to describe intraindividual growth given the results from the IGM analyses of the Fall 1990 and Fall 1991 cohorts. It was hypothesized that the relative distribution of student's among the previously identified patterns of change would be significantly different between the original and the follow-up samples. The nature of this difference was expected to come in the form of a greater proportion of students' true growth characterized as declining (both straight and fluctuating), with fewer students characterized as having no change and increasing change (again both straight and fluctuating).

Table 17. Results of ANOVAs of demographic differences in linear-cubic dualism change groups (N = 53).

Variable	Change Group	n	Mean	SD	F	p	r^2
SAT-Verbal							
	Fluctuating Decline	11	511	87	1.52	.22	.09
	Fluctuating Increase	12	495	42			
	Straight Decline	15	505	81			
	Straight Increase	10	555	60			
SAT-Quantitative							
	Fluctuating Decline	11	547	93	0.44	.72	.03
	Fluctuating Increase	12	553	89			
	Straight Decline	15	555	67			
	Straight Increase	11	584	85			
High School GPA							
	Fluctuating Decline	11	3.64	0.23	2.63	.06	.16
	Fluctuating Increase	11	3.49	0.27			
	Straight Decline	13	3.30	0.40			
	Straight Increase	11	3.52	0.33			
Campus Activities							
	Fluctuating Decline	13	2.15	1.07	3.33	0.03	.17
	Fluctuating Increase	13	1.38	1.04			
	Straight Decline	15	2.87	1.77			
	Straight Increase	12	1.83	0.94			
Course Difficulty							
	Fluctuating Decline	13	2.62	0.35	1.42	.25	.08
	Fluctuating Increase	13	2.50	0.60			
	Straight Decline	15	2.82	0.49			
	Straight Increase	12	2.80	0.40			

Table 18. Results of ANOVAs of differences in mean anxiety and mean mood measures as a function of linear-cubic dualism change groups (N = 109).

Variable	Change Group	n	Mean	SD	F	p	r^2
State Anxiety							
	Fluctuating Decline	26	41.91	9.91	0.74	0.53	.02
	Fluctuating Increase	28	40.73	7.89			
	Straight Decline	26	38.78	8.97			
	Straight Increase	29	42.08	9.50			
Trait Anxiety							
	Fluctuating Decline	26	42.49	8.19	0.23	.88	.01
	Fluctuating Increase	28	42.48	8.82			
	Straight Decline	26	40.76	9.61			
	Straight Increase	29	42.03	8.26			
Pleasant Mood							
	Fluctuating Decline	26	8.50	6.64	0.73	.54	.02
	Fluctuating Increase	28	8.88	5.94			
	Straight Decline	26	10.58	5.46			
	Straight Increase	29	8.23	7.16			
Aroused Mood							
	Fluctuating Decline	26	22.48	3.00	0.70	.56	.02
	Fluctuating Increase	28	22.85	2.77			
	Straight Decline	26	23.32	2.53			
	Straight Increase	29	22.24	3.31			
Positive Mood							
	Fluctuating Decline	26	8.78	2.80	0.56	.64	.02
	Fluctuating Increase	28	9.00	3.11			
	Straight Decline	26	9.52	2.26			
	Straight Increase	29	8.52	3.42			
Negative Mood							
	Fluctuating Decline	26	14.81	2.00	0.03	.99	< .01
	Fluctuating Increase	28	14.88	1.57			
	Straight Decline	26	14.74	1.95			
	Straight Increase	29	14.85	1.75			
Overall Mood							
	Fluctuating Decline	26	3.69	2.92	1.11	.35	.03
	Fluctuating Increase	28	4.41	2.63			
	Straight Decline	26	4.63	4.49			
	Straight Increase	29	3.34	3.77			

Table 19. Results of ANOVAs of differences in mean locus of control, mean ambiguity tolerance and mean multiplicity as a function of linear-cubic dualism change groups (N = 109)

Variable	Change Group	n	Mean	SD	F	p	r^2
Internal LOC							
	Fluctuating Decline	26	21.45	4.35	0.65	0.58	.02
	Fluctuating Increase	28	21.38	3.80			
	Straight Decline	26	22.59	3.83			
	Straight Increase	29	21.17	4.30			
Chance LOC							
	Fluctuating Decline	26	21.04	3.36	0.93	.43	.03
	Fluctuating Increase	28	21.94	3.28			
	Straight Decline	26	21.05	2.78			
	Straight Increase	29	20.62	2.76			
Powerful Others LOC							
	Fluctuating Decline	26	15.24	4.24	1.13	.34	.03
	Fluctuating Increase	28	17.13	3.53			
	Straight Decline	26	15.62	3.64			
	Straight Increase	29	16.35	4.80			
Ambiguity Tolerance							
	Fluctuating Decline	13	17.09	1.90	1.94	.14	.11
	Fluctuating Increase	13	16.81	1.79			
	Straight Decline	15	18.19	1.93			
	Straight Increase	12	16.47	2.38			
Multiplicity							
	Fluctuating Decline	26	79.60	5.94	2.92	.04	.15
	Fluctuating Increase	28	77.96	7.97			
	Straight Decline	26	81.75	7.45			
	Straight Increase	29	74.46	3.34			

These hypotheses were guided by the predictions made by Perry's scheme concerning later college development, and the consistent findings of very few dualists in studies of epistemological development using sophomore samples (Baxter Magolda & Porterfield, 1988).

A matrix describing the distribution of the linear and cubic growth in dualism of students participating in the follow-up is presented in Figure 10. As with the Fall 1990 and 1991 cohorts, 28 sophomores[16] were classified into one of five patterns of change in dualism scores: no linear change, straight decline, straight increase, fluctuating decline, and fluctuating increase. Table 20 presents the comparison of the relative distribution of students among the five change patterns found in the original first-semester and subsequent follow-up samples. A χ^2 analyses testing the equality of the relative distribution of these samples revealed that they were significantly different, $\chi^2 = 157.96$, $p < .001$. In addition, the effect size estimated from this indicated a strong relationship between sample and distribution of change patterns, $r = .87$. From Table 19 it appears that, in terms of the proportion of students, there was a general decrease in the number of students reporting an increase in dualism when followed into their sophomore year. Because of the small size of the follow-up sample, no additional inferential parametric analyses (i.e., ANOVA) were performed.

Discussion

Comparison of Methods for Analyzing Change

In light of the series of analyses of change just described, what can be said about their relative merit? Does one approach offer the "best" method for the analysis of change? If so, why? And if not, what guidelines can be used to steer researchers towards the approaches that best suit their needs? One way to judge the relative merits of each approach is to refer the performance of each to the summary of their advantages and disadvantages described in Table 1 (Section II). With a cursory examination, this table reveals a definite bias in favor of

---

[16] One participant in the follow-up study was excluded from the IGM analyses as the result of having missing data from two time points in the Fall 1990 study. With only three data points a cubic model could not be fit for this subject.

Figure 10. Matrix crossing linear and cubic types for change in students' dualism scores from the first semester in college to second semester Sophomore year (N=28).

Linear \ Cubic	-3	-2	-1	0	1	2	3	
-3	0	0	0	0	0	0	0	0
-2	0	0	0	0	0	1	0	1
-1	0	0	0	2	6	0	0	8
0	0	0	3	9	2	0	0	14
1	0	0	3	1	0	0	0	4
2	0	0	0	0	0	0	0	0
3	1	0	0	0	0	0	0	1
Total	1	0	6	12	8	1	0	28

Table 20. Comparison of the relative distribution of growth pattern groups obtained during the first semester in college (N = 180) and in the Spring 1992 follow-up (N = 28).

	Fall 1990 & Fall 1991 Cohorts		Spring 1992 Follow-Up	
Pattern of Change in Dualism	n	Proportion	n	Proportion
Fluctuating Decline	26	14%	7	25%
Straight Decline	26	14%	2	7%
No Change	71	39%	14	50%
Straight Increase	29	17%	1	4%
Fluctuating Increase	28	16%	4	14%

approaches nearer to the bottom of the table. But, given the results of this longitudinal study, is this bias warranted?

Each of the advantages and disadvantages of the difference score and the residual change score were born out in this study. Even though these measures were easily computed from two waves of data (i.e., Time 1 and Time 4), it was shown in later nonlinear analyses that the amount and type of change in dualism scores across the semester were greatly underestimated by these two wave designs. Among other things, this inability to accommodate multiwave information may also lie at the heart of "what ails the difference score" and related two wave measures of change (Rogosa, 1988, p. 188).

Results from the examination of one cross-time correlation matrix revealed that a segmented measurement of time is limited in its ability to address questions of individual change over time. Although the cross-time correlation matrix allows for the analysis of multiwave data, its inability to provide useful information about change reveals it limits. Perhaps this approach does have a place in the measurement of change, but I would argue that developmentalists would be best served by treating this technique as an EDA (Tukey, 1977) procedure more helpful for screening data or providing hypotheses about possible underlying true growth than actually measuring change.

The remaining two approaches provided the most information regarding change in epistemological orientation during students' first semester in college. Repeated measures ANOVA was successful in at least examining the data for differences over time, although its typical approach of partitioning variance among cells resulted in time effects being measured only between cells adjacent in time. Like the cross-time correlation matrix, it too could be used successfully as an exploratory device with an eye to generating possible hypotheses to be tested with a more adequate ability to measure individual change.

The modified individual growth modeling technique used here revealed both the strengths and limitations of this general approach. Increasingly more detailed information regarding individual growth was obtained with each addition growth model fit to the data. This illustrates

the ideographic nature of growth modeling, as each individual is allowed to "speak" for him or herself as much as possible. However, some of that unique information was lost when individuals were classified into groups representing different patterns of change. It appears that in choosing among techniques for measuring change, it is perhaps best to rely on several techniques by integrating them. This approach was taken by using analysis of variance to examine differences in grouped estimates of true growth in dualism over students' first semester.

In a final point, much of the difficulty of analyzing change in the constructs chosen here lies in the variability in both the magnitude and direction of change in epistemological orientation. Many methodological treatments of the measurement of change deal exclusively with "clean" data sets constructed to be illustrative and pedagogical. However, when dealing with "less-than-clean" data, the unexpected often happens and much of the methodological machinery begins to whine. As much as anything else, these real data experiences can help to challenge existing methods and therefore improve them.

Epistemological Orientation During the Transition to College

Based upon the results of the various approaches to measuring change, two major conclusions can be drawn about students' epistemological development during the transition to college. The first of these is that a majority of students do not significantly change their attitude towards knowledge during their first semester in college. Although this finding could be a function of the imprecision of the instrument used to measure dualism, it seems more likely that much of what may count for change in Perry's scheme does not occur with time but occurs as the result of later conceptualization of their development. This is to say that development may not be understood forwards, but rather backwards (Kierkegaard, 1967): a student's attitude towards epistemological issues may be independent of the way in which she actually assumes certain epistemological stances towards knowledge and education.

Freeman (1984) arrives at much the same conclusion regarding the epistemological standing of current approaches to life-span developmental psychology. According to Freeman

(1984), when life-span theories like Perry's are viewed as a type of historical analysis, it become clear that they built upon individuals' making meaning through reflection, interpretation, and the narrative reconstuction of experience. In addition, this interpretation seem more plausible when it is stressed that much of Perry's own work and others (Kitchener, 1983; Baxter Magolda, 1990) involves students reflecting back on their educational experiences, and in a very real sense "making meaning" out of those experiences.

Although few significant differences among students' growth were found to be related to any demographic, emotional, or personality variables, a number of patterns were suggested by these data. The first of these patterns was the finding that initial level of ambiguity tolerance was strongly related to interindividual differences in dualism at each of the four assessments. This provides some tentative support for the notion of epistemological dualism being related to a larger authoritarian personality factor.

Second, given the constraints of running multiple inferential tests, there was a suggestion of a possible relation between students' levels of campus activity and their epistemological growth: students presenting decreasing levels of dualism tended to be more involved in activities outside of the classroom. This suggested relationship agrees with the Perry's view of the importance of the pluralism of social experiences as a factor in students' epistemological development. Third, although a similar suggested relationship was found between students' multiplicity and their change in dualism, the questionable psychometric properties of the scale measuring multiplicity make this finding rather inconclusive.

Examining the relation between epistemological orientation and both affect and mood, average levels of anxiety and negative dimensions of mood appeared in general to correspond with hypothesized differences. It is important to note that these average levels were not large enough to be significant above what would be expected from sampling error. Last, a follow-up investigation of the epistemological growth by the sophomore year found a difference between change measured within the initial semester in college and change measured up until the middle of students' second year.

V. CONCLUSION

By comparing the ability of various methodological approaches to measure change in students' epistemological development, this research has shown that many of the conclusions of recent discussions on the measurement of change (Burchinal & Applebaum, 1991; Rogosa, 1988; Willett, 1988) have merit when applied to multiwave data. Further, by examining patterns of change (Belsky & Rovine, 1990), the superiority of the individual growth modeling approach was especially apparent in its ability to describe interindividual differences in both the magnitude and direction of change in students' beliefs about the nature of knowledge.

This benefit has added importance for the empirical investigation of Perry's scheme of development. Although in the totality of the Scheme change may be best characterized along some more complex developmental function (i.e., Perry's metaphor of the expanding helix), the use of the linear-cubic growth model provided empirical support for the recursive nature of development within the transition from dualism to multiplicity originally theorized by Perry (1970, 1981). One question that arises from this finding is to what extent is a linear-cubic model of change valid for describing the remaining developmental transitions in Perry's scheme? In future research on student epistemological development, effort needs to be placed on collecting sufficient multiwave data in order to allow for the investigation of such complex models of growth as used in this study.

In light of this contribution, however, the ability of this investigation to successfully address other issues was limited by several problems in its design. The primary one of these was its circumscribed focus on change occurring across a single semester. As a result, a large proportion of students reported no substantial change in epistemological orientation during that time. Although following a sample of these students into their sophomore year revealed significant differences in the distribution of patterns of change compared to students just assessed in their first semester, the follow-up sample was too small to provide any strong

conclusions about what other psychological domains might be related to epistemological development across the first two years of college.

A second limitation of this study was its reliance on the Checklist of Educational Views as the sole measure of epistemological orientation. Although its construct validity was supported in part by its consistent negative correlation with ambiguity tolerance, the assessment of epistemological development would have been greatly enhanced by the inclusion of more recently developed, qualitative instruments designed to measure Perry positions, like Moore's (1982) Measure of Intellectual Development or Baxter Magolda and Porterfield's (1985, 1988) Measure of Epistemological Reflection.[17] Future examinations of epistemological development during the transition to college would be enhanced by the use of both approaches to assessment.

Turning to the general methodological implications of this investigation, comparisons among these five techniques for measuring change revealed that the individual growth modeling approach supported Rogosa's (1988) claim that it is a "hero" of sorts in the analysis of longitudinal change. The validity of this claim was seen in the ability of IGM to retain an ideographic focus on change and to allow for the specification of increasingly complex models of true growth. But like most heroes, IGM has its weakness as well. In the case of the data presented here, the inability of standard IGM (as well as most other approaches investigated here) to differentiate between interindividual differences in both the magnitude and direction of change led to the use of a grouping procedure in which some information about individual growth was lost. That is not to say that this is an inevitable result in all or even some other domains of developmental research; however, as stressed in Section III, care should be taken to choose growth models that are appropriate for the phenomena being examined.

In all, this research argues for a heightened awareness of the importance of methodological issues in studies of human development. Although most techniques for measuring change can be

---

[17] As part of planned longitudinal study of epistemological development across the college years, MER responses were obtained from participants in the Fall 1991 cohort but not used in any of the results presented here.

easily conducted, the interpretation of what is actually measured and in what sense it corresponds to some underlying developmental process is not as straightforward an issue. It is hoped that the research presented in this dissertation will enable other developmentalists to be more cognizant of the challenges still facing the measurement of psychological change.

LIST OF REFERENCES

Adorno, T. W., Frenkel-Brunswick, E., Levinson, D. J., & Stanford, R. N. (1950). The authoritarian personality. New York: Harper & Row.

Albrecht, F. M. (1960). The new psychology in America, 1880-1895. Unpublished doctoral dissertation, Johns Hopkins University, Baltimore, MD.

Aldwin, C. M., Spiro, A. I., Levenson, M. R., & Bossé, R. (1989). Longitudinal findings from the Normative Aging Study: 1. Does mental health change with age? Psychology and Aging, 4, 295-306.

Baillargeon, R. (1987). Object permanence in 3 1/2 and 4 1/2 month old infants. Developmental Psychology, 23, 655-664.

Baker, R. W., & Siryk, B. (1984). Measuring adjustment to college. Journal of Counseling Psychology, 31, 179-189.

Baldwin, J. M. (1902). Development and evolution. New York: Macmillian.

Baldwin, J. M. (1968). Mental development in the child and the race: Methods and processes (3rd ed.). New York: A. W. Kelley. (Original work published 1906)

Baxter Magolda, M. B. (1990). The impact of the freshman year on epistemological development: Gender differences. The Review of Higher Education, 13, 259-284.

Baxter Magolda, M. B., & Porterfield, W. D. (1985). Measure of epistemological reflection. Unpublished manuscript, Miami University, Oxford, OH.

Baxter Magolda, M. B., & Porterfield, W. D. (1988). Assessing intellectual development: The link between theory and practice. Alexandria, VA: American College Personnel Association.

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

Belsky, J., & Rovine, M. (1990). Patterns of marital change across the transition to parenthood: Pregnancy to three years postpartum. Journal of Marriage and the Family, 52, 5-19.

Bereiter, C. (1963). Some persisting dilemmas in the measurement of change. In C. W. Harris (Ed.), Problems in measuring change (pp. 3-20). Madison, WI: University of Wisconsin Press.

Bochner, S. (1965). Defining tolerance of ambiguity. The Psychological Record, 15, 393-400.

Bryk, A. S., & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. Psychological Bulletin, 101, 147-158.

Bryk, A. S., & Raudenbush, S. W. (in press). Hierarchical linear models for social and behavioral research: Applications and data analysis methods. Newbury Park, CA: Sage.

Buchanan, T. (1986). MacDonald's AT-20: A literature review and assessment of a measure of tolerance of ambiguity. Unpublished manuscript, Baylor University, Waco, TX.

Buchanan, T. (1991, April). Predicting change in postpartum depression: An individual growth curve analysis. Paper presented at the biennial meeting of the Society for Research in Child Development, Seattle, WA.

Bunyan, J. (1939). Pilgrim's progress. New York: Frederick A. Stokes. (Original date 1728)

Burchinal, M., & Applebaum, M. I. (1991). Estimating individual developmental functions: Methods and their assumptions. Child Development, 62, 23-43.

Case, R. (1985). Intellectual development: A systematic reinterpretation. New York: Academic Press.

Chaplin, W. F. (1984). State-Trait Anxiety Inventory. In D. J. Keyser & R. C. Sweetland (Eds.), Test critiques (pp. 626-632). Kansas City, MO: Test Corporation of America.

Charlesworth, W. R. (1992). Darwin and developmental psychology: Past and present. Developmental Psychology, 28, 5-16.

Collins, L. M., & Horn, J. L. (Eds.) (1991). Best methods for the analysis of change: Recent advances, unanswered questions, future directions. Washington, DC: American Psychological Association.

Commons, M. L., Richards, F. A., & Armon, C. (Eds.) (1984). Beyond formal operations: Late adolescent and adult cognitive development. New York: Praeger.

Constantinople, A. (1969). An Eriksonian measure of personality development in college students. Developmental Psychology, 1, 357-372.

Cremin, L. A. (1980). American education: The national experience, 1783-1876. New York: Harper & Row.

Cremin, L. A. (1988). American education: The metropolitan experience, 1876-1980. New York: Harper & Row.

Cronbach, L. J., & Furby, L. (1970). How we should measure change—or should we? Psychological Bulletin, 74, 68-80.

Cutrona, C. E. (1982). Transition to college: Loneliness and the process of social adjustment. In L. A. Peplau, & D. Perlman (Eds.), Loneliness: A source book of theory, research, and therapy (pp. 291-301). New York: Wiley.

Factbook FY 91. (1991). Durham, NH: University of New Hampshire System.

Feldman, K. A., & Newcomb, T. M. (1969). The impact of college on students. San Francisco: Jossey-Bass.

Festinger, L. (1957). A theory of cognitive dissonance. Evanston, IL: Row, Peterson.

Flavell, J. H. (1963). The developmental psychology of Jean Piaget. New York: D. Van Nostrand.

Fletcher, J. M., Francis, D. J., Pequegnat, W., Raudenbush, S. W., Bornstein, M. H., Schmitt, F., Brouwers, P., & Stover, E. (1991). Neurobehavioral outcomes on diseases of childhood: Individual change models for pediatric human immunodeficiency viruses. American Psychologist, 46, 1267-1277.

Francis, D. J., Fletcher, J. M., Stuebing, K. K., Davidson, K. C., & Thompson, N. M. (1991). Analysis of change: Modeling individual growth. Journal of Consulting and Clinical Psychology, 59, 27-37.

Freeman, M. (1984). History, narrative, and life-span developmental knowledge. Human Development, 27, 1-19.

Frenkel-Brunswik, E. (1949). Intolerance of ambiguity as an emotional and perceptual personality variable. Journal of Personality, 18, 108-143.

Gilligan, C. (1977). In a different voice: Women's conceptions of the self and of morality. Harvard Educational Review, 47, 418-517.

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

Gilligan, C., & Attanucci, J. (1988). Two moral orientations: Gender differences and similarities. Merrill Palmer Quarterly, 34, 223-237.

Gilligan, C., & Murphy, J. M. (1979). Development from adolescence to adulthood: The philosopher and the dilemma of the fact. New Directions for Child Development, No. 5, 85-99.

Gilligan, C., Murphy, J. M., & Tappan, M. B. (1990). Moral development beyond adolescence. In C. Alexander, & E. Langer (Eds.), Higher stages of human development (pp. 208-225). New York: Oxford University Press.

Glass, G. V., & Hopkins, K. D. (1984). Statistical methods in education and psychology (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Hays, W. L. (1981). Statistics (3rd ed.). New York: Holt, Rinehart & Winston.

Heffernan, J. M. (1975). An analytical framework for planning and research in higher education. Liberal Education, 61, 493-503.

Heider, F. (1958). The psychology of interpersonal relations. New York: John Wiley & Sons.

Hertzog, C., & Rovine, M. (1985). Repeated-measures analysis of variance in developmental research. Child Development, 56, 787-809.

Hilgard, E. R. (1987). Psychology in America: A historical survey. New York: Harcourt Brace Jovanovich.

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. Developmental Psychology, 27, 236-248.

James, W. (1987). Pragmatism. In B. Kuklick (Ed.), William James: Writings, 1902-1910 (pp. 479-624). New York: Library of America. (Original work published 1907)

Jay, G. M., & D'Augelli, A. R. (1991). Social support and adjustment to university life: A comparison of African-American and White freshman. Journal of Community Psychology, 19, 95-108.

Kierkegaard, S. (1967). [Entry 1025, April 15, 1838]. In H. V Hong & E. H. Hong (Trans. & Eds.), Søren Kierkegaard's journals and papers (Vol. 1). Bloomington, IN: University of Indiana Press.

King, P. M. (1978). William Perry's theory of intellectual and ethical development. New Directions for Student Services, No. 4, 35-51.

King, P. M., Kitchener, K. S., Davidson, M. L., Parker, C. A., & Wood, P. K. (1983). The justification of beliefs in young adults: A longitudinal study. Human Development, 26, 106-166.

King, P. M., Kitchener, K. S., & Wood, P. K. (1991, November). Moral and intellectual development beyond the college years: A ten-year study. Paper presented at annual meeting of the Association for Moral Education, Athens, GA.

Kirk, R. E. (1990). Statistics: An introduction (3rd ed.). New York: Holt, Rhinehart & Winston.

Kirton, M. J. (1981). A reanalysis of two scales of tolerance of ambiguity. Journal of Personality Assessment, 45, 407-414.

Kitchener, K. S. (1982). Human development and the college campus: Sequences and tasks. New Directions for Student Services, No. 20, 17-45.

Kitchener, K. S. (1983). Cognition, metacognition, and epistemic cognition: A three-level model of cognitive processing. Human Development, 26, 222-232.

Kitchener, K. S. (1986). The Reflective Judgment model: Characteristics, evidence, and measurement. In R. A. Mines, & K. S. Kitchener (Eds.), Adult cognitive development: Methods and models (pp. 76-91). New York: Praeger.

Knefelkamp, L. L. (1974). Developmental instruction: Fostering intellectual and personal growth in college students. Unpublished doctoral dissertation, University of Minnesota, Minneapolis.

Knefelkamp, L. L., & Slepitza, R. (1976). A cognitive-developmental model for career development: An adaptation of the Perry scheme. The Counseling Psychologist, 6 (3), 53-58.

Kohlberg, L. (1981). Essays on moral development (Vol. 1): The philosophy of moral development. San Francisco: Harper & Row.

Kohlberg, L. (1984). Essays on moral development (Vol. 2): The psychology of moral development: The nature and validity of moral stages. New York: Harper & Row.

Kohlberg, L., & Ryncarz, R. A. (1990). Beyond justice reasoning: Moral development and consideration of a seventh stage. In C. N. Alexander, & E. J. Langer (Eds.), Higher stages of human development: Perspectives on adult growth (pp. 191-207). New York: Oxford University Press.

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

Levenson, H. (1973). Multidimensional locus of control in psychiatric patients. Journal of Consulting and Clinical Psychology, 41, 397-404.

Levenson, H. (1981). Differences among internality, powerful others, and chance. In H. M. Lefcourt (Ed.), Research with the locus of control construct, Vol. 1. Assessment methods (pp. 15-63). New York: Academic Press.

Linn, R. L., & Slide, J. A. (1977). The determination of the significance of change between pre- and post-testing periods. Journal of Educational Research, 47, 121-150.

Lord, F. (1963). Elementary models for measuring change. In C. W. Harris (Ed.), Problems in measuring change (pp. 21-38). Madison, Wi: University of Wisconsin Press.

MacDonald, A. P., Jr. (1970). Revised scale for ambiguity tolerance: Reliability and validity. Psychological Reports, 26, 791-798.

Mayer, J. D., & Gaschke, Y. N. (1988). The experience and meta-experience of mood. Journal of Personality and Social Psychology, 55, 102-111.

Mayer, J. D., & Salovey, P. (1988). Personality moderates the interaction of mood and cognition. In K. Fiedler, & J. Forgas (Eds.), Affect, cognition, and social behavior (pp. 87-99). Toronto: C. J. Hogrefe.

McCall, R. B. (1977). Challenges to a science of developmental psychology. Child Development, 48, 333-344.

McCartney, K., Buchanan, T., & Jordan, E. A. (in preparation). The role of maternal characteristics in the growth of children's language: An application of individual growth modeling. Unpublished manuscript, University of New Hampshire, Durham, NH.

McDermott, J. J. (1976). The culture of experience: Philosophical essays in the American grain. New York: New York University Press.

Mines, R. A., King, P. M., Hood, A. B., & Wood, P. K. (1990). Stages of intellectual development and associated critical thinking skills in college students. Journal of College Student Development, 31, 538-547.

Moore, W. S. (1982a). The measure of intellectual development. Unpublished manuscript. Center for Applications for Developmental Instruction, Farmville, VA.

Moore, W. S. (1982b). William Perry's cognitive-developmental theory: A review of the model and related research. Unpublished manuscript, University of Maryland, College Park, MD.

Moore, W. S. (1991). Perry network cumulative bibliography and copy service catalog for the Perry scheme of intellectual and ethical development. Unpublished manuscript, Center for the Study of Intellectual Development, Olympia, WA.

Murphy, J. M., & Gilligan, C. (1980). Moral development in late adolescence and adulthood: A critique and reconstruction of Kohlberg's theory. Human Development, 2, 77-104.

Nesselroade, J. R. (1991). Interindividual differences in intraindividual change. In L. M. Collins & J. L. Horn (Eds.), Best methods for the analysis of change: Recent advances, unanswered questions, future directions (pp. 92-105). Washington, DC: American Psychological Association.

Nezu, A. M., Nezu, C. M., & Nezu, V. A. (1987). Depression, general distress, and causal attribution among university students. Journal of Abnormal Psychology, 95, 184-186.

O'Donnell, J. M. (1985). The origins of behaviorism: American psychology, 1870-1920. New York: New York University Press.

Perry, W. G., Jr. (1968). Patterns of development in thought and values of students in a liberal arts college: A validation of a scheme (Contract No. SAE-8973). Washington, DC: U.S. Department of Health, Education, and Welfare, Office of Education, Bureau of Research (ERIC Document Reproduction Service No. ED 024 315).

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

Perry, W. G., Jr. (1977). Comments, appreciative and cautionary. The Counseling Psychologist, 6 (4), 51-52.

Perry, W. G., Jr. (1981). Cognitive and ethical growth: The making of meaning. In A. W. Chickering (Ed.), The modern American college (pp. 76-116). San Francisco: Jossey-Bass.

Phares, E. J. (1976). Locus of control in personality. Morristown, NJ: General Learning Press.

Piaget, J. (1926). The language and thought of the child. New York: Harcourt, Brace.

Piaget, J. (1932). The moral judgment of the child. London: Kegan Paul.

Piaget, J. (1952). The origins of intelligence in children. New York: Norton.

Piaget, J. (1969). Judgment and reasoning in the child. London: Routledge & Kegan Paul. (Original work published 1928)

Piaget, J. (1972). Intellectual evolution from adolescence to adulthood. Human Development, 15, 1-12.

Piaget, J. (1983). The child's conception of the world. London: Routledge & Kegan Paul. (Original work published 1929)

Richards, F. A., & Commons, M. L. (1990). Postformal cognitive developmental theory and research: A review of its current status. In C. N. Alexander, & E. J. Langer (Eds.), Higher stages of human development (pp. 139-161). New York: Oxford University Press.

Riegel, K. F. (1973). Dialectical operations: The final period of cognitive development. Human Development, 16, 346-370.

Roberts, M. L., & Nesselroade, J. R. (1986). Intraindividual variability in perceived locus of control in adults: P-technique factor analyses of short-term change. Journal of Research in Personality, 20, 529-545.

Rogosa, D. (1980). A critique of cross-lagged correlation. Psychological Bulletin, 88, 245-258.

Rogosa, D. (1988). Myths about longitudinal research. In K. W. Schaie, R. T. Campbell, W. Meridith, & S. C. Rawlings (Eds.), Methodological issues in aging research (pp. 171-209). New York: Springer.

Rogosa, D., Brandt, D., & Zimowski, M. (1982). A growth curve approach to the measurement of change. Psychological Bulletin, 92, 726-748.

Rogosa, D. R., & Willett, J. B. (1985). Understanding correlates of change by modeling individual differences in growth. Psychometrika, 50, 203-228.

Rousseau, J. J. (1979). Emile; Or, On education. New York: Basic Books. (Original work published 1672)

Ryan, M. P. (1984a). Conceptions of prose coherence: Individual differences in epistemological standards. Journal of Educational Psychology, 76, 1226-1238.

Ryan, M. P. (1984b). Monitoring text comprehension: Individual differences in epistemological standards. Journal of Educational Psychology, 76, 248-258.

SAS Institute, Inc. (1985). SAS® user's guide: Statistics, Version 5 edition. Cary, NC: SAS Institute Inc.

Sayer, A. G., & Willett, J. B. (1991, April). Impact of chronic illness on change and adjustment: Integrating growth modeling and covariance-structure analysis to investigate individual differences in growth. Paper presented at the biennial meeting of the Society for Research in Child Development, Seattle, WA.

Schaie, W. (1965). A general model for the study of developmental problems. Psychological Bulletin, 64, 92-107.

Shakow, D., & Rapaport, D. (1964). The influence of Freud on American psychology. New York: Macmillian.

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). Biometrika, 52, 591-611.

Singer, J. D., & Willett, J. B. (1991). Modeling the days of our lives: Using survival analysis when designing and analyzing longitudinal studies of duration and the timing of events. Psychological Bulletin, 110, 268-290.

Snarey, J. R. (1985). Cross-cultural universality of social-moral development: A critical review of Kohlbergian research. Psychological Bulletin, 97, 202-232.

Sokal, M. M. (1987). James McKeen Cattell and mental anthropometry: Nineteenth-century science and reform and the origins of psychological testing. In M. M. Sokal (Ed.), Psychological testing in American society, 1890-1930 (pp. 21-45). New Brunswick, NJ: Rutgers University Press.

Spielberger, C., Gorsuch, R., Lushene, R., Vagg, P., & Jacobs, G. (1983). The State-Trait Anxiety Inventory. Palo Alto, CA: Consulting Psychologists Press.

Stephenson, B. W., & Hunt, C. (1977). Intellectual and ethical development: A dualistic curriculum intervention for college students. The Counseling Psychologist, 6 (4), 39-42.

Stine, W. W. (1989). Meaningful inference: The role of measurement in statistics. Psychological Bulletin, 105, 147-155.

Tabachnick, B. G., & Fidell, L. S. (1989). Using multivariate statistics (2nd ed.). New York: Harper & Row.

Tappan, M. B. (1989). Stories lived and lives told: The narrative structure of late adolescent moral development. Human Development, 32, 300-315.

Touchton, J. G., Wertheimer, L. C., Cornfield, J. L., & Harrison, K. H. (1977). Career planning and decision-making: A developmental approach to the classroom. The Counseling Psychologist, 6, (4) 42-47.

Tukey, J. W. (1977). Exploratory data analysis. Reading, MA: Addison-Wesley.

Vidal, F., Buscaglia, M., & Voneçhe, J. J. (1983). Darwinism and developmental psychology. Journal of the History of the Behavioral Sciences, 19, 81-94.

Waterman, A. S., & Waterman, C. K. (1971). A longitudinal study of changes in ego identity status during the Freshman year in college. Developmental Psychology, 5, 167-173.

Widick, C. (1977). The Perry Scheme: A foundation for developmental practice. The Counseling Psychologist, 6, 4, 35-38.

Widick, C. C. (1975). An evaluation of developmental instruction in a university setting. Unpublished doctoral dissertation, University of Minnesota, Minneapolis.

Wilkinson, W. K., & Schwartz, N. H. (1987). The epistemological orientation of gifted adolescents: An empirical test of Perry's model. Psychological Reports, 61, 976-978.

Willett, J. B. (1987-1991). Analysis of multiwave growth data using individual growth modeling [Computer program]. Cambridge, MA: Harvard Graduate School of Education

Willett, J. B. (1988). Questions and answers in the measurement of change. In E. Z. Rothkopf (Ed.), Review of research in education (pp. 345-422). Washington: American Educational Research Association.

Willett, 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.

Willett, J. B. (1991, April). Concepts and models in the measurement of individual change: That was then, this is now. Paper presented at the biennial meeting of the Society for Research in Child Development, Seattle, WA.

Willett, J. B., Ayoub, C. C., & Robinson, D. (1991). Using growth modeling to examine systematic differences in growth: An example of change in the functioning of families at risk of maladaptive parenting, child abuse, or neglect. Journal of Consulting and Clinical Psychology, 59, 38-47.

Wohlwill, J. F. (1973). The study of behavioral development. New York: Academic Press.

APPENDICES

APPENDIX A

EXAMPLE OF SAS® 6.0 COMPUTER PROGRAM FOR CONDUCTING GROWTH CURVE ANALYSIS
(SOURCE: WILLETT, 1987/1991)

```
DATA STUDY23;
 INFILE '1STSEM.DAT';
 INPUT #1 COHORT $ SUBJID
        #2 AGE GENDER MAJOR RESIDE MOCC FOCC HSSIZE SAT_V SAT_Q HSPGA HOUSING
          JUNEOR FCAMP FDAYS FYSEM ACADEM SPECINT GREEK MUSART POLITIC RELIG
          SERVICE CLBSPRT VARSPRT WORKSTUD OTHER COURSE1 COURSE2 COURSE3
          COURSE4 COURSE5 ACTIVE COURSDIF
        #3 TIME1 $ INTERN1 CHANCE1 OTHERS1 STATE1 TRAIT1 MOOD1 PLEAS1 AROUS1
          POS1 NEG1 CLEV1 MEOS1 AMBIG1
        #4 TIME2 $ INTERN2 CHANCE2 OTHERS2 STATE2 TRAIT2 MOOD2 PLEAS2 AROUS2
          POS2 NEG2 CLEV2 MEOS2 AMBIG2
        #5 TIME3 $ INTERN3 CHANCE3 OTHERS3 STATE3 TRAIT3 MOOD3 PLEAS3 AROUS3
          POS3 NEG3 CLEV3 MEOS3 AMBIG3
        #6 TIME4 $ INTERN4 CHANCE4 OTHERS4 STATE4 TRAIT4 MOOD4 PLEAS4 AROUS4
          POS4 NEG4 CLEV4 MEOS4 AMBIG4
     #7 TIME5 $ INTERN5 CHANCE5 OTHERS5 STATE5 TRAIT5 MOOD5 PLEAS5 AROUS5
          POS5 NEG5 CLEV5;

IF COHORT='S3' THEN MOOD1=MOOD1-10;
IF COHORT='S3' THEN MOOD2=MOOD2-10;
IF COHORT='S3' THEN MOOD3=MOOD3-10;
IF COHORT='S3' THEN MOOD4=MOOD4-10;

IF COHORT='S2' THEN ID=90000+SUBJID;
IF COHORT='S3' THEN ID=91000+SUBJID;

DATA STUDY23;
 SET STUDY23;
 IF NMISS (OF CLEV1 CLEV2 CLEV3 CLEV4) GE 2 THEN DELETE;

PROC SORT DATA=STUDY23;
 BY ID;

DATA MULT_X;
 SET STUDY23;
 KEEP ID CLEV1 CLEV2 CLEV3 CLEV4;
```

```
DATA UNIV_X;
 SET MULT_X;
 ARRAY SCORES{4} CLEV1 CLEV2 CLEV3 CLEV4;
 ARRAY TIMES{4} _TEMPORARY_ (1 2 3 4);
  DO I = 1 TO 4;
       X = SCORES{I};
       T = TIMES{I} - 2;
       TIME = TIMES{I};
       T2  =  T**2;
       OUTPUT;
  END;

PROC REG COVOUT NOPRINT OUTEST=TEMP1 DATA=UNIV_X;
TITLE 'INDIVIDUAL GROWTH MODEL REGRESSION ANALYSES: CLEV';
 MODEL X = T;
 BY ID;

DATA TEMP2;
 SET TEMP1;
 IF _TYPE_ = 'PARMS';
 SLOPE = T;
 KEEP ID SLOPE;

DATA TEMP3;
 SET TEMP1;
 IF _TYPE_ = 'COV' AND _NAME_ = 'T';
 V_SLOPE = T;
 SE_SLOPE = SQRT(V_SLOPE);
 KEEP ID SE_SLOPE;

PROC REG NOPRINT OUTEST=TEMP4 DATA=UNIV_X;
 MODEL X = T / SELECTION = RSQUARE SSE MSE;
 BY ID;

DATA TEMP4;
 SET TEMP4;
 W_SSE = _SSE_;
 W_MSE = _MSE_;
 W_DFE = _EDF_;
 W_RSQ = _RSQ_;
 KEEP ID W_SSE W_MSE W_DFE W_RSQ;

DATA STUDY23;
 MERGE STUDY23 TEMP2 TEMP3 TEMP4;
 BY ID;

PROC REG NOPRINT COVOUT OUTEST=TEMP6 DATA=STUDY23;
 MODEL SLOPE=CLEV1;

DATA TEMP6;
 SET TEMP6;
 IF _TYPE_='PARMS';
  B_MSE= _RMSE_**2;
 KEEP B_MSE;
```

```
DATA STUDY23;
 IF _N_=1 THEN SET TEMP6;
 SET STUDY23;
  WP = B_MSE/(B_MSE + SE_SLOPE**2);

PROC PRINT DATA=STUDY23;
TITLE 'CLEV GROWTH MODELING DATA';
 VAR ID CLEV1 CLEV2 CLEV3 CLEV4 SLOPE SE_SLOPE W_RSQ WP;

RUN;
```

APPENDIX B

FORMULAE FOR COMPUTING MEAN, LINEAR, QUADRATIC AND CUBIC ORDINAL COMPONENT SCORES (SOURCE: M. ROVINE, PERSONAL COMMUNICATION, FEBRUARY 10, 1992)

Let X_1, X_2, X_3, and X_4 represent some dependent measure obtained at four equally spaced points in time.

The following contrast coefficients are used to to specify linear, quadratic, and cubic models:

	1	2	3	4
Linear	-3/SQRT(20)	-1/SQRT(20)	1/SQRT(20)	3/SQRT(20)
Quadratic	1/2	-1/2	-1/2	1/2
Cubic	-1/SQRT(20)	3/SQRT(20)	-3/SQRT(20)	1/SQRT(20)

The mean, linear, quadratic, and cubic ordinal component scores (OCS) for X over time are obtained using the following formulae:

Linear OCS = (X_1* Linear 1) + (X_2* Linear 2) + (X_3* Linear 3) + (X_4* Linear 4)

Quadratic OCS = (X_1* Quadratic 1) + (X_2* Quadratic 2) + (X_3* Quadratic 3) + (X_4* Quadratic 4)

Cubic OCS = (X_1* Cubic 1) + (X_2* Cubic 2) + (X_3* Cubic 3) + (X_4* Cubic 4)