

*Existing student development assessment techniques are reviewed and evaluated in terms of developmental assessment issues, measurement methods, and costs.*

## *Student Development Assessment Techniques*

*Robert A. Mines*

The student development practitioner has been challenged to use developmental theories for programming since the early to mid-1970s. In addition to applying developmental models in programming, the student development practitioner is also faced with issues of accountability and the call to use sophisticated evaluation techniques (Mines, Gressard, and Daniels, 1982). To meet these demands, the practitioner must turn to student development assessment techniques. This chapter reviews the existing techniques, introduces the practitioner to special issues related to developmental assessment, and outlines the next phase of instrument development needed in the field. This chapter assumes a pragmatic/utilitarian perspective on the use of assessment techniques. Within the framework of pragmatism and direct service utility, the existing instruments are evaluated by more stringent standards than are necessary for research purposes alone. The techniques not only need to be reliable and valid, they must be easy to administer, score, and interpret, or else they will not be of much help to the people who must conduct the administrative and program evaluations.

### **Developmental Assessment Issues**

The assessment of developmental phenomena poses additional problems beyond those encountered within the traditional state-trait,

achievement testing, or behavioral assessment perspectives. Kitchener (see her chapter in the present volume) suggests that the length of time for cognitive stage change or task resolution almost precludes that a single student services level one program (Wirtz and Magrath, 1979) would have a significant impact on a selected developmental domain. Development does not end with graduation from college, so we are only measuring a specific segment of the developmental process. The data in the area of social-cognitive development support the assumption that stages are complex, not simple (see, for example, Mines, 1980; Rest, 1976). Any individual may manifest various stage levels in a given context and across content areas (that is, developmental decalage). Furthermore, the assumption that individuals progress through stages but do not regress is not supported. The attainment or resolution of developmental tasks is a complex phenomenon consisting of cognitive attitudinal and reasoning process changes as well as behavioral changes (for example, Chickering's vector of freeing of interpersonal relationships as discussed in Mines, 1978). Assessing developmental task resolution requires a complex multilevel approach.

In summary, the assessment of developmental stage change or task resolution presents general problems, such as the slowness of developmental change, the complex stage phenomenon, the issue of decalage (that is, stage change in one content area but not in another) and the assessment of cognitive and behavioral interaction for developmental task resolution. Test format and scoring variations also contribute to problems in assessing developmental change.

### **Assessment Formats for Measuring Development**

Developmental researchers and student services practitioners are faced with several psychometric dilemmas. Are we trying to assess a developmental process (that is, a social-cognitive stage), or are we trying to "capture a developmental moment" as a landmark or sign of a given individual's development? Any time we try to reduce a process phenomenon to a static description we will introduce measurement error into our assessment.

The methodology format determines the type of information we will obtain. Rest (1976) distinguishes between preference, comprehension, and spontaneous use (production) of stage responses. An individual will prefer a higher stage response than he or she can comprehend or produce. It follows that one can comprehend a higher stage response than one can produce. The most conservative and taxing measure of stage level(s) is to have the individual produce his or her stage process in responding to the test items. Various formats have been used that elicit

preference (Likert-type scales), comprehension (asking subjects to paraphrase or match statements), or production (open-ended or structured interviews). Rest (1976) states that "different tasks — manifest the acquisition of new ideas at different points of consolidation" (p. 202). The variety of formats eliciting different levels of stage acquisition limits the meaningfulness of inferences that can be made about stage level without a consideration of the task employed.

**Production Formats.** Open-ended or semistructured interview or sentence completion formats require students to produce a "stage typical" response. These formats assume that students cannot produce a response higher than their stage level. The students can produce lower stage responses. Typically, these assessment data are rated by two or more raters to produce a stage score.

There are two major advantages of this technique. It provides an open-ended data source for the refinement of the theory. It can be rated reliably (see, for example, Loevinger, 1976; Mines, 1981; Rest, 1976). The production format has its primary utility in basic developmental research. It is time consuming in terms of interviewer and/or rater time. It can also be costly to learn to rate the data or have the data rated (such as the Reflective Judgment Interview, Mines, 1980). This type of assessment format is neither practical nor efficient for most student services programming evaluation.

**Preference and Comprehension Formats.** Preference or comprehension formats are usually presented in a Likert-type scale or a multiple choice format. This type of format is used when the theory and stage typical responses have been identified and the purpose of the assessment is to "systematically inventory a subject's reactions to a standardized set of stimuli statements" (Rest, 1976, p. 201). In other words, when the purpose is to classify a person's stage level, a preference or comprehension test is preferable.

The advantage of this type of format is that it uses a consistent set of stage typical statements that minimizes the problem of having a subject wander or an interviewer using his or her clinical judgment to pursue ambiguous responses. The objective format does not allow any determination of the underlying developmental process used to arrive at a decision. The type of format used for a given developmental assessment technique affects the stage classification of a student.

### **Scoring Methods**

Another area of controversy in developmental assessment that has a direct impact on stage classification is the scoring method. There are a variety of scoring schemes in use. These vary from using the high-

est scored stage, the modal level of stage usage (Loevinger, 1976), the percentage of highest stage exhibited (Rest, 1976), the predominant stage plus the second most used stage — which is then averaged by stage and across raters (Mines, 1981) — the use of cutting scores that use cumulative distributions of stage typical responses (Loevinger, 1976), or using a strong scalogram analysis that presents tasks that are stage typical in and of themselves (Fischer, Hand, and Russell, forthcoming).

Each of these methods has its liabilities, and only a few have any compelling merit beyond their initial heuristic value. The use of the highest stage score assumes that the student is motivated to produce his or her highest stage responses, which is not always the case (Mines, 1980). A major problem with the use of the highest stage is the assumption of a simple stage model. None of the major theorists (Fischer, Hand, and Russell, forthcoming; Loevinger, 1976; Rest, 1976) assumes a simple stage model. Furthermore, they argue that development is uneven and varies across content domains, so that using a single stage score does not represent the complexity of the phenomenon. The use of the modal stage response has the same simple stage assumption problems. The modal stage response underestimates the highest stage production or comprehension level of the subject. The implementation of the percentage of highest stage level produced is a positive move toward capturing the complexity of stage properties. The sole use of the highest stage percentage ignores the percentage of lower stage responses exhibited. The use of multiple stages in determining a score is used, for example, in reflective judgment scoring rules (see King, 1977). However, the complexity of the stage level is then diluted by averaging the stage levels across two raters (Mines, 1981). This results in a conservative estimate of stage functioning that is also affected by subject motivation and decalage problems across the test items. Loevinger (1976) addresses the complex stage scoring problem by using ogive rules of cumulative distributions. The use of these cutting scores is an important refinement in assessing developmental stages. Unfortunately, without a breakdown of stage typical responses, it is still difficult for the practitioner to use ogive rules in any sophisticated manner, because he or she is still left with a stage score that does not convey the intricacy or the interplay of the different stage skills.

Fischer, Hand, and Russell (forthcoming) have offered a variety of innovative scoring procedures in the cognitive development domain that may have potential in other social-cognitive areas as well as in developmental task models. Fischer and his colleagues have suggested that a strong scalogram could be used. This procedure predicts a sequence of

steps in the acquisition of a developmental skill. A separate task is designed to assess each step. Each student's performance should then fit a Guttman Scale. This method eliminates the scoring algorithm problems discussed previously. It also eliminates the problem of using one developmental task to differentiate all of a model's developmental stages. "When every developmental stage is assessed independently, the assumption (the use of a single developmental task) is no longer a problem since it becomes a hypothesis to be tested" (Fischer, Hand, and Russell, forthcoming, p. 25). To date, such an approach to student development assessment does not exist in a format that would be useful to student services practitioners. Fischer and his colleagues present two other techniques/scoring methods that hold promise and are discussed in the recommendations for instrument development section of this chapter.

### **Student Development Assessment Techniques**

Test format and scoring method issues have been briefly discussed to provide a conceptual framework for evaluating current student development assessment techniques. Due to limited space, selected theoretical domains and techniques are reviewed. The areas and instruments reviewed are: intellectual development (the Measurement of Intellectual Development [MID], the Reflective Judgment Interview), moral development (the Defining Issues Test), ego development (Loevinger and Wessler's Sentence Completion Test), and developmental task models (that is, Chickering's Vectors, the Student Development Task Inventory, and the "Iowa Instruments").

#### ***The Measurement of Intellectual Development (MID)***

King (1978) notes that over eight different assessment methods have been used to determine the cognitive development position scores corresponding to Perry's (1970) theory of intellectual and ethical development. While the original interview methodology continues to be used to replicate and extend Perry's work (Clinchy, 1981; Mentkowski, 1981), it is time consuming and expensive, especially for applied situations. The major work in developing a practical alternative instrument was initiated by Knefelkamp (1974) and Widick (1975), who developed a paper-and-pencil measure that could be used both for research purposes and for feedback to faculty about how their students made meaning in the classroom. The instrument they created is called the Measure of Intellectual Development and, as the title implies, it focuses on the intellectual assets of the Perry model (positions 1-5).

**Purpose and Nature of the Instrument.** The MID is a semi-structured generation/production instrument designed to assess cognitive stage level in three specific domains: decision making, careers, and classroom learning. Each domain takes approximately fifteen minutes to complete. The MID essays are rated independently by two trained raters who then confer to reach a consensus rating. Each essay is given a three-digit rating that allows for an "elongation" of the Perry scheme (Knefelkamp, 1974). This scoring system provides stable position ratings as well as two transitional steps between each stable position. Standardized normative data are not available currently, although the instrument has been used with a variety of traditional-aged and nontraditional-aged, as well as graduate, students at numerous institutions throughout the country. The general cross-sectional results indicate freshmen to be largely in transition between positions 2 and 3, while juniors and seniors are in transition between positions 3 and 4 and in stable position 4. Sex differences have not been found. However, the authors point out that the instrument is *not* appropriate for international students, as cross-cultural differences (especially language) are likely to produce significant confounding effects.

**Reliability.** The MID is more reliably scored than most instruments assessing complex phenomena. The reliability studies conducted to date fall into one of three categories: correlations with interview ratings, correlations with external experts, and interrater reliability data of various kinds (see Table 1).

Clearly, the absolute agreement criterion is the most stringent, but, considering the complexity of assessing the degree to which a given subdominant position is reflected in transitional essays, the authors of the MID consider the dominant position agreement the most reasonable measure of interrater reliability.

**Validity.** A variety of approaches have been taken in addressing the issue of validity for the MID. Table 2 shows correlations representing convergent/divergent validity data between the MID and other construct and developmental models. The MID correlates moderately well with conceptual level (.51), with the Defining Issues Test (DIT) in one study (.45, Meyer, 1977), but not in another (.13, Wertheimer, 1980), and with a measure of ego development (.30) by Wertheimer (1980). The correlations shown in Table 2 suggest that, while there is some overlap among the various developmental dimensions, they seem to be reasonably distinct domains. The relationships with empathy and locus of control are consistent with the implications of the Perry (1970) scheme and provide additional evidence to suggest the validity of the MID.

**Table 1. Measure of Intellectual Development—Reliability Data**

	<i>Mentkowski (1981)</i>	<i>Moore (1982)</i>	<i>Slepitz (1976)</i>	<i>Wertheimer (1976)</i>	<i>Widick (1975)</i>	<i>Stephenson, Hunt (1977)</i>	<i>Allen (1982)</i>	<i>Meyer (1977)</i>
Correlations with interview ratings	NA*	NA	.74	.77	NA	NA	NA	NA
Correlations with expert raters	.45 <sup>1</sup> .76 <sup>2</sup>	NA	NA	NA	NA	NA .73 – .87 <sup>2</sup>	.42 – .64 <sup>1</sup>	NA
Interrater agreement	.43 <sup>1</sup> .74 <sup>2</sup>	.58 <sup>1</sup> .83 <sup>2</sup>	.93	.93	.35 – .62 <sup>1</sup> .87 – 1.00 <sup>2</sup>	.82	NA	.91 – .93

\*NA: Not Applicable

<sup>1</sup>Absolute agreement

<sup>2</sup>Dominant position agreement

**Table 2. Validity Data: Comparisons Between MID Ratings and Various Constructs/Models**

	<i>Widick (1975)</i>	<i>Wertheimer (1980)</i>	<i>Allgire (1977)</i>	<i>Meyer (1977)</i>	<i>Mason (1978)</i>	<i>Viesar (1978)</i>	<i>Bogar (1981)</i>
Ego Development (Loevinger)	NA*	.30	NA	NA	NA	NA	NA
Moral Development (Kohlberg, DIT)	NA	.13	NA	.45	NA	NA	NA
Conceptual Level (Schroder, Driver, and Streufert, 1967)	.51	NA	NA	NA	NA	NA	NA
Empathy	NA	NA	NA	NA	.14, .32	.13	.31 <sup>1</sup>
Locus of Control	NA	NA	NA	NA	-.19, -.54	NA	NA
Information Responsiveness <sup>2</sup>	NA	NA	.42	NA	NA	NA	NA

NA\*: Not Applicable

<sup>1</sup>Attitudes Towards Women Scales (AWS)

<sup>2</sup>measured by the WAIS



Face validity, criterion group differences, and experimental enhancement studies are other methods of establishing the validity of the MID. The MID's particular focus on classroom learning and student-generated open-ended responses lends critical face validity to the instrument. Cross-sectional and short-term longitudinal studies summarized by Moore (1982) reflect expected freshman-senior differences predicted by the Perry scheme. In several experimental studies, the MID has been used to examine differences in pre-post gains related to developmentally designed classroom experiences. The differential gains in the theoretically predicted direction add further credence to the instrument's relation to its underlying construct (see Table 3).

The major strength of the MID is that it has incorporated a standard structure in a written response format for a specific content domain during the theory development phase of the Perry (1970) research. It yields production type responses that are heuristically rich. It can be consistently scored by trained raters with an acceptable percentage of agreement. The MID can easily be administered in a group situation, which is a plus for the student services practitioner. The MID can be scored at the Center for Applications of Developmental Instruction for a reasonable fee (\$3.00 per protocol). See the final chapter in this volume for the complete address. The MID has resolved one aspect of the decade issue in adult intellectual development by focusing on specific content domains (that is, classroom, career). Student services practitioners should use the MID in programs designed to impact on these areas. The validity of using it in other areas is untested at this time. Further components could be developed for areas such as residence services, counseling centers, and so on.

The major liabilities of the MID are its scoring system and the time expense of learning to use the scoring system. The scoring system

**Table 3. Experimental Enhancement Studies  
Using the MID**

	<i>Knefelkamp Widick (1975)</i>	<i>Stephenson, Hunt (1977)</i>	<i>Touchton and others (1977)</i>	<i>Clement and others (1977)</i>
Average Stage Movement	.85 <sup>1</sup> , .79 <sup>2</sup>	.85 <sup>3</sup> , .42 <sup>4</sup> , .12 <sup>4</sup>	.59 <sup>3</sup> , .39 <sup>5</sup> , .17 <sup>4</sup>	.54

<sup>1</sup>Dualist treatment

<sup>2</sup>Relativist treatment

<sup>3</sup>Experimental treatment

<sup>4</sup>Control group

<sup>5</sup>Quasi-experimental group

underestimates the level of complexity the individual is capable of understanding but is a reasonably accurate estimate of a person's ability to *produce* specific levels of cognitive complexity, as is often required in classroom settings. Users of the MID need to recognize that students may well be able to comprehend or express a preference for a higher level of cognitive complexity. The MID also does not differentiate among the domains of cognitive complexity, specific cognitive skills, epistemology, and metaphysical assumptions about reality. These domains appear to be separate yet interrelated in adult intellectual development (Kitchener and King, 1981). The scoring system does not represent the complexity of stage interaction that appears in early adulthood; rather, the rating criteria for the MID are based on narrowly defined, specific content domains that assume a single stage model within each domain.

The next steps in the continuing development of the MID will examine the relationship between the MID ratings of Perry (1970) position and critical thinking or writing skills. What, for example, are the necessary concomitant skills for true contextually relativistic reasoning? Additional work on the rating and scoring system is planned to clarify the distinctions between stage/position cues and style/topology cues. Also, more work is planned to collect normative data from nontraditional learners—that is, older students and noncollege groups.

### *Reflective Judgment Interview (RJI)*

The reflective judgment model of young adult intellectual development has a great deal of promise. It is partially based on Perry's (1970) model but does not have the conceptual confounding with intellectual and identity concerns that Perry's model has. Over ten dissertations and research projects have been completed to date using the reflective judgment interview (RJI). (See, for example, Brabeck, 1980; Kitchener and King, 1981; Mines, 1980; Strange, 1978; Welfel, 1979; and Wood, 1980).

*Purpose and Nature of the Instrument.* The RJI is copyrighted by Karen Kitchener and Patricia King. Information regarding it can be obtained from Kitchener at GCB 112, School of Education, University of Denver, Denver, CO 80208. The RJI is a semistructured interview and elicits production-type data to provide a description of the subject's intellectual stage functioning. The interview is administered individually and takes about one hour to complete. During the interview, one of four dilemmas (the dilemmas are of a social issue nature and can be found in King, 1977) are presented verbally while the subject follows along on an

identical written copy. After each dilemma is read, the interviewer asks a series of standard probe questions. The subject's responses are tape recorded separately. All of the tape recordings (that is, four per subject) are transcribed and rated blind by certified raters.

Kitchener and King require that anyone using the RJI use certified interviewers and raters or become certified themselves in order to assure comparability of results across the studies. At this time, there is no charge for the certification process other than mailing and phone expenses. Certified raters have charged from \$7.00 to \$25.00 per subject to rate the RJI. The major expenses of the RJI are the costs of audio tapes, recording equipment, transcription of tapes, duplicating costs, and time.

The RJI assumes that reflective judgment stages are complex, as opposed to simple stages. This assumption is not realized in the current scoring rules. The current scoring rules require the raters to include a major stage response and a second major stage response for each dilemma that are then averaged across dilemmas and across raters, resulting in a single stage score. In a critique of this psychometrically conservative treatment of the RJI raw scores, Mines (1980) notes that the interview responses are more complex than the averaging of the dilemma scores indicates.

The RJI has been developed on traditional-age high school, college, and graduate student populations and on two noncollege populations (see Lawson, 1980; Shoff, 1979). Normative data in and of itself has not been compiled, although Welfel (1981) has reviewed the data in a study-by-study format. The majority of the research has been done at large state universities (for example, the University of Minnesota, the University of Iowa, the University of Utah), with the exception of Brabeck's (1980) study on a small liberal arts campus.

**Reliability.** As an interview procedure requiring the use of two raters, the RJI and rating rules are more reliably scored with a high interrater agreement than most techniques that assess complex processes. The RJI has moderate to high internal consistency as determined by the dilemma total correlations. (See Table 4.) The Cronbach alpha estimate of reliability indicates the interrelationship among a set of items comprising a scale. The higher the value, the more likely that the items are measuring the same construct. Measures of equivalence or internal consistency are high using Cronbach's alpha. (See Table 5.) The Interjudge reliability of the RJI has ranged from a high of .96 (King, 1977) to moderate .53 to .63 (Strange, 1978). (See Table 6.) Finally, the interrater agreement scores of the RJI are moderately high for the first round of rating. (See Table 7.)

**Table 4. Dilemma-Total Correlations**

Dilemma	King (1977)	Strange (1978)	Shoff (1979)	Mines (1980)	Brabeck (1980)
1	.90	.67, .78, .45	.62	.71	.77
2	.89	.66, .55, .48	.83	.75	.76
3	.91	.57, .62, .38	.74	.81	.73
4	.92	.78, .42, .58	.84	.75	.78

**Validity.** The validity of the RJI is intertwined with the validity of the reflective judgment model. To validate a developmental model, cross-sectional and longitudinal data are needed as well as convergent and divergent validity data. The majority of the research on the construct and the instrument has been cross-sectional (Kitchener, 1981). These studies support the age and education trends in the theoretically predicted direction. One longitudinal study has been completed (King and others, 1982). Stage change was in the appropriate direction. The convergent and divergent validity studies have focused on four major competing constructs: formal operations, verbal ability, general academic aptitude, and critical thinking skills (Brabeck, 1981). In Brabeck's (1981) review, she notes that formal operations (King, 1977), verbal reasoning (Kitchener, 1977), general academic ability (see Brabeck, 1980; Mines, 1980; Welfel, 1979), and critical thinking skills (Brabeck, 1980; Mines, 1980) do not account for differences in reflective judgment scores and probably represent a necessary but not sufficient relationship to advanced intellectual development. This advanced intellectual development could be due to a combination of additional cognitive skills development (Mines, 1980) and changes in epistemological assumptions across stage levels (Brabeck, 1980; Kitchener, 1981).

**Summary and Evaluation.** The RJI generates theoretically rich data that has strong heuristic value for continued research on intellectual development. The RJI is sound psychometrically and has adequate reliability and good interrater agreement. The implementation of a certification process for raters and interviewers ensures that variations in ratings across studies are more likely to be due to actual affects than to

**Table 5. Internal Consistency**

Overall Cronbach's alpha	King (1977)	Strange (1978)	Shoff (1979)	Mines (1980)	Brabeck (1980)
	.96	.63 to .85	not rep.	.71	.77

**Table 6. Overall Interjudge Reliability**

<i>King (1977)</i>	<i>Strange (1978)</i>	<i>Shoff (1979)</i>	<i>Mines (1980)</i>	<i>Brabeck (1980)</i>
.96	.53 to .63	.51	.83	.90

rating or interviewing error. This will enhance the comparability of the studies, thus minimizing problems that have existed, for example, in the Perry (1970) research.

The RJI is expensive in terms of training, time of administration, transcription time and costs, and time and cost of rating. It is more appropriate for student services research than for program evaluation. The interview format is semistructured and allows for the stimulus to be presented to the subject in various ways. However, the responses may be due to factors other than differences in stage level. Brabeck (1980) and Mines (1980) have both suggested that ambiguities exist in the rating rules at the middle and upper ends of the model. Brabeck (1980) notes that the individual dilemmas of the RJI warrant further examination of their measurement characteristics, as they have a low to moderate intercorrelation.

The next step in RJI development and refinement involves the resolution of conceptual as well as measurement issues. The major conceptual issue involves the concept of stage and the representation of the stage(s) in the scoring scheme. The current scoring system needs to be revised, as it underestimates the highest stage of production by the subject and also reduces the complexity of the response to a single stage score through the averaging of the dilemma scores. The scoring rules need to reflect the complexity of the model and of the optimal stage usage of the subject.

The unit of analysis in the rating of the RJI also needs to be clearly defined. It is not clear whether each thought, each sentence, each paragraph, or the total transcript constitutes the phenomenon that is rated. As the rating procedure currently exists, raters have focused on all of the above in varying degrees and then come up with a rating for the dilemma. Given this variability, it is a credit to the scoring rules that the interrater agreement is as high as it is.

**Table 7. First Round Interrater Agreement**

<i>King (1977)</i>	<i>Strange (1978)</i>	<i>Shoff (1979)</i>	<i>Mines (1980)</i>	<i>Brabeck (1980)</i>
.77	.70, .76, .64	.74	.71	.76

Currently an objective reflective judgment test does not exist. However, two groups of researchers, Schmidt and Davis at the University of Minnesota and Strange and King, Bowling Green State, are independently developing objective format reflective judgment tests. The development of an objective format is necessary to provide a practical intellectual development assessment technique that the student services practitioner can use in program evaluation.

The development of an objective instrument needs to answer the question: Is the test going to have a preference, comprehension, or production format? The answer to this question has a direct bearing on the construction of items and the stage property inferences that can be made.

The issue of recognition versus production tests was played out in the moral reasoning research between Kohlberg's (1971) interview and Rest's (1976) Defining Issues Tests. Rest (1976) concluded that the recognition test is measuring something other than Kohlberg's model of moral reasoning. It remains to be seen whether the assessment techniques of the reflective judgment model will follow a similar metamorphosis.

Finally, the existing RJI data needs to be reanalyzed to further our understanding of the psychometric properties of the existing instrument. Brabeck (1980) has suggested that the psychometric properties of the individual dilemmas be investigated further, as it appears that they may be eliciting various stage level responses. She has also called for the investigation of alternative scoring methods.

The RJI has been invaluable as an exploratory instrument. As Kitchener notes in her chapter in the present volume, what is needed in assessing developmental change is not a global or epochal assessment but a finer-grained assessment. The next step in intellectual developmental stage assessment is to look at specific cognitive skills that are part of a given stage (Fischer, Hand, and Russell, forthcoming; Mines, 1980) and to develop instruments that can assess epistemological perspectives.

## **Moral Development**

### ***The Defining Issues Test (DIT)***

The DIT is designed to assess moral development from a cognitive stage perspective. The DIT has one of the more substantial data bases of the current developmental assessment techniques. Rest (1979) has an ongoing instrument refinement project at the University of Min-

nesota. He requests in the DIT manual that individuals who use the DIT send him copies of the data to be added to the data file.

**Purpose and Nature of the Instrument.** The DIT is a recognition test rather than a production test. It therefore produces higher stage levels than an interview or sentence completion format. The DIT consists of a long form (six stories) and a short form (three stories). The DIT contains an instruction page and three or six stories that reflect various moral dilemmas. The subject is asked to mark his or answers directly on the test, although machine scorable answer sheets can be developed

The DIT can be administered to students in ninth grade or above. The long form can be administered in a large group setting, and fifty minutes is usually sufficient for completing it, with most people completing it in thirty to forty minutes. The short form takes fifteen to thirty minutes to complete. The DIT is an objectively scored instrument and can be hand scored by clerical help or by a computer-scoring program provided by Rest (1979). If these two options are not feasible, a scoring service is offered by the Minnesota Moral Research Project that is relatively inexpensive (contact Rest for current prices).

The scoring systems provide a profile of the percentage of each stage level marked by the subject, plus the *P* score, which is the percentages of principled responses, an empirical weighted scored called the *D* score, *A* scores, which are an antiestablishment rating, a reliability check called the *M* score, and a consistency check. The *P* score and the *D* are the most often used for research purposes. Rest (1979) also recommends use of a stage profile approach. A positive feature of the DIT is its use of the *M* score to identify subjects who mark items on the basis of the "pretentiousness rather than the meaning" (Rest, 1979, p. 52). The consistency check is a monitor on the usability of the subject's responses. This gives an indication of the seriousness with which the subject approached the test.

**Reliability.** The DIT manual (Rest, 1979) reports test-retest reliabilities across several studies as generally in the high .70s or .80s and Cronbach's alpha internal consistency indices in the high .70s. The reliabilities for the specific stage scores are more moderate in the .50s and .60s. Rest strongly suggests that caution needs to be used in interpreting stage scores because of lower reliabilities. As the DIT is an objective instrument with objective scoring methods, interjudge agreement, reliability, or interviewer consistency are not of concern. Alternative forms of the DIT do not exist yet.

**Validity.** The DIT appears to have adequate criterion validity. Rest (1979) reports that, in six studies, significant differences in scores

have been found across age and education levels in the theoretically predicted direction. Several longitudinal studies have been completed that report significant upward stage changes over four and six years. Furthermore, this upward stage movement cannot be accounted for by generational or cohort specific effects. The convergent-divergent validity data indicates correlations up to the .60s and .70s with Kohlberg's (1971) various test versions. The DIT has correlations of .20 to .50 with measures of cognitive development and intelligence (Rest, 1979). It is usually nonsignificant or inconsistent with personality, and attitude measures as well as demographic variables such as sex, socioeconomic class, and political party.

Rest (1979) also provides some unique validation data not available on most of the developmental instruments. In a review of sixteen experimental enhancement studies, none of the control groups and only two of the experimental groups showed any gains in moral development. In all cases, the movement shown was less than in the longitudinal studies (Rest, 1979). This suggests that the DIT is measuring underlying conceptual organization and not just a specialized content. A second approach was to have subjects "fake good" and "fake bad." Subjects were able to "fake bad" and lower their scores but were not able to increase their scores by "faking good." The last approach reviewed implementing scaled techniques. Rest (1979) reports a study by Davison, Robbins, and Swanson (1978) in which the average scale values of items, when grouped according to their theoretical stage, were in the predicted order. Stage two had the lowest value, stage three the next, and so on up the model.

*Summary and Evaluation.* Rest (1979) provides an impressive array of data on the reliability and validity of the DIT. The extent and variety of studies on the DIT suggest that the DIT is a solid measure of moral reasoning. The ease of administering and scoring makes the DIT an attractive instrument for student services practitioners who are interested in this aspect of cognitive development. As with all of the cognitive stage models, change is slow and single programs probably will not affect stage change as measured by the DIT. The DIT provides a comprehensive scoring procedure that reflects the complexity of the stages. The major limitation of the DIT (which is true for all the cognitive stage instruments) is the inability to measure the fine-grained aspects of the stages, such as stage specific cognitive skill change or epistemological and metaphysical changes related to moral reasoning. The development of assessment techniques that reflect these stage specific changes will be of greater utility for evaluating social-cognitive stage change.



## Ego Development

### *Sentence Completion Test*

The Sentence Completion Test (SC) is used to assess ego development level as described by Loevinger's (1976) model. This is the third area of social-cognitive stage assessment that will be reviewed. Other models, such as Marcia's (1966), may also be of value to student services practitioners but are not reviewed here due to space limitations.

***Purpose and Nature of the Instrument.*** Loevinger and Wessler (1970) describe ego development as the master trait. All other forms of development, such as intellectual, moral, the self-system, interpersonal relationships, character development, and others are subsumed under the concept of ego development. One of the basic assumptions of the SC is that there are "coherent meanings in experience. . . . [This] is the essence of the ego or ego functioning, rather than one among many equally important ego functions. The ego maintains its stability, its identity, and its coherence by selectively gating out observations inconsistent with its current state. (One man's coherence, however, is another man's gibberish" (Loevinger and Wessler, 1970, p. 8).

The SC is a production test that attempts to identify qualitative differences in ego level. The SC consists of thirty-six incomplete sentence stems. The subject is asked to complete the sentences in any way he or she wants and takes fifteen to thirty minutes to complete the test. A number of different forms for girls, women, boys, and men are available in Loevinger and Wessler (1970). Thus the cost is minimal, involving institutional reproduction costs. The scoring of the SC requires two trained raters. Loevinger and Wessler (1970) provide a self-training guide, or one can be trained at Washington University in St. Louis. The SC can be administered in large groups. The age range is from eleven through adulthood. The original test development samples were all female. A scoring manual for males has not been published yet. Loevinger and Wessler (1970) suggest adapting the rules for females until the manual for males is published. The basic scoring system uses ogive rules or cutting points for determining stage level. The ogive rules take stage complexity into account but end up assigning a single stage score to the subject.

***Norms.*** The original sample was a cross-section of females ranging in age from eleven to fifty-plus, from a variety of races, marital statuses, and educational and socioeconomic statuses. The norms for males have not been published. Subgroup norms are not available.

**Reliability.** Loevinger and Wessler report interrater agreement by pairs of raters that range from 60-86, 63-91, and 65-94 percent across three pairs of raters. The medians for the percentage agreement across the rater pairs were 77, 78, and 81. The interrater correlations for the three rater pairs were .49-.88, .53-.93, and .56-.96. The median reliabilities were .75, .76, and .76, respectively. It is important to note that these reliabilities and agreements were obtained between raters that had little or no background in psychology. The correlations of four self-trained raters with a composite-trained rater were .60-.89, .57-.98, and .44-.89 across the thirty-six items. The median correlations were .78, .79, .85, and .76. The Cronbach alpha coefficients were .91 for a composite-trained rater over 543 subjects and .92, .90, and .88 for the composite-trained rater and two self-trained raters on a subsample of 100 subjects.

Loevinger and Wessler (1970) report extensive information on the reliability and percent agreement of raters on the total protocol ratings. The percent agreements across two samples were .50-.72, and .53-.80 with medians of .61 and .72 for ten pairs of raters. The percent agreements within half a stage level were 88-97 and 91-100, with a median of 94 for both samples. The interrater correlations ranged from .78 to .93, with a median of .85. A final indice of overall rater reliability is the correlations of trained and self-trained raters with the criterion total protocol rating and with the composite rater ogive score. The correlations across five trained raters for the criterion total protocol rating ranged from .90 to .96, and for the composite rater ogive score the range was .86 and .92, with the criterion total protocol rating .83 and .93 with the composite rater ogive score.

The SC has a high degree of interrater reliability and agreement. The internal consistency of the SC is high, although the alphas across subsamples were not reported. These alphas would be expected to be lower. One of the outstanding aspects of Loevinger and Wessler's (1970) interrater agreement and reliability data is the performance of the self-trained raters. This reflects the extent to which the authors have gone in refining the scoring manual and the advantage of objective criteria for assigning levels to a protocol.

**Validity.** The validation of a developmental stage measure is a multifaceted, ongoing endeavor, as has been noted. The SC appears to measure a unitary dimension structure. Factor analysis results indicate that the first factor correlates at .999 and with the sum of the item ratings (Loevinger and Wessler, 1970).

The construct validity information indicates that the SC total protocol ratings correlate .58 and .61 with an open-ended interview. Ini-

tial cross-sectional studies suggest progressive age differences across the ego levels. The SC score correlates .45 with IQ scores for black sixth grade boys and .47 for the girls.

**Summary and Evaluation.** The SC has good interrater reliability and internal consistency. The initial cross-sectional validity data are promising. For the student services practitioner, the SC is easy to obtain and administer. The low cost of the SC makes it attractive. It is somewhat time consuming to learn to rate the SC.

The major drawback in using the SC comes not from the instrument but from the nature of the developmental phenomena it assesses. Change in ego development is a slow process. In one study, a full stage change had not occurred in one year (King, Mines, and Barratt, 1979). Thus the construct may not be too useful for student services programming purposes.

The scoring rules for the SC are an important improvement over the use of the mode, mean, or highest stage approaches. Ogive rules do not represent the stage complexity that probably exists in ego development. Refining this aspect of the rules will be an important advancement.

Loevinger (1976) calls ego development the master trait, but all of the threads of ego development do not represent the whole. It will be important for the student services practitioner to separate out some of the threads in order to do any program planning with this model and to develop specific assessment techniques for those subsets of ego development.

### **Developmental Task Models**

Developmental task models represent another approach to the conceptualization of adult development. Developmental tasks are culturally specific events that occur at approximately the same time in the life of a given age cohort. A task must be successfully resolved for a person to develop the experiential foundation to resolve later developmental tasks, although, to date, there is little or no evidence of a longitudinal nature to support this assumption.

Developmental task models present a complex measurement picture. The models present a specific goal that young adults must attain (for example, freeing interpersonal relationships). The attainment of this goal requires multilevel changes. For example, in Chickering's (1969) vector of freeing of interpersonal relationships in order to move from dependence to independence to interdependence as well as increasing one's tolerance for diversity, it is reasonable to assume the indi-

vidual must experience a shift in cognitive complexity. Cognitive complexity as viewed by Perry (1970) or Kitchener and King (1981) is necessary if we are to move from banal stereotypic views of others to a sophisticated appreciation for individuals from diverse backgrounds. Cognitive stage change is implied in being able to process the interrelations by assessing the compromises and benefits necessary to relate interdependently. Also implicit in the resolution of this task are changes in ego and moral development levels. Ego development changes are necessary for the self-awareness necessary to act interdependently. Moral development changes reflect the implicit social contracts of dependent relationships versus interdependent relationships. The assessment of cognitive stage change and/or attitudinal change is one aspect of assessing a developmental task.

Developmental tasks can also be considered from a skills perspective. Each task has certain skills that the individual must possess in order to resolve the task. For example, the freeing of an interpersonal relationships vector requires that the individual possess certain communications skills to be able to function interdependently. Assertiveness skills, conflict mediation skills, and basic communications skills, such as those mentioned by Egan (1982), are skill categories that would be helpful in living interdependently. However, we do not have any data regarding specific skills required for task resolution. As with the cognitive stage assessment problems, the identification and assessment of task-related skills may provide us with a more refined understanding of the components of developmental task resolution.

The third aspect of the developmental tasks is behavior. Perhaps we should assess the behaviors related to task resolution. The assumption is that if one exhibits task-related behavior, one has resolved the task. This perspective has some appeal, in that it eliminates the problems of assessing cognitive stage, skills, and attitudes and goes directly to the behavioral manifestations. This approach is probably sufficient if the behaviors can be identified and one is only concerned with a yes or no diagnosis of task resolution. Unfortunately, the resolution of a developmental task is a process, not an event. The process occurs over time and, as noted previously, involves cognitive and skill changes as well as behavioral changes. Assessing only behavior tells the student services practitioner nothing about the why of an individual's progress in task resolution or programming needs. Behavioral assessment does not provide us with any clues as to how or even where to intervene. The limits of using a behavioral assessment approach to task resolution must be kept in mind.

The best approach to developmental task assessment would

integrate the cognitive stage, skills, and behavioral dimensions. The second best alternative would be to identify specific skills related to task resolution as well as behavioral manifestations of the task. The least desirable approach would be to assess only one dimension or even one aspect of the dimensions. We do not have the research base at this time to know what the relationship of cognitive complexity or attitudinal changes is to behavioral indicators of task resolution. In a related area, moral development research indicates that moral stage level is not a very good predictor of behavior (Rest, 1976).

There has been relatively little developmental task instrument development. Chickering's (1969) vectors of young adult development had received consistent attention from student services practitioners and researchers. This section will review the instrumentation designed to assess Chickering's vectors.

Chickering's model consists of seven vectors on developmental tasks that college age young adults need to resolve. The seven vectors are: achieving competence, managing emotions, becoming autonomous, establishing identity, freeing interpersonal relationships, clarifying purposes, and developing integrity. (See Chickering, 1969, for a complete description). The major instrument development (which focuses on behavioral aspects) has been done by Winston, Miller, and Prince (1979) at the University of Georgia. A second group of instruments focusing on attitudinal aspects of freeing of interpersonal relationships, clarifying purpose, and establishing identity were developed at the University of Iowa by Mines (1978), Barratt (1978), and Erwin and Delworth (1978), respectively. These instruments are still preliminary, although cross-sectional and convergent validity data (Rodgers, 1982) and four-year longitudinal change data (Hood, 1982) have been reported on the Iowa instruments. As the Iowa instruments are in the early instrument development phases, they will not be reviewed in this chapter. The Student Development Task Inventory-2 (Winston, Miller, and Prince, 1979) is reviewed in the next section, as it has received the most development to date.

### ***Student Development Task Inventory-2 (SDTI-2)***

***Purpose and Nature of the Instrument.*** The SDTI-2 "represents a sample of behaviors which students can be expected to demonstrate when they have satisfactorily achieved certain developmental tasks" (Winston, Miller, and Prince, 1979, p. 6). The SDTI-2 consists of three scales that assess the developmental tasks of developing autonomy, developing purpose, and freeing of interpersonal relationships. The

three major task areas are each divided into three subtasks. The SDTI-2 consists of 140 items that are marked true or false. It takes twenty to thirty minutes to complete. The answer sheet consists of two parts, an original and a carbon copy for the student. The SDTI-2 is designed to assess the development of individual college students from seventeen to twenty-three years old. It is intended to be used to assess behaviors related to task resolution or for stimulating discussion around personal growth issues. The SDTI-2 can be administered individually or in groups. The cost for a sample packet is \$7.00. The answer sheets and booklets cost 35¢ and 75¢ each.

**Norms.** The authors specifically state that the SDTI-2 is to be used for individual students and there is no need to establish norms or to attempt interpretations by making references to the performances of groups of students. The SDTI-2 was developed in a systematic manner through four phases. The SDTI-2 was refined on a sample ( $N = 497$ ) of college students at twenty colleges and universities across the country. The SDTI-2 samples have consisted of eighteen to twenty-two year old college students, of both sexes, white and black racial backgrounds, and freshman through senior class standing.

**Reliability.** The SDTI-2 manual reports test-retest and internal consistency reliability information. The two-week test-retest correlations for the total inventory were .90. The range of test-retest reliability coefficients for the scales was from .85 to .93, with the majority clustering around .90. The Cronbach's alpha internal consistency coefficients for the total inventory and scales were: total inventory, .90; developing autonomy, .78; developing purpose, .85; and developing mature interpersonal relationships, .73. The SDTI-2 subscales' alpha coefficients ranged from .45 to .78, with four of the nine having alphas at .51 or below. The subscale reliabilities are low, and they should be interpreted with great caution.

**Validity.** The SDTI-2 has contrasted group and concurrent validity data. Four contrasted groups were identified by residence hall staff members: active daters, nondaters, joiners, and social isolates. These groups were used to validate the subscales of the developing mature interpersonal relationships scale. Joiners and active daters scored higher on the scale score than did the nondaters and social isolates. The joiners also scored significantly higher than the social isolates on the mature relationships with peers subscale but not on the other two subscales. The active daters scored significantly higher than the nondaters on the intimate relationships with the opposite sex subscale but not on the other two subscales. Students who had not developed appropriate skills in relating to peers or were not dating were identified by the scales.

The concurrent validity studies correlated the Study Habits, Family Independence, and Peer Independence subscales of the College Student Questionnaire (CSQ) and the Adult Form I of the Career Development Inventory with SDTI-2. The Study Habits Scale correlated significantly but moderately with the Developing Autonomy scale, Instrumental Autonomy subscale, Developing Purpose scale, Appropriate Educational Plans subscale and the Nature Lifestyle Plans subscale. Family Independence correlated significantly but in the low moderate range with the Emotional Autonomy and Mature Career Plans subscales. Finally, the Peer Independence scale had a significant low moderate correlation with the Emotional Autonomy subscale. These correlations were all in the theoretically predicted direction, thus giving initial support to those SDTI-2 scales.

The Crystallization, Specification, and Implementation scales of the Career Development Inventory all had significant low moderate correlations with the Developing Purpose scale. The Implementation scale also correlated with the Developing Autonomy scale.

Significant age differences were only found on the Appropriate Educational Plans and the Mature Career Plans subscales. The only sex differences were found on the Developing Mature Interpersonal Relationships scale and the Tolerance subscale. Females scored higher than males on both scales. There were no differences between blacks and whites on any of the scales. The widest range of differences were found according to class standing. Significant differences were found on the Developing Autonomy and Developing Purpose scales as well as the Instrumental Autonomy, Interdependence, Appropriate Educational Plans, and Mature Career Plans subscales. The authors suggest these findings highlight the impact the college environment has upon an individual's development.

**Summary and Evaluation.** The SDTI-2 is one of the first major instrument development efforts designed to assess young adult task development. It approaches the assessment problem by focusing on behaviors that should be task related. The instrument has good internal consistency, and the initial validity studies are promising.

The question of how to assess developmental tasks becomes central in evaluating the SDTI-2. As noted earlier, the attainment of a developmental task involves cognitive complexity and attitudinal as well as behavioral changes. The SDTI-2 only uses behavioral indicators; thus we lose a significant aspect of the developmental change. The second area that needs further refinement is the scale development. The authors discuss scale and subscale differences, yet the scales and subscales are not orthogonal. This raises the question of how well the SDTI-2

discriminates among the various groups. At this point only the scales should be used to make inferences. The scoring system is one of the major liabilities of the SDTI-2. The behaviors are summed to give a scale score. This would be fine if the scales were Guttman-type scales; however, they are not. The scales are treated like modified Likert-type scales. The reason this distinction is important is that statistically significant differences may be found but the differences may not be developmentally significant. For example, class standing differences were found on the Developing Purpose scale with the means as follows: freshmen, 27.92; sophomores, 29.91; juniors, 31.07; and seniors, 34.03. If these were Guttman scales, we would be able to infer which behaviors the upperclassmen had mastered that the underclassmen had not. As the scoring system now exists, we cannot infer anything other than that upperclassmen report more developmentally appropriate behaviors. We do not know at this point whether or not the task has been resolved. These three areas set the stage (so to speak) for the next phase of developmental task instrument development.

### **Conclusions and Future Directions**

The current state of the art in young adult developmental stage and task assessment is not refined to the extent that student services practitioners can easily or meaningfully use the techniques. The field is posed on the brink of a new generation of assessment techniques. The areas for refinement and the specifics for refinement of the new generation of techniques are discussed in depth in Hanson's chapter. These techniques will have to incorporate complex stage scoring systems. Multilevel assessment for developmental tasks is necessary. Some of the more innovative assessment ideas have evolved from Fischer, Hand, and Russell's (forthcoming) work. They call for the implementation of strong scalogram analysis and the use of practice manipulations on developmental skills/tasks to determine maximum developmental performances. The new generation of instruments will incorporate specific skills related to a given stage as opposed to global descriptors. Finally, the new generation of instruments will be cost-effective and have utility for the student services practitioner who is faced with the challenge of developmental programming and evaluation.

### **References**

- Allen, K. Unpublished master's thesis, University of Maryland-College Park, 1982.
- Allgire, N. E. "Developmental Instruction and Information Responsiveness: Their Effects on Cognitive Development and Empathy." Unpublished doctoral dissertation, University of Maryland, 1977.



- Barratt, W. R. "Developing Purposes Inventory." *Technical Report*. Iowa City: College of Education, University of Iowa, 1978.
- Bogar, C. "The Relationship Between Attitudes Toward Women and Cognitive Complexity in Undergraduate College Students." Unpublished thesis, University of Maryland-College Park, 1981.
- Brabeck, M. "The Relationship Between Critical Thinking Skills and the Development of Reflective Judgment." Unpublished doctoral dissertation, University of Cincinnati, 1980.
- Chickering, A. W. *Education and Identity*. San Francisco: Jossey-Bass, 1969.
- Clement, L., Agar, J., Sherman, M., and Sobol, J. "Paraprofessionals—Development of Our Systems and Our Human Resources: A Model for Training." Paper presented at the University of Maryland Student Affairs Conference, University of Maryland-College Park, March, 1977.
- Clinchy, B. "A Longitudinal Study of the Cognitive Development of Women." Paper presented at the Perry Conference, St. Paul, Minnesota, June, 1981.
- Davison, M. L., Robbins, S., and Swanson, D. "Stage Structure in Objective Moral Judgments." *Developmental Psychology*, 1978, 14, 137-146.
- Egan, G. *The Skilled Helper*. Monterey, Calif.: Brooks/Cole, 1982.
- Erwin, T. D., and Delworth, U. "Erwin-Delworth Identity Scale." *Technical Report*. Iowa City: College of Education, University of Iowa, 1978.
- Fischer, K. W., Hand, H. H., and Russell, S. L. "The Development of Abstractions in Adolescence and Adulthood." In M. Commons (Ed.), *Beyond Formal Operations*. New York: Praeger, forthcoming.
- Hood, A. B. "Student Development on Three Vectors Over Four Years." Paper presented at the meeting of the American College Personnel Association National Convention, Detroit, Michigan, March, 1982.
- King, P. M. "The Development of Reflective Judgment and Formal Operational Thinking in Adolescents and Young Adults." Unpublished doctoral dissertation, University of Minnesota, 1977.
- King, P. M. "William Perry's Theory of Intellectual and Ethical Development." In L. Knefelkamp, C. Widick, and C. Parker (Eds.), *New Directions for Student Services: Applying New Developmental Findings*, no. 4. San Francisco: Jossey-Bass, 1978.
- King, P. M., Mines, R. A., and Barratt, W. R. "Longitudinal Study of Developmental Changes During the Freshman Year." Paper presented at the American College Personnel Association National Convention, Los Angeles, March, 1979.
- King, P. M., Kitchener, K. S., Davison, M. L., Parker, C., and Wood, P. K. "The Justification of Beliefs in Young Adults: A Longitudinal Study." Unpublished manuscript, School of Education, University of Denver, 1982.
- Kitchener, K. S. "Intellectual Development in Late Adolescents and Young Adults: Reflective Judgment and Verbal Reasoning." Unpublished doctoral dissertation, University of Minnesota, 1977.
- Kitchener, K. S. "Does Development Occur in Reflective Judgment in the Young Adult Years?" Paper presented at the meeting of the American College Personnel Association, Cincinnati, Ohio, April, 1981.
- Kitchener, K. S., and King, P. M. "Reflective Judgment: Concepts of Justification and Their Relationship to Age and Education." *Journal of Applied Developmental Psychology*, 1981, 2, 89-116.
- Knefelkamp, L. "Developmental Instruction: Fostering Intellectual and Personal Growth in College Students." Unpublished doctoral dissertation, University of Minnesota, 1974.
- Kohlberg, L. "Stages of Moral Development." In C. M. Beck, B. S. Crittenden, and E. V. Sullivan (Eds.), *Moral Education*. Toronto: University of Toronto Press, 1971.

- Lawson, J. "The Relationship Between Graduate Education and the Development of Reflective Judgment: A Function of Age or Educational Experience." Unpublished doctoral dissertation, University of Minnesota, 1980.
- Loevinger, J. *Ego Development: Conceptions and Theories*. San Francisco: Jossey-Bass, 1976.
- Loevinger, J., and Wessler, R. *Measuring Ego Development*. Vol. I: *Construction and Use of a Sentence Completion Test*. San Francisco: Jossey-Bass, 1970.
- Marcia, J. E. "Development and Validation of Ego Identity Status." *Journal of Personality and Social Psychology*, 1966, 3, 551-558.
- Mason, K. E. "The Effects of Developmental Instruction on the Development of Cognitive Complexity, Locus of Control, and Empathy in Beginning Counseling Graduate Students." Unpublished master's thesis, University of Maryland, 1978.
- Mentkowski, M. "Using the Perry Scheme of Intellectual and Ethical Development as a College Outcomes Measure." Paper presented at the Perry Conference, St. Paul, Minnesota, June, 1981.
- Meyer, P. "Intellectual Development: An Analysis of Religious Content." *Counseling Psychologist*, 1977, 6 (4), 47-50.
- Mines, R. A. "The Mines-Jensen Interpersonal Relationship Inventory." Paper presented at the American College Personnel Association National Convention, Detroit, March, 1978.
- Mines, R. A. "An Investigation of the Developmental Levels of Reflective Judgment and Associated Critical Thinking Skills in Young Adults." Unpublished doctoral dissertation, University of Iowa, 1980.
- Mines, R. A. "Psychometric Aspects of the Reflective Judgment Interview Procedures." Paper presented at the meeting of the American College Personnel Association, Cincinnati, Ohio, May, 1981.
- Mines, R. A., Gressard, C. F., and Daniels, H. "Evaluation in Student Services: A Metamodel." *Journal of College Student Personnel*, 1982, 23, 195-201.
- Moore, W. S. "The Measure of Intellectual Development: A Brief Review." Unpublished paper, Center for Applications of Developmental Instruction, University of Maryland-College Park, 1982.
- Perry, W. G., Jr. *Forms of Intellectual and Ethical Development in the College Years*. New York: Holt, Rinehart and Winston, 1970.
- Rest, J. R. "New Approaches in the Assessment of Moral Development." In T. Lickona (Ed.), *Moral Development and Behavior*. New York: Holt, Rinehart and Winston, 1976.
- Rest, J. R. *Revised Manual for the Defining Issues Test*. Minneapolis: Moral Research Projects, University of Minnesota, 1979.
- Rodgers, R. "Chickering's Vectors and Their Development in Women." Paper presented at the meeting of the American College Personnel Association, Detroit, Michigan, March, 1982.
- Schroder, H., Driver, M., and Streufert, S. *Human Information Processing*. New York: Holt, Rinehart and Winston, 1967.
- Shoff, S. P. "The Significance of Age, Sex, and Type of Education on the Development of Reasoning in Adults." Unpublished doctoral dissertation, University of Utah, 1979.
- Slepitz, R. L. "The Validation of a Stage Model of Career Counseling." Unpublished master's thesis, University of Maryland, 1976.
- Stephenson, B. W., and Hunt, C. "Intellectual and Ethical Development: A Dualistic Curriculum and Intervention for College Students." *Counseling Psychologist*, 1977, 6 (4), 39-42.
- Strange, C. C. "Intellectual Development, Motive for Education, and Learning Styles During the College Years: A Comparison of Adult and Traditional Age College Students." Unpublished doctoral dissertation, University of Iowa, 1978.
- Touchton, J. G., Wertheimer, L. C., Cornfeld, J. L., and Harrison, K. H. "Career Planning and Decision Making: A Developmental Approach to the Classroom." *Counseling Psychologist*, 1977, 6 (4), 42-47.

- Viesar, K. G. "The Use of Cognitive Complexity, Characterological Type, Empathy, and Leaderless Group Discussion Measures as Evaluative Indices in the Selection of Orientation Advisors." Unpublished master's thesis, University of Maryland, 1978.
- Welfel, E. R. "The Development of Reflective Judgment: Its Relationship to Year in College, Academic Major, and Satisfaction with Major Among College Students." Unpublished doctoral dissertation, University of Minnesota, 1979.
- Welfel, E. R. "The Reflective Judgment Scores of College Students: A Closer Look." Paper presented at the meeting of the American College Personnel Association National Convention, Cincinnati, Ohio, April, 1981.
- Wertheimer, L. C. "A New Model and Measure for Career Counseling: Incorporating Both Content and Processing Aspects of Career Concerns." Unpublished master's thesis, University of Maryland, 1976.
- Wertheimer, L. C. "Relations Among Developmental Dimensions in Jane Loevinger's Model of Ego Development." Unpublished doctoral dissertation, University of Maryland, 1980.
- Widick, C. "An Evaluation of Developmental Instruction in a University Setting." Unpublished doctoral dissertation, University of Minnesota, 1975.
- Winston, R. B., Jr., Miller, T. K., and Prince, J. S. *Assessing Student Development*. Athens, Ga.: Student Development Associates, 1979.
- Wirtz, P. G., and Magrath, D. S. "Creating Quality Programs." In M. J. Barr and L. A. Keating (Eds.), *New Directions for Student Services: Establishing Effective Programs*, no. 7. San Francisco: Jossey-Bass, 1979.
- Wood, P. K. "The Analysis of the Structural Relationship Between Two Tests of Critical Thinking and Reflective Judgment." Unpublished master's thesis, University of Iowa, 1980.

*Robert A. Mines is an assistant professor of counseling psychology, School of Education, and senior staff psychologist, Counseling Services, at the University of Denver. Previously he was the acting coordinator of the Iowa Student Development Project at the University of Iowa. He has been involved in consulting and research on developmental stage and task issues in young adults.*