

Construct Validity of the Sentence Completion Test of Ego Development

Jane Loevinger
Washington University

Evidence for the construct validity of the Sentence Completion Test of Ego Development (SCT), some of it previously unpublished, is reviewed. The substantive component is substantiated both by intuitive plausibility and by the fact that rating skill can be communicated through written instructions. Internal consistency is shown by alpha coefficients of about .85. Evidence for sequentiality is provided by studies showing cross-sectional gains with age during adolescence, similar profiles of item scores for adjacent total protocol ratings, substantial correlations over long time spans, gains proportional to retest intervals during adolescence, gains following theory-relevant interventions of several months' length, and better comprehension of lower than of higher stages. The SCT has substantial correlation with tests of related conceptions, such as Kohlberg's test of moral maturity, and with measures of ego level applied to other behavior samples. Correlations with isolated traits characteristic of specific stages range from negligible to positive. Studies of external validity in general yield positive results, though theory does not sharply define what results should be considered positive. At lower ego levels there is behavioral evidence confirming the syndromes; at higher levels differential evidence lies in attitudes and ideas. Overall, the test has adequate validity for research purposes when administered and scored with sufficient care, but not for individual diagnosis unsupported by other data. Evidence for construct validity is substantial, but falls short of clear proof of sequentiality.

This paper reviews studies bearing on the construct validity of the Sentence Completion Test of Ego Development (SCT), emphasizing the logic of construct validation and previously unpublished data. Studies covered in Hauser's (1976) review of the ego development concept and test will be referred to only briefly. The original discussions of construct validity (e.g., Cronbach & Meehl, 1955; Loevinger, 1957) ignored tests of developmental variables; therefore, the present review covers new ground methodologically.

The Construct of Ego Development

The term ego development is ambiguous, covering both the specific set of stages measured with the SCT and the broader domain of the developmental typologies, such as those of Kohlberg, Perry, Selman, and Blasi (Loevinger, 1976). There has been speculation on relations among the constructs within the domain, but only minimal research. Since the stages have been fully delineated elsewhere (Hauser, 1976; Loevinger, 1966, 1976; Loevinger & Wessler, 1970), they can be described briefly.

The earliest stage that can be judged from the test is the Impulsive Stage, normal in early childhood and not infrequent as high as the sixth grade. At this stage the child is egocentric, dependent, and demanding, and lives in a con-

ceptually oversimplified world. When problems are perceived, they are external and often located in a place; thus, the child can run away from them. Work is onerous, not an opportunity. Rules are either not perceived or are seen as the arbitrary exercise of authority.

The next stage is Self-Protective. Originally it was called Opportunistic, but that term was discarded as too pejorative, particularly for children passing through the phase in normal time—grade school or junior high years. The child in this stage tends to be less impulsive, more wary, and more self-sufficient than before, but rules are obeyed for his or her own short-term advantage. Other people are blamed when trouble occurs. Staying out of trouble and not getting caught are major preoccupations.

The next, or Conformist, stage is characterized by conformity to rules for their own sake. The person identifies himself or herself with the group. Inner life is perceived, but only in banal and conventional terms. Moralistic clichés are favored expressions.

The modal level for high school graduates is probably the transition from the Conformist to the Conscientious Stage. Rather than assuming, as the Conformist does, that “one size fits all,” the person here perceives personal problems and alternative possibilities in situations. Self is seen as somewhat apart from and differentiated from the group. Work is an opportunity.

At the Conscientious Stage, the person has a richly differentiated inner life and a sense of the many differences in traits among people. The person adheres to a set of beliefs that have been self-evaluated, rather than ones adopted by the group, aspires to achievement as measured by his or her own standards rather than worldly success, and feels responsible for others.

The person at the next stage, called Autonomous, has many of the characteristics of the previous stage. The aspiration for achievement is partially supplanted by a desire for mutuality and self-fulfillment. Inner conflict is accepted and coped with. Individual differences are not only perceived but cherished. Other

people’s autonomy is respected. There is a theoretical higher stage, called Integrated, resembling Maslow’s self-actualizing person. Table 1 summarizes the stages of ego development measured by the SCT.

Numbering the stages is inappropriate, as there are one or two prior stages, which are omitted because they have no representation in the SCT. Moreover, the construct is open to revision and expansion, in which case stage numbers would cause confusion. (For example, Kohlberg, 1973, is now saddled with the awkward concept of a “Stage 4 ½.”) Many authors have numbered these stages, but they have done so in inconsistent ways.

The alternative, naming the stages, also can lead to misunderstanding, as no human capacity springs into being full blown at one stage and disappears on passage to the next. For example, a response indicating or advocating conformity may or may not be classified at the Conformist Stage. Each stage is assumed to be characterized by a more or less coherent structure of thought, feelings, and perceptions. The Conformist Stage might be defined as the stage at which conformity, nonconformity, and anticonformity are central, though often unacknowledged, problems of existence. Such a definition lifts the conception out of the realm of simple conventionality and into what might be called *meta-conformity* (Loevinger, 1976). There is no easy way to reduce such a complex conception to objective indicators.

The Test and its Manual

The scoring manual for the SCT (Loevinger & Wessler, 1970; Loevinger, Wessler, & Redmore, 1970) includes item manuals for 36 specific sentence stems from the forms for women and girls, together with instructions for arriving at a total protocol score after rating all 36 item responses. A set of graduated exercises, appropriate for graduate or advanced undergraduate students, permits persons to train themselves in the technique.

Table 1
Stages of Ego Development Measured by the SCT

Stage	Code	Score ^a	Impulse Control		Interpersonal Mode		Conscious Preoccupations	
			Impulsive	Control	Egocentric, dependent	Mode	Bodily feelings	
Impulsive	I-2	0						
Self-Protective	I-2/Δ	1						
	Δ	2						
	Δ/3	3						
Conformist	I-3	4		Opportunistic		Manipulative, wary		"Trouble," control
	I-3/4	5		Respect for rules		Cooperative, loyal		Appearances, behavior
Conscientious	I-4	5		Exceptions allowable		Helpful, self-aware		Feelings, problems, adjustment
		6		Self-evaluated standards; self-critical		Intense, responsible		Motives, traits, achievements
	I-4/5	7		Tolerant		Mutual		Individuality, development, roles
Autonomous	I-5	8		Coping with conflict		Interdependent		Self-fulfillment, psychological causation
		10				Cherishing individuality		Identity

Note: ^a Some studies use a different origin, so comparison of item sum scores may require a translation.

The manual also gives variant forms: one for men, a revised form for women deleting a few nonproductive items, and forms for boys and for girls eliminating a few items referring to sex. New, closely comparable forms for boys and girls are now used (Loevinger, no date). Instructions are that the test should always consist of 36 items; thus, the stems are assumed roughly equivalent in psychometric value. Some investigators have used the method with forms containing only 12, 15, or 18 stems; reliability estimates for the 36-item version do not apply to abbreviated forms. Where items on the men's or boys' forms are identical to items for which there is a women's manual, good results have been obtained using it as a guide (see below).

A 12-item Japanese form has been used by Kusatsu (1977). The form for men has been used in Curaçao, translated into Papiamentu, Spanish, and Dutch by Lasker, Pinedo, and their associates (Lasker, 1978). Snarey and Blasi (1978) have used a Hebrew form. Vetter (1978) has used a German form. Limoges (1978) has translated the forms into French.

Tentative scoring manuals for stems that occur on the forms for men and boys have been worked out by several investigators (Holt, no date; Love, no date; Redmore, Loevinger, & Tamashiro, 1978). Lasker and his associates (1978) have an auxiliary manual, which is applicable mainly to men, covering the transition from the Self-Protective to the Conformist Stage, a level slighted in the original manual.

Evidently the method for constructing the scoring manual (Loevinger & Wessler, 1970, chap. 2) is communicable. That is not trivial, for revision of the scoring manual is the method for revising the details of the conception (Loevinger, 1979).

Construct Validity

Substantive Component

The substantive component of validity (Loevinger, 1957) covers what is called content validity, construed not as an alternative kind of

validity but as an aspect of construct validity. Content or substantive validity must be evaluated more or less subjectively (Ebel, 1956): Does the content of the test, including the task and scoring method, reasonably appear to index the putative trait?

The rationale for the SCT is that ego development is, or reflects, the person's frame of reference (Loevinger, 1966). Therefore, an unstructured test, permitting the respondent to supply his or her own frame of reference, is appropriate. At the same time, by providing 36 discrete answers and partially restricting the domain of the answers, the test is psychometrically simpler than other projective tests.

The central questions are whether the content of the responses classed at each level make coherent sense and whether the succession of stages is logically convincing. Since specific responses are listed in the scoring manual as classed at each level, it may appear that specific behaviors represent each level. However, there are often contradictory responses at a level. Thus, in response to "Raising a family . . .," answers such as "is good," "is bad," or "is not good" would all receive the rating Impulsive (Loevinger, Wessler, & Redmore, 1970, p. 2). In response to "My conscience bothers me if . . .," "I did something intentionally," and "I offend some one, unintentionally" are both classed at the Conscientious Stage (pp. 438–439). Thus, the method tries to reach and to gauge the structure of the response. The intuitive argument for the coherence of the construct (Loevinger, 1976; Loevinger & Wessler, 1970) is similar to other content validity presentations.

Studies of interrater reliability, in addition to their technical psychometric interest, testify to the communicability, hence the coherence, of the underlying construct. Major studies were conducted using a Sealed Sample, called that because none of the protocols were read by any staff member during construction of the pre-publication manual, for which it served as cross-validation (Loevinger & Wessler, 1970). It included 543 women and girls, diverse as to

source, age, and social class, randomly intermixed. Four experienced raters, two per item, independently rated all 543 responses to each item. On an 11-point scale, shown in Table 1, median complete agreement between raters per item was 77%. The median correlation between the two ratings per item was .75. In rating total protocols, five raters took part, and each protocol was scored by three of them. Each of the 10 pairs of raters scored between 143 and 200 protocols in common. The median correlation between raters for total protocol scores was .86. Median complete agreement was 61%; median agreement within a half stage, 94%. In a later sample, complete agreement on total protocol scores was 71%, and the other values were approximately the same as in the Sealed Sample.

In order that their results could be assumed to be valid for scorers trained in other research settings, Loevinger and Wessler trained some scorers using only the prepublication manual. (The published version differs only slightly as a result of feedback from these and related studies; hence it is presumably at least as good as, and probably better than, the prepublication version.) The "self-trained raters" were not permitted to discuss the task or to ask questions, nor were they given additional oral instructions before the experiment was over. The first two self-trained raters were advanced graduate students: One of them had had brief experience with an earlier form of the manual; the other had no relevant experience. The second pair of self-trained raters were a freshman and a sophomore with no knowledge of the theory and no relevant experience. By all criteria, these four raters were indistinguishable from the more experienced raters who had been working together for many months. Thus, the construct is coherent enough to be conveyed by available written materials to at least some people.

Evidence that a manual constructed for women and girls is as reliable when it is used for men and boys, and evidence that trained raters do about as well without an explicit item manual

as with one, are further evidence for the coherence of the construct, hence for substantive validity. The data from two studies of both high school and junior high school boys (Hoppe, 1972; Cox, 1973), involving 108 and 86 cases, respectively, are pertinent to these points. Two-thirds of the items were the same as those in the form for women or differed only in gender, so that the women's item manual could be used; one-third of the items were entirely new and had to be rated with only an intuitive understanding of the construct for guidance.

Five experienced raters, each paired with all of the others, were used in Hoppe's study; six raters, each paired with all of the others at least twice, were used in Cox's study. For 24 items rated with a more or less appropriate item manual, the median complete agreement was 72% and 65%, a small loss of accuracy as compared to the 77% of the Sealed Sample. For the 12 new items, median complete agreement was 56% and 54%. Median agreement within a half stage, the finest discrimination that can be made on the scale, was 89% and 84% for items rated with a manual, 90% and 84% for items rated without a manual. Thus, disregarding disagreements of a half stage, experienced raters do as well without an item manual as with one. (These data are reported in Redmore and Waldman, 1975, not in the original theses.) Needless to say, the manual is the chief device for conveying the tacit component of the conception; so these results hold only for trained raters.

Structural Component

The primary evidence for the structural component of validity is homogeneity or internal consistency. For the Sealed Sample, using the ratings of the "composite trained rater" (that is, the compromised item ratings of experienced raters, two of whom rated each item), coefficient alpha for the 36-item test was .91. A principal components analysis yielded an eigenvalue of 8.8 for the first principal component and a value of 1.2 for the second principal component. The sec-

ond component could not be interpreted intelligibly, and its eigenvalue was consistent with considering it a chance deviation from unity. Scaled scores computed for the first principal component correlated essentially 1.0 with the sum of item ratings and correlated .96 and .97 with the first-component scaled scores for each of the first two self-trained raters. On the basis of these and related data, it was concluded that the test measures a unitary dimension (Loevinger & Wessler, 1970).

The source of that unity, however, is methodological; a single conception guided the construction of all item manuals. Of course, using the total protocol rating to improve the item manuals guarantees that they will be as consonant as possible with the general conception; this finding is thus almost too obvious to mention. Nonetheless, various investigators have challenged it, seeking subsets of items that measure separate aspects of ego development.

Lambert (1972) had available SCTs and scores on Kohlberg's (1964) test of moral maturity from a diverse sample ranging in age from 11 to 60 years, all American-born whites. Subsamples included 35 11-year-old middle-class school children; 24 15-year-old middle-class school children; 15 middle-class men, aged 20 to 27 (Kohlberg's longitudinal samples); and 16 persons aged 26 to 60 in the psychology department of a psychiatric institution who had or were working for advanced degrees. Two remaining samples were culled from larger groups to overrepresent

high ego levels as measured by the SCT; five were chosen from 15 members of the social service department of the psychiatric institution, and 14 were chosen from 53 candidates for a master's degree in pastoral counseling at a Catholic university.

Lambert analyzed 22 SCT stems that were identical, or identical except for gender, on all forms of the test used in his study. For these 22 items, the correlations with the Kohlberg global score ranged from .34 to .65. (The correlations of the items with the total protocol ratings for the Sealed Sample of Loevinger and Wessler ranged from .35 to .60.) The five best items, using a stepwise multiple regression analysis, yielded a multiple correlation with the Kohlberg score of .82. Approximately the same results were obtained no matter which of the 22 items was arbitrarily chosen as the first step. From these and other analyses, Lambert concluded that there was no "moral" subset of SCT items.

Blasi (1971) tested 109 black sixth-grade boys and girls from an inner-city school. He separated the 36 items into the 13 or 14 that seemed to tap "responsibility" and the remainder. He gave the SCT before and after his experiment (described below), with about 5 to 7 months between testings. He computed the item sum score separately for responsibility items, both pretest and posttest. As Table 2 shows, he found no evidence for separable blocks of items. (These data are not reported in his thesis.)

Table 2
Correlations of Item Sum Scores on Responsibility
Items and Remaining Items
(Data of Blasi, 1971)

	Pretest Other	Posttest	
		Responsibility	Other
Pretest scores			
Responsibility items	.52	.50	.50
Other items		.48	.49
Posttest scores			
Responsibility items			.60

Sequentiality

For a variable or dimension to qualify as developmental, the stages must at least develop in sequence. Kohlberg (1969), in defending his conception of moral development, puts most of his validation chips on demonstrations of sequentiality. Sequentiality asserts that the stages always occur in the same order, that no stage can be skipped, and that progress is ordinarily irreversible. The alternative hypothesis is that the order of the stages is at times reversed and that the order is dictated by external factors, rather than by the internal logic of the developmental continuum itself. A rigorous test of sequentiality would require frequent testing (at least as frequent as the least time needed to make measurable progress), error-free measurement, and no effects of measurement per se. Those conditions are impossible to meet. In particular, frequent measurement is likely to be resented and hence to result in poor validity for retests. There are six different kinds of evidence for sequentiality of the SCT, all more or less supportive of the claim, but none are totally convincing. Table 3 summarizes the evidence for sequentiality in the SCT.

Age differences. The most obvious, but equivocal, evidence for sequentiality is the demonstration of stage progression with increasing age during childhood and adolescence. Informal evidence to that effect is the usual starting point for developmental stage conceptions. Since, however, many kinds of changes are occurring simultaneously, these observations cannot be critical for ego development as a separate domain.

Loevinger and Wessler (1970, p. 50) reported a study of Toronto school children, 100 girls and 91 boys, in four age groups: 9–10, 11–12, 13–14, and 17–18. Protocols were randomly intermixed for scoring, with indications of age and grade removed. Although the distributions of ego level for successive ages overlapped, the ogive curves showing the cumulative distributions did not cross. Thus, the entire distribution shifts upward with age.

Similar cross-sectional differences were found by Coor (1970) in a study of black inner-city children in St. Louis schools in Grades 7, 8, and 9; however, she found a slight, nonsignificant decrease in Grade 10 as compared to students in Grade 9. She tested 28 students in the seventh grade, 18 or 19 in the other grades, and about equal numbers of boys and girls. In a study of 107 boys in a private school in St. Louis county, Hoppe (1972) found a clear cross-sectional progression in ego level between Grades 8, 9, and 11.

Distribution of stage usage. A second kind of evidence for sequentiality comes from the greater resemblance in distribution of stage usage for adjacent, as compared to nonadjacent, stages. If every person's responses all fell at one level, this criterion would not be usable; however, all investigators agree that whatever the testing method used, people show variability in the level of behavior exhibited. Table 4 shows for the Sealed Sample the median frequency of item responses at each level for each level of total protocol score. (Only a random 20% of cases were used at the Conformist, Conscientious, and Conscientious-Conformist levels. All cases were used at the other levels. Because medians rather than means were used, the frequencies do not total 36, though in every case there were 36 ratings.)

Again, this criterion is somewhat tautological. The manual construction technique is based on the assumption that this relation will hold. In fact, partly because there is more use of Impulsive Stage responses in the Self-Protective Stage, it was differentiated from the Conformist Stage.

Longitudinal studies. The most important kinds of evidence for sequentiality come from longitudinal studies, and they are of two sorts: (1) correlation over a long period of time, say at least a year, and (2) demonstration that ego level either increases or stays constant over time. In longitudinal studies of the SCT cited here, test and retest were randomly intermixed and rated without knowledge of which was pretest and which posttest and without knowledge of which two tests belonged to the same subject.

Table 3
Evidence for Sequentiality

<i>Topic and Author</i>	<i>Sample</i>	<i>Conditions</i>	<i>Results</i>	<i>Raters</i>	<i>Sampling</i>
Cross-Sectional Age Differences					
Loevinger & Wessler	191 Toronto M & F; 10-18 years old		Increased by grade	2	R
Coor	83 black inner-city M & F; Grades 7, 8, 9, 10		Increased to Grade 10	2	R
Hoppe	107 prep school M; Grades 8, 9, 11		Increased by grade	2	R
Comparison of Item Distributions of Different Stages					
Loevinger & Wessler	196 F, 12 to 60 years old, from Sealed Sample		Adjacent stages more similar than nonadjacent stages	2	H, X
Retests in Longitudinal Studies					
Redmore & Loevinger	Several high school samples; 230 M & F	1½- to 6-year interval	Most samples, signif. gain & signif. <i>r</i> ; gain approx. proportional to interval	2	R
Loevinger	Several classes at technical institute; 115 M, 19 F	2- & 4-year interval	Most samples, signif. gain after 2 years; no further gain after 4 years, signif. <i>r</i>	2	R
Change Following Intervention					
Blasi	33 M & 37 F; black inner-city, Grade 6	2-week role-playing experiment; +1- & +2-stage training	No change	2	X
Sullivan	14 high school jrs. & srs. in experiment; 14 in each control group	9-month course	Signif. rise in experimental group only	?	R
Exum	24 jr. college freshmen in experiment; 24 in control group	9-month course	Signif. rise in experimental group only	?	R
Lasker	About 100 adult workers	Group experience, about 6 months; posttest, 9 months to 2 years after pretest	Signif. rise below I-3, not above	2	R
Asymmetry of Comprehension					
Redmore	Various; 146 M & 88 F	Faking "high" & "low"	Most can fake low; only slight rise for high	2	R
Blasi	(See above)	Role-playing	Many children unable to enact higher roles	2	X
Bonneville	86 white F; Grades 6, 8	Role-playing	No evidence of asymmetry	2	X

Notes. F means female; M means male.

Raters: 2 means more than one rater for every protocol.

Sampling: R means quasi-random within specified group; X means subjects chosen using SCT to maximize variability; H means heterogeneous combination of samples.

Table 4
Median Frequency of Item Ratings by Total Protocol Rating in Sealed Sample
(Data of Loevinger & Wessler, 1970)

Total Protocol Rating	N	Item Rating								
		I-2	Δ	$\Delta/3$	I-3	I-3/4	I-4	I-4/5	I-5	I-6
I-2	17	7	2	1	22	2	1			
I-2/ Δ	11	4	3	1	20	4	2			
Δ	11	2	7	1	17	5	3			
$\Delta/3$	22	2	3	1	22.5	4	3			
I-3*	30	1	1	0	23	7	4			
I-3/4*	29				17	10	7	1		
I-4*	26				11	8.5	12.5	2.5		
I-4/5	29				7	7	13	6	2	
I-5	14				6.5	6	12	5.5	4	
I-6	7				2	3	9	10	7	1

Note: * means random 20%

Before results of longitudinal studies can be discussed, however, a baseline of short-term reliability must be established. Although retest reliability is not directly germane to construct validity (Loevinger, 1957), in longitudinal studies retest effects are germane to interpreting the second test. Presumably, the longer the time between test and retest, the more the second test is like another first test (as classical reliability theory has unjustifiably assumed all retests are).

Redmore and Waldman (1975) report several small studies of short-term retest effects. There appears to be a tendency for scores to decline on retest after a week or two, averaging about half a stage. This conclusion may apply particularly to scores above the Conformist Stage, since the elaborations and qualifications that establish a response as being at high levels are what the bored subject is apt to omit on retest. A psychology class for whom the retest was introduced as a study of test reliability showed no decline. Thus, for most subjects under most testing circumstances, a retest within a short time cannot be assumed to reflect change in ego level so much as change in response to the test per se. For the one group that showed no average change, about 60% showed no change in score, with most of the remaining cases going up or down half a stage (that is, from a stage to the adjacent transitional level or vice versa).

Redmore and Loevinger (1979) report longitudinal results for eight adolescent samples, drawn from inner city black schools of relatively low socioeconomic status, a private boys' preparatory school, and two suburban schools of intermediate socioeconomic status. The first test was given variously between Grades 6 and 11; retest intervals ranged from 1½ to 6 years; some samples were retested twice. All but one sample were retested in Grade 12. Various selective factors operated with respect both to the original samples and to those students available for retest, to whom the report is restricted. The value of the study is two-fold: (1) each retested subject served in effect as his or her own control and (2) the samples covered the entire socioeconomic spectrum of the St. Louis metropolitan area.

All samples showed a mean rise in ego level; for six of the eight samples the rise was statistically significant. The mean gain ranged from almost a stage and a half, between Grades 6 and 12 for an inner-city sample, to about a sixth of a stage, between Grade 11 and the freshman year of college for a preparatory school sample. All samples showed a positive correlation between test and retest; of the 14 correlations, 10 were significantly different from zero. Retest correlations ranged from .14 to .92, averaging around .5. Time between test and retest did not seem to account for variation in the size of the coeffi-

cient; variability of the sample appeared to be more important. Ego gains tended to level off towards the end of high school and the beginning of college.

Loevinger (1978) reported longitudinal data for a series of samples at Worcester Polytechnic Institute. All samples showed a mean gain in college, averaging about half a stage; however, there was no evidence of a gain between the end of the sophomore and the end of the senior year. Most samples showed a positive correlation between test and retest; half those correlations were significantly different from zero.

These substantial retest correlations over long time periods support the construct validity of the SCT and, more specifically, sequentiality. However, theory does not predict exactly how high the correlation should be over an interval in which there is substantial growth. With a single exception, every sample in which there was not a significant gain over a given period did show a significant test-retest correlation. Some samples showed both a significant gain and a significant test-retest correlation. Taking into account the results of the reliability studies, the longitudinal studies are consistent with the hypothesis of sequentiality, though not decisive evidence for it.

Change following intervention. Somewhat different from longitudinal studies are studies of change following a theory-relevant intervention. Blasi's (1971, 1976) experiment was similar to Turiel's (1966) study of moral reasoning, but designed independently. Blasi reasoned that if sequentiality held, children trained to reason at the next higher stage ("plus one") would be more affected by their training than those exposed to reasoning two stages higher than their current functioning ("plus two"). He pretested sixth-grade children in inner-city schools, then constituted multiple groups of between five and eight children at the Impulsive, Self-Protective, and Conformist levels, and a single group at the Conscientious-Conformist level.

Blasi devised stage-appropriate moral dilemmas, which the children in each group enacted, followed by inquiry and discussion. The pro-

ceedings in each group were recorded independently by Blasi and one of two other observers; one of his observers was trained in ego development theory, but the other one was not. Each group met for 10 sessions, 1 class period every school day for 2 weeks. The behavior of each child was recorded after each session by each of the observers. Subsequently, they were post-tested with the SCT, as was a control group at each level given no training. All tests were then rescored (by new raters), with pretests and post-tests randomly intermixed.

No statistically significant changes were shown, so that the hypothesis that "plus one" training is more effective than "plus two" could not be tested. It is doubtful that 2 weeks is long enough to give a fair test to such training techniques, given the theoretical stability of ego level. Although this study failed as an intervention study, other parts of the data are supportive of construct validity, particularly the observations of the children's behavior (see below).

A series of studies by Mosher and Sprinthall and their collaborators (1971) are better than Blasi's in taking place over a longer time, but many lack the precision of design and the articulation between theory and experimental intervention. Typically, students take part in a course that lasts one semester or one quarter; the course includes both a didactic and a practicum aspect, some substantive content, and some attempts to give them insight into their problems. Mosher and Sprinthall call such curricula "deliberate psychological education." The curricula have included peer counseling courses in high school, high school children assisting in teaching elementary school children, a course in improvisational drama, and a course in child development that involved a nursery school practicum.

Subsequent studies have included a woman's studies program at the college level (Erickson, 1974, 1975). Some of the studies showed significant gains, some did not, but neither result is decisive. Absence of carefully matched control groups, coupled with the fact that most of the

studies took place in an age range where growth is normally expected, vitiate the value of those showing gains. Since retests normally decline when given a short time after the original test, studies showing no gain or even decline are also ambiguous. (In these studies, the term was 12 or 16 weeks; and though the retest interval was not always specified, it seemed generally to be from some time after the first class to some time before the last class.)

Some studies showed significant loss in the (quasi) control group, compared with no change in the experimental group, but it is difficult to interpret such results. For example, Hurt (1974) found that half of his subjects with a posttest decrease were judged independently to be showing either test fatigue or hostility. Most of those showing test fatigue or hostility were lower on SCT posttest than on pretest; none were higher on posttest. In addition, the curricula were not intimately tied to theory; their descriptions at junior high school, high school, and college reveal similar elements.

One of the best intervention studies is that of Sullivan (1975), whose high school course lasted a full school year and included four phases: (1) moral discussions, (2) counseling and empathy training, (3) moral psychology and philosophy, and (4) leading moral discussions with elementary school children and setting up a board for discipline issues in the high school. Fourteen juniors and seniors took part; there were two comparison groups—another psychology class and a physical science class. All groups were pretested and posttested with the SCT and the Kohlberg Moral Judgment test given in written form. Analysis of covariance verified the rise in both tests at the .001 level of significance in the experimental group, whereas neither comparison group changed significantly on either test.

Exum (1977) produced another successful program in “deliberate psychological education” at a small junior college in Minnesota. The experimental group registered for a three-quarter course in psychology. The control group was randomly selected from the remainder of

the student body. Half of each group was given the opportunity to work 4 hours a week as peer tutors or counselors; they were randomly selected from a pool of “good” students. That was called the “high involvement” condition. No one declined to participate. The “low involvement” condition for the experimental group involved teaching the communication skills being learned in class to a friend. There were 12 students in each cell of the 2 x 2 design. The experimental curriculum involved a number of elements, such as signing learning contracts, reading assignments, viewing films, learning communication skills, and keeping journals, plus a weekly meeting of each student with the instructor. The format of the class changed from structured to unstructured; the role of the instructor faded into that of a participant; and the student, correspondingly, had to shift from a passive to an active role.

On pretest with the SCT, the groups had no significant differences. On posttest, the control groups showed a slight decline, the experimental groups a gain. As predicted, on posttest the high-involvement experimental group showed the highest mean SCT scores, midway between the Conscientious Stage and the Conscientious-Autonomous transition; the low-involvement experimental group was next, with a mean at the Conscientious Stage. The high-involvement control group had a mean at the Conscientious-Conformist transition, and the low-involvement control group was marginally lower. In contrast to the SCT, the Fundamental Interpersonal Relationship Orientation scales and the Self-Actualization scale of the Personal Orientation Inventory showed no significant changes between pretests and posttests and no significant differences among groups on posttest.

Lasker (1977) and his associates have been experimenting at several companies in Curaçao with an organizational development program based on ego development theory. They found the modal pretest for workers at the Self-Protective-Conformist transition and the modal level for staff at the Conscientious-Conformist trans-

ition. Their training program, which lasted several months, appeared to achieve a modest rise in mean ego level for those initially below the Conformist Stage, but there was no evidence of success in raising ego level above the Conformist Stage. Lasker speculated that when there are few persons in the organization above the Conscientious-Conformist level, a ceiling effect may be represented.

Asymmetry of comprehension. A sixth line of evidence for sequentiality comes from asymmetry of comprehension, that is, persons understand the thinking of ego levels lower than their own, but not that of much higher levels. This is the basis for much of Rest's (1973; Rest, Turiel, & Kohlberg, 1969) work on moral judgment. In a set of studies (Redmore, 1976) the SCT was administered under normal conditions, simply asking subjects to complete the sentences. It was followed later with a retest, asking them to make a good (or bad) impression or to answer as would a person of high (or low) ego level. In the latter case, one or several lectures on ego development intervened between test and retest; however, in order not to spoil the test, sentence completions were not discussed. Several small studies included a total of 146 men and women—college students and university graduate students, plus a few persons on the staff of the psychiatry department of a general hospital (all of whom had at least one college degree).

Of those asked to make a bad impression or to answer as would a person of low ego level, most lowered their score on the SCT, often to the Self-Protective Stage. Of those asked to make a good impression or to answer as would a person of high ego level, most either stayed at the same level or raised their score one half a stage. However, a considerable number, particularly of the women, lowered their score when trying to make a good impression. (This occurred particularly where the highest scores on pretest were obtained by women; so it could not be ascertained whether the conclusion that scores tend to be lowered by trying to make a good impression was true of women or of persons of high ego

level.) Psychologically sophisticated persons, particularly those with clinical skills, can often raise their scores by approximately half a stage as compared with their spontaneous responses; hence, the SCT may be an inappropriate instrument for use with them. Redmore concluded that the ideal condition for testing is one of interest in the task without pressure to conform to social norms or expectations.

Blasi's study (1976) provided evidence for a lack of comprehension of roles belonging to higher levels. This took the form of rejecting or disclaiming or being unable to enact higher roles or, less often, misunderstanding or distorting the story so as to eliminate uncongenial roles. However, the study was not primarily designed to demonstrate asymmetry of comprehension, so there were no control groups trained at the subjects' own level or lower. Bonneville (1978), adapting Blasi's methods, tried to devise a crucial test of asymmetry of comprehension. Her results were negative, perhaps because her test proved to be more a test of intelligence than of ego level (see below).

External Component

Classical validity referred to the correlation of a test with an external criterion of what it is supposed to measure. The present context illustrates why that definition broke down and was abandoned, for what single external criterion could serve for the SCT? The concept of ego development is new, and it does not correspond to single behaviors or to membership in easily identified groups; indeed, any evidence that it did would be considered an indication that the concept is superfluous. The external component of test validity is evaluated in terms of the complete pattern of its correlations with other variables, in the light of the nomological net of theoretical predictions (Cronbach & Meehl, 1955). Table 5 summarizes data on the correlation of SCT scores with other measures.

Ego level in other behavior samples. Measures of ego level applied to other behavior

samples should correlate positively with the SCT. The higher the coefficients, the greater the evidence of validity.

Use of the SCT as a measure of ego development was originally undertaken as a method of validating the Family Problems Scale (Loevinger, 1962), an objective test whose chief cluster, called Authoritarian Family Ideology, was hypothesized to measure ego level. For a sample of 204 single women, including high school and college students and student nurses, Nettles (Loevinger & Nettles, 1963) found a correlation between Authoritarian Family Ideology and a preliminary version of the SCT of $-.32$, the negative correlation being the expected direction. Shumate (1969), using 104 junior college students and a later version of the SCT scoring manual, found a correlation of $-.22$; however, a correlation of $-.35$ between age and Authoritarian Family Ideology makes the correlation between ego level and authoritarianism less impressive. Since, however, age is easily measured, its correlations with other variables will not be attenuated by unreliability, unlike the correlations of psychological variables.

Lucas (1971) recorded interviews with 26 women students in an elementary psychology class of a public college; they also took the SCT. The interviewer and a second rater, who worked only from the transcribed interviews, judged each interview for ego level without knowledge of the SCT. As there was no scoring manual for the interviews, they could only be rated intuitively and hence not as reliably as the test. The two ratings of ego level of interviews correlated $.81$; they agreed completely or within half a stage on 25 of the 26 subjects. The correlation between ego level as judged from the interview and the compromised SCT ratings was $.58$ and $.61$ for the two raters.

Farrell (1974), using an open-ended interview modeled after that of Perry (1970), studied freshmen male students at an engineering school. The SCT had been given as part of fall registration; there was a good deal of confusion on the part of the incoming students, and their

cooperation was not ideal. Thirty of his subjects were a random sample from the class; 20 cases were added by scanning the SCTs with names covered and picking those that looked especially high or low. A secretary then supplied Farrell with the names selected by the two processes, so that in interviewing he did not know which were random and which were extreme cases. He found a correlation of $.32$, significant at the $.05$ level, between the SCT (item sum score) and the interview as rated for ego level; the SCT total protocol ratings correlated $.32$ with the same interviews rated according to Perry's scheme for stages or "positions" of ethical-intellectual development. Farrell had hypothesized that the SCT would be more closely related to Perry's scheme at the lower levels, where the descriptions have similar elements, than at higher levels, where Perry's scheme emphasizes the element of commitment not prominent in the ego development scheme. No such curvilinearity was evident in the scatter diagram. Defects in research design may have been responsible for the correlations being lower than expected, particularly the lapse of several months between the SCT and the interview, months that are eventful for college freshmen.

Brinkerhoff (1971) interviewed 37 sixteen-year old girls from the junior class of a Protestant parochial school; he found a correlation of $.48$ between ego level estimated from interviews and from the SCT. Apparently he used a single rater, whose training was not described. He appears to have selected cases to maximize variability on the SCT.

Correlation with other traits. One validation paradigm involves scanning the descriptions of ego stages for traits that have been measured in other contexts. Tests of those traits are then correlated with the SCT. The expected relationship may be positive, negative, or curvilinear, depending on the stage where the trait is expected to peak in relation to the stages found in the sample.

Shumate (1969) predicted that persons high on internal control as measured by Rotter's I-E

Table 5
Correlation of SCT with Other Measures

<i>Type of Study and Author</i>	<i>Measure</i>	<i>Sample</i>	<i>Results</i>	<i>Raters</i>	<i>Sampling</i>
Interview					
Lucas	Global rating	26 college psych F	$r = .58$ & $.61$	2	R
Farrell	Global rating	47 M; engineering	$r = .32$	2	X
Brinkerhoff	Global rating	37 parochial 16-year-old F	$r = .48$	1	X
Atkins	Gendlin's Experi- encing Scale	24 college F > I-4 23 college F < I-4	Signif. differences between groups	1?	X
Hopkins	Marcia's Ego Identity Status	69 college F	Expected order of identity statuses confirmed	2	R
Rock	Self-Confronta- tion Score	50 college F	$r = .52$	1	X
Hauser	Ratings on: Warm-Available	100 high school F	Signif. relation: Postconformist highest	2	R
	Active-Spon- taneous Sexy		Postconformist highest Preconformist highest		
Objective Tests					
Loevinger & Nettles Shumate	Family Problems Scale	204 single F	$r = .32$	2	H
Strodtbeck, Ronchi, & Hansell	Family Problems Scale	104 jr. college M & F	$r = .22$	2	R
	Rotter's I-E Test	181 inner city high school M & F in Neighborhood Youth Corps	n.s.	2	R
	Rotter's I-E Test		SCT as moderator variable for gains in efficacy	2?	R
Schenberg	Personal Orienta- tion Inventory Scales	111 high school & 61 jr. college M & F	5 signif. r 's, highest = $.29$	2	R
Rock	Heath's Perceived Self Questionnaire	50 college F	$r = .44$	1	X
Snarey & J. Blasi	Morris's Philosophy of Life Test	42 M & F kibbutz- niks; different seniority groups	Helpfulness, Cramer's $V = .44$; Scientific Solutions, Cramer's $V = .46$ n.s.	1	R
Redmore & Waldman	Mach V	34 M & F; Grade 9	n.s.	2	R
Fischer	Adaptation of Mach Scale	41 M & 39 F; black parochial, Grades 6, 7, 8	$r = .46$, M; $r = .07$, F; F peak at Δ	1	R
	Social Desirabil- ity Scale		Peak in Con- formist range		
Hoppe	Self-report of conformity	108 prep school M; Grades 8, 9, 11	Peak in Con- formist range	2	R
Harakal	Inventories of conformity	110 F; very heterogeneous	Peak in Con- formist range	2	H

continued on next page

Table 5. continued

<i>Type of Study and Author</i>	<i>Measure</i>	<i>Sample</i>	<i>Results</i>	<i>Raters</i>	<i>Sampling</i>
Projective Tests					
Redmore	TAT nAch TAT nAff	59 F; nursing	n.s. n.s.	2	R
Lasker	TAT nAch	380 M; 25-35 years old	Signif. <i>r</i> , hierarchical arrangement of themes, peak at I-4	2	R
Rock	TAT nPower TAT Self-Interpretation Score	50 college F	Peak at Δ/3 <i>r</i> = .53	1	X
Lamb	Feffer's Role-Taking Test	96 jr. high civics M & F	<i>r</i> = .24 for maximum role-taking	2	R
Adams & Shea	Marcia's Ego Identity Test (sentence completions)	136 M & 158 F; college	Signif. χ^2	?	R
Other Developmental Stage Tests					
Lambert	Kohlberg's Moral Maturity Test	109 M & F; heterogeneous	<i>r</i> = .60, age partialled out	1	H, X
Sullivan, McCullough, & Stager	Kohlberg's Test	60 M & 60 F; 12, 14, 17 years old	<i>r</i> = .40, age partialled out	2	R
Zielinski	Hunt's Conceptual Level Questionnaire Carkuff's Empathy Test	40 M & F; graduate counseling	<i>r</i> = .23, age partialled out <i>r</i> = .46, communicate empathy; signif. <i>r</i> with gain in discriminating empathy	2	R
Behavioral Measures					
Cox	Helping behavior Service activities	79 black & white M; Grade 8	n.s. n.s.	2	R
Blasi	Responsibility level rating	35 black F & 32 black M; Grade 6	<i>r</i> = .54, F <i>r</i> = .56, M	2	R
Goldberger	Behavioral ratings, etc.	16-year-old college; varying <i>N</i> 's, up to 266	Behaviors confirm predictions	?	R
Hoppe	No. of demerits (nonconformity)	108 prep school M	Curvilinear relation	2	R
Frank & Quinlan	Delinquent behavior; status offenses	66 F black & Puerto Rican teenagers	Deviant behaviors, especially fighting, signif. more frequent at low stages	1	R
Powitzky	Type of federal offense	112 M prisoners	Most predictions confirmed	2	R

Scale would be higher in ego level than those who tested as externals; the prediction was not confirmed in his sample of 104 junior college students.

Schenberg (1973) predicted a positive correlation between ego level and several tests of time perspective and several scales of Shostrom's Personal Orientation Inventory. His samples were 111 high school students and 61 junior college students. Correlations of the SCT with various time perspective measures were mostly low positive, but not statistically significant; however, the convergent validity of the time perspective measures was essentially nil. The 12 correlations between the SCT and the Personal Orientation Inventory scales were positive, five of them significant at the .05 level. The highest was a correlation of .29 with a scale called Constructive Nature of Man.

Cox (1974) predicted more "helping" behavior with higher ego level; her experiment used 79 eighth-grade boys and involved helping another boy, a confederate, complete a puzzle. The prediction was not confirmed; the boys were more responsive to direct modeling of helping. Among 34 black students, there was a significant correlation between SCT and sports participation; among 45 white students there was a significant correlation between SCT and participation in educational activities. There was no relation between SCT and participation in school or community service activities. There was a trend in the data, although there were not enough cases to test it, that those of low ego level were more responsive to modeling than those of median or high ego level. A possible reason for failure to find a correlation between altruism and ego level is that in the school setting, helping another person with a school-type task may be thought of as cheating rather than as helping.

Lamb (1971) predicted a correlation between ego level and role-taking as measured by a test adapted from Feffer (Feffer & Jahelka, 1968), in which students wrote stories about pictures. For her sample of 96 junior and senior high school civics students, there was a correlation of .24

(significant at the .02 level) between SCT and the maximum role-taking score, but no significant correlation with mean role-taking. The relation probably was attenuated by testing procedures, which permitted students to take time from one test to finish the other. Thus, involvement in one test may have left inadequate time to respond fully to the other.

Redmore and Waldman (1975) studied 34 ninth-grade suburban children, finding no correlation between SCT and a form of the Machiavellianism (Mach) scale, Mach V, (Christie & Geis, 1970). However, Fischer (1973), using an age-appropriate adaptation of the scale with sixth-, seventh-, and eighth-grade black parochial school students, found a positive correlation between Mach and SCT among 41 boys, none of whom tested higher than the Conscientious-Conformist transitional level on the SCT. Among a similar sample of 39 girls, however, Mach showed a decided peak at the Self-Protective Stage, as predicted. Goldsmith (1971) tested 57 first- and third-year law school students, 47 of them men. He found a slight negative correlation between SCT and Mach V scores; however, scores were heavily concentrated at the Conscientious Stage. The single case at the Self-Protective Stage in his sample had the highest Mach score. These results are fragmentary and equivocal. Perhaps a manipulative Machiavellian attitude is more characteristic of the person who remains at the Self-Protective Stage into maturity than of the youngster who is traversing the stage at an appropriate age.

Redmore (1969) predicted that those at the Conformist Stage on the SCT would be comparatively high in need for affiliation, as measured by the Thematic Apperception Test (TAT), whereas those at the Conscientious Stage would be relatively higher in need for achievement. The prediction was not confirmed for a group of 59 student nurses. Those at the Conscientious Stage gave significantly more affiliation responses to TAT pictures of females than to pictures of males, an effect not found at lower levels; they also showed a trend to give more

achievement responses to male pictures, but these effects had not been predicted.

Lasker (1978) studied the achievement motive on the island of Curaçao in the Netherlands Antilles. He began by working with McClelland's motivation training groups in India, Appalachia, and Curaçao. On the basis of observations in those settings, he predicted that the number and quality or categories of achievement themes in TAT stories would be related to ego level. He studied 380 persons, mostly men from 25 to 35 years of age, drawn chiefly from 7 diverse companies and 1 governmental agency. The sample reflected the heterogeneity of Curaçao with respect to socioeconomic status, race, religion, language, and ethnic origin. Testing occurred in small groups in Dutch, English, Spanish, or Papiamentu, according to the person's preference. Six TAT pictures were used, adapted from McClelland's standard set to make them more appropriate for Curaçaoans. The stories were scored by standard procedures for needs for achievement, power, and affiliation. Responses to the SCT were translated into English for scoring.

As predicted, Lasker found that the mean need for achievement score was low, below the Conformist Stage, and rose sharply to the Conscientious Stage, probably peaking there. The result continued to hold when the stories were corrected for length. Further, Lasker showed that the standard categories into which achievement themes have been classed form a Guttman-type scale. The same hierarchical arrangement of achievement categories held in other data obtained within the U.S., including McClelland's original data. The power motive, by contrast, peaked at the Self-Protective-Conformist transition, and the affiliation motive showed still a different stage pattern.

Several studies have been concerned with correlations between ego level and measures of empathy and insight, which might be presumed to differentiate among levels, particularly at the high end of the scale. Zielinski (1973) ad-

ministered the SCT to a group of 40 graduate students in a beginning course in counselor education. They were also given Carkhuff's index of communication and his index of discrimination of empathic understanding before and after the course, which was a 20-hour training program adapted from Carkhuff's model. On pretest there was a correlation of .46, highly significant, between SCT and ability to communicate empathic understanding, but no significant relation with ability to discriminate empathic understanding. There was, however, a significant relation between the SCT and gains in ability to discriminate empathy. Obviously, a significant relation between SCT and both pretest and gain scores cannot be expected; hence Zielinski's results—with SCT significantly related to one index on pretest and to gains in the other—is evidence for the construct validity of the SCT.

To maximize variability in ego level, Rock (1975) selected 50 undergraduate women from a larger pool tested with the SCT. Two experiments were performed with each subject. In the first experiment, each woman was recorded while talking about a neutral topic, "My first job"; when the tape was played back, she was asked what her voice and manner of speech told about her as a person. Her responses were taped, the tapes transcribed, and the transcriptions rated by a scoring system ("Self-Confrontation") designed for the study. In the second experiment, each woman subject was requested to tell stories to six TAT cards according to standard instructions. Afterwards, the theory of projection was explained to her, and her stories were read back to her one at a time. She was asked to interpret each story in terms of the influence of her perceptions and experiences. Her responses were scored by an adaptation of the TAT self-interpretation score of Luborsky (1953). Then she was given Heath's (1968) Perceived Self Questionnaire, which is a measure of personal maturity (like the SCT) and the information subtest from the WAIS, as a measure of intelligence.

The two measures of self-insight correlated .53 and .52 with the SCT, highly significant, as predicted; indeed, they correlated almost as highly with ego level as they did with each other, .60. The correlations remained highly significant when intelligence or age was partialled out. (Surprisingly, age correlated .40 with the SCT in this college group.) Heath's test of maturity correlated .44 with the SCT. Rock's conclusions, based on detailed analysis of data from his measures of self-insight, follow.

Subjects at the Conformist and preconformist levels tended to be unreflective about their experience or to be reflective in nonpsychological terms. The shift to psychologically minded self-reflection occurred in a minority of subjects at the Conscientious-Conformist transition and a majority of those at the Conscientious Stage. Almost all of those at the Individualistic level and the Autonomous Stage saw themselves in complex psychological terms, and they often had a dynamic understanding of the reasons for their behavior and their inner conflicts. Rock also concluded that Heath's Perceived Self Questionnaire was less valid as a measure of personal maturity than the SCT and more influenced by defensiveness.

Atkins (1976) selected 24 high ego level women (above the Conscientious Stage) and 23 (relatively) low ego level women (below the Conscientious Stage, mostly at the Conformist or Conscientious-Conformist levels) from 120 University of Chicago graduate and undergraduate students and University of Illinois at Chicago Circle undergraduates. Each woman was interviewed by a graduate student or a professional who had been trained in client-centered therapy. The women were instructed to talk, as if they were in psychotherapy, about a problem related to being a woman. Then they filled out some questionnaires. High ego level women showed higher experiencing levels on the interview, rated on Gendlin's Experiencing Scale (Klein, Mathieu, Gendlin, & Kiesler, 1969), than did low ego level women. In particular, 42% of the high SCT women had a modal rating at Ex-

periencing Stage 4 in at least one segment of the interview, whereas only 4% of the low SCT women did; Experiencing Stage 4 and higher is considered the "therapeutic range."

On a postinterview questionnaire, the women were asked to describe briefly the problem that they had chosen to discuss. The low (Conformist) women tended to write about control of emotions, concrete aspects of women's role, and home, family, and mother. Highs tended to write about their relations with others, abstract aspects of women's role, identity, and independence. In describing their conception of psychotherapy and its purpose, lows tended to stress working out solutions; realizing, alleviating, and coping with feelings; and adjustment. Highs discussed self-understanding; the self as hidden or internally blocked; the relationship between therapist and patient; patterns of behavior and being stuck in them; and connections between thoughts, feelings, and actions.

Correlation with other developmental conceptions. Another set of studies compares the SCT with other developmental stage conceptions. Farrell's comparison of ego level with Perry's conceptual-ethical level can be considered in that category.

Lambert (1972) studied the relation between the SCT and Kohlberg's measure of moral judgment maturity, using a manual for the Kohlberg test which is now, however, obsolete. In order to tap the full range of both scales, he used six subsamples varying widely in age and in education (see above). The correlation between the SCT and the global rating on the Kohlberg was .80 over the entire sample, a value inflated by the wide age range. For a subsample of 69 eleven- and fifteen-year-olds, the correlation was .46, a value diminished by the restricted range. Using his figures, the correlation for the entire range with age partialled out is .60.

Sullivan, McCullough, and Stager (1970), using the data from Toronto school children that Loevinger and Wessler used to show age progression in ego level, found a correlation of .66 between the SCT and Kohlberg's test (again,

a now obsolete version); with age partialled out, the correlation became .40. Considering the age groups separately, for those about 12 years old, the correlation between SCT and Kohlberg scores was .19; about 14 years old, .48; about 17 years old, .54. These groups included 20 boys and 20 girls at each age level, selected randomly from the subjects at that age. For the same sample, Hunt's Conceptual Level Questionnaire (Hunt & Halverson, 1964) correlated .56 with the SCT, .23 with age partialled out.

Hopkins (1977), studying a group of 69 college women, found a significant relation between the SCT and ego identity status on Marcia's test of Erikson's identity typology. In order of increasing ego level, the identity groups were Diffusion, Foreclosure, Moratorium, and Achievement. This result is consistent with the expected developmental progression. Adams and Shea (1979) gave the SCT and a sentence completion version of Marcia's test to students at Utah State University. They found the same order as Hopkins did, with a highly significant chi-square.

Behavioral and group membership correlates. Suppose it is theorized that ego level should correlate with Trait X. If Trait X is measured as defined by another investigator in his or her own theoretical context, the very aspects related to ego level may be omitted. If, on the other hand, Trait X is measured by a new scale devised to reflect exactly the aspects most related to ego level, the construct is being validated in a much narrower sense; in effect, virtually the same scale is being applied to two different behavior samples. (I owe this paradox to A. Blasi.) One way to transcend Blasi's paradox is to seek behavioral manifestations of the several levels. A related criterion uses membership in certain predictable groups as an index of some ego levels. Since, as a rule, direct translation of ego level into specific behaviors is not expected, these correlations ought not to be high. Such topics as conformity, delinquency, responsibility, and occupational choice have been investigated in relevant studies.

Blasi's study (1971, 1976) illustrates one way of handling his paradox. Blasi delineated a 10-point scale of levels in the development of responsibility, using stages of ego development as one major source. After each role-playing session (see above), his subjects were rated by the experimenter and the observer on a 3-point scale for role-playing ability. Also, a checklist for behaviors characteristic of the several levels was filled out by each rater for each child. At the end of the 2 weeks of training, each child was given a global rating for role-playing ability and general level of responsibility.

The correlation of the SCT with responsibility level was .56 for boys and .54 for girls. The correlations of SCT with IQ were lower (.46 and .50), and with role-playing much lower (.29 and .06). Since this was an ad hoc unstandardized scale for responsibility level, it can be assumed to be much less reliable than the IQ; hence this is supportive of validity. Interpretation of the results is clouded, as Blasi pointed out, by the fact that the substance of the stories was different for children of different ego levels; however, a 2×3 analysis of variance of the responsibility ratings showed a highly significant effect of ego level and no significant effect of treatment or treatment by ego level interaction, again confirming the validity of the SCT.

Another aspect of Blasi's study provides behavioral validation of the SCT; that is the record of the behavior of the children in groups sorted for ego level by the SCT. Unfortunately, the study was not planned with that use of the data in mind, so the evidence is not crucial. One of the two observers who recorded the children's behavior knew the experimental design and was not blind to the ego level of the group. What renders the data impressive, nonetheless, is that the description of the various levels had been derived up to the time of his study almost entirely from older subjects, and almost entirely from white females. Blasi's subjects were all black; at the lower age limit for the test, 11 or 12 years; and both boys and girls. Even more, the descriptions of levels had been inferred almost entirely

from sentence completions, with only rarely an occasional bit of feedback concerning the actual behavior of some subject. Thus, the jump from patterns of sentence responses to patterns of behavior was a large one and, on the whole, impressively successful (Blasi, 1976).

The behavior characteristic of the several levels was as follows: The Impulsive groups were different from the others; they tended to be lower intellectually, were poor in discrimination of feelings, lacked insight into motives, had short attention spans, were restless, misbehaved, disrupted the sessions, and giggled thoughtlessly. Although they had no difficulty seeing themselves as causes of negative behaviors, they tried to shift the blame as did the Self-Protective children, but were more naive and absurd in their reasons. Their role-playing was limited to actions and stripped of details. Their greatest interest was aroused by a chance to attack each other or "rat" on each other to the experimenter. They saw behavior as dependent on authorities and on use of punishments to check misdemeanors. In enacting the role of authority, they recommended disproportionately harsh punishments.

The Self-Protective children role-played effectively, especially the role of the sneaky, opportunistic hero. Children of higher level also played this role well, but tended to indicate a disapproval of it. The Self-Protective children were concerned with being found out and with protecting themselves. As heroes, they thought that sneaky solutions were appropriate and smart; when caught, they demanded indulgence and leniency. As authorities, they were rigidly authoritarian and harshly punitive. They recognized the differences in the roles, but could not coordinate the points of view. They could not portray shame, being instead sullen and defiant. Nor could they portray the authority trying to induce shame, resorting instead to screaming. They could not change even when they saw that yelling was arousing anger instead of shame.

The Conformist children recognized the function of rules more than did the children of lower

levels, but they clung to literal interpretation of the rules, with their responsibility limited to obedience. Rather than just avoiding punishment, they saw love for and loyalty to authority as motivation for obedience. Although they did not express anger or hostility to authority, they denigrated authorities for liberal interpretation of rules, which they interpreted as two-faced or indecisive. The small group of children at the Conscientious-Conformist level were like the Conformist children, but even more disapproving of sneaky and pragmatic attitudes and more sensitive to conflict between independence and obedience.

Hauser (1978) gave the SCT to 100 suburban high school girls who volunteered for a study of interpersonal relations. His procedure was to give each girl a structured interview, elicit her fantasies to TAT cards, and then follow with another structured interview. The interviewers were eight young men—four black and four white—assigned equally to the several ego levels. Videotaped segments of the interviews were rated on scales of Warm-Available, Active-Spontaneous, and Sexy. On the initial interview the postconformist girls (Conscientious Stage or higher) were higher on Warm-Available than the Conformists, who were higher than the preconformists; these differences were accentuated in the postfantasy interviews. The same order held for ratings of Active-Spontaneous. On the Sexy scale, however, the preconformist girls were rated highest.

Hoppe (1972; Hoppe & Loevinger, 1977) reasoned that measures of conformity should peak at median ego levels, falling off among preconformist and postconformist subjects. His subjects were 41 eighth graders, 37 ninth graders, and 30 eleventh graders in a private school for boys. The prediction was confirmed for two measures of conformity—a self-report measure and a nonreactive measure (number of demerits recorded in the headmaster's office). Relationships remained significant with age and grade held constant. Results for a peer rating for conformity were ambiguous, partly for method-

ological reasons. An adaptation of Willis's experimental measure of conformity, an Asch-type measure, did not correlate with any of the other variables.

Fischer (see above), using the Social Desirability scale from his adaptation of the Mach scale, found a predicted peak in the Conformist range, at the Conformist Stage for girls and at the Self-Protective-Conformist transition for boys. However, these trends were not statistically significant.

Harakal (1971) found a curvilinear relation between the SCT and several indices of conformity, as measured by the Interpersonal Style Inventory (Lorr & Youniss, 1969) and the Schedule of Interpersonal Responses (Kinnane, Lorr, & Suziedelis, 1969). The former is a true-false self-report inventory, the latter an objective test asking the subject to predict another person's response in an interpersonal situation. All the statistically significant differences were in the predicted direction. Her sample of 110 adolescent and young adult women was extremely heterogeneous, ranging from institutionalized delinquents and high school students to law school students and women with college degrees working for a national service organization. Of the 32 delinquent girls, 13 were rated Impulsive or Self-Protective, as compared with only 4 of the 23 high school girls. The extreme diversity of the subsamples, evidently used to ensure a wide range of ego levels, clouds any conclusions from this study.

Frank and Quinlan (1976) compared delinquent adolescent girls with two control groups matched for age, social class, and ethnic background (black and Puerto Rican). The delinquents were all institutionalized status offenders, their most frequent offenses being running away, "incurability," and sexual delinquency. Length of stay in the institution at the time of testing was from 2 days to 2 years, with a median of 5 months. One control group was drawn from participants in a settlement house recreational program in the same neighborhood; an additional control group was formed by 16

girls in the settlement house leadership training program. The interviewer recruited subjects by "hanging around" with them in informal settings, in order not to appear to be part of the administration. Only one of the subjects invited to participate did not do so. Girls were tested individually with the SCT, administered orally, and the Peabody Picture Test of Intelligence. Their self-report of deviant behaviors was also obtained; these reports were corroborated by social workers and counselors. SCT raters did not know group membership.

Intelligence did not differ significantly among the three groups, though it correlated significantly with ego stage ($r = .31, p < .05$). The delinquent girls differed significantly in ego level from the other two groups, which did not differ from each other; this remained true when intelligence was covaried out. There were a large number of girls from all three groups at the Self-Protective Stage; at the Impulsive Stage, however, there were 12 of the delinquents but only one of each control group. The overall number of deviant behaviors correlated $-.45$ with ego stage; fighting alone correlated $-.52$. Other behaviors with a significant correlation with ego stage were running away, promiscuity, homosexuality, and alcohol abuse. While the latter behaviors confirm the description of the lowest ego levels, they are redundant with the fact of delinquency in these status offenders.

Cresswell and Lacks (no date) tested a group of 30 delinquent adolescent girls from a suburban juvenile detention center and a control group of 30 girls from a suburban junior high and high school of about the same age, race, achievement scores, family structure, and socioeconomic status. The delinquent group included 22 girls rated Impulsive or Self-Protective on the SCT; the control group had only 6. Because of their youth, this delinquent group did show a predominance of preconformists, unlike Harakal's and Frank and Quinlan's samples; however, delinquency again proved compatible with Conformist and postconformist ratings on the SCT.

Powitzky (1975) studied 112 young male offenders, contrasting those incarcerated for different federal offenses. He hypothesized that they would range in order of increasing ego level as follows: car thieves, bank robbers, opiate offenders, marijuana offenders, embezzlers, and conscientious objectors. (Opiate offenders were men sentenced one or more times for offenses related to heroin, cocaine, or opium. Conscientious objectors were men sentenced one or more times for violation of the Selective Service Act and also listed in *The Peacemaker*, a peace movement publication that published a monthly list of resisters in prison or under prosecution, excluding Jehovah's Witnesses. Most subjects came from Seagoville, a minimum security federal prison; most bank robbers came from El Reno, a medium security federal prison. Since many embezzlers and conscientious objectors refused to be tested, cases were sought in other prisons as well.)

The prediction that the groups would differ in ego level was confirmed ($p < .01$); and the predicted order held up with the exception that embezzlers did not score lower than conscientious objectors, considered here the equivalent of a normal control. Because of confounding of variables, it is difficult to interpret many of Powitzky's findings, such as the significant relation between ego level and number of prior offenses: car thieves tend to be repeaters and tend to have the lowest ego level. One reason for some surprisingly high ego levels is that for most offenses there are distinct types of offenders. For example, car thieves may be of an impulsive type or members of an organized crime ring. Also, most cases were drawn from minimum security institutions, which implies the selection of more controlled and trustworthy men. None of the 22 conscientious objectors scored below the Conformist Stage, whereas 21 of the 92 other offenders did.

In another study of delinquents, 107 inmates of a county adult correctional facility outside a major midwestern city were given the SCT, half in 1972 and half in 1974 (Mikel, 1974). Staff

members (not including guards) filled out an adjective check list describing the men they knew well. On this basis, descriptive portraits of the various levels could be drawn. (The adjectives descriptive of a level were not necessarily used in all or even a majority of instances at that level, but tended to be used more often than for men of other levels.) As the numbers were small and the study was primarily exploratory, tests of significance were not performed. The pre-conformists as a group were considered rebellious, unrealistic, immature, and untrustworthy. Men rated at the Impulsive Stage on the SCT were characterized as impulsive, easily provoked, and undisciplined; those at the Self-Protective Stage were more likely to be seen as gregarious and critical of procedures. Those at the Conformist and Conscientious-Conformist level were seen as adjusted, not rebellious, and dogmatic. Post-conformist men were seen as truthful and receptive to counseling. Overall, there were 39 men (36%) at the preconformist levels, 54 (51%) at the Conformist and Conscientious-Conformist levels, and 14 (13%) at postconformist levels. Although no control group was obtained, this seems to indicate an overrepresentation of preconformist subjects for a group of late adolescents and young adults.

Two studies have related the SCT to vocational choice or adjustment. Sheridan (1974) asked vocational adjustment coordinators in 4 Houston area school districts to identify the 25 educable mentally retarded clients who were most likely to succeed in employment and the 25 who were least likely to succeed. The two groups so chosen did not differ significantly in age (16 to 19.5 years), ethnic composition, measured IQ, or sex. They did differ significantly with respect to the SCT and to Kohlberg's test of moral reasoning. Even with this restricted range (Impulsive to Self-Protective-Conformist transitional level), the two tests correlated .67, a highly significant value.

Schain (1974) gave the SCT to 100 women in second year master's degree programs—55 from fields traditionally open to women, such as

social work, guidance, and counseling and 45 from the schools of business and government, with such majors as business administration, urban and regional planning, and health care. Contrary to her expectation, the women in business had a higher level than those in the social-supportive fields. One interpretation of this finding is that it takes a higher ego level for a woman to break out of the mold of traditional expectations for her social role. These results must be replicated, however, as Schain's subjects were given a packet of tests to do at home, which is not a favorable circumstance for control of extraneous variables. Also, there may be other ways to interpret her results, since her groups were not a random sample from traditional and nontraditional fields.

Goldberger (1977) has used the SCT as a measure of personal growth at Simon's Rock College, a small liberal arts college in New England which encourages early admission; entering freshmen are usually 16 or 17 years old. She recorded a decline in the number of preconformists and a rise in the number of postconformists after the freshman year; the trends were more marked after the senior year. There was no consistent relation between academic ability or achievement in the freshmen year and ego level. There were, however, marked differences in behavior. Postconformist students as a group were more serious about their studies, were rated by faculty (who did not know their SCT scores) as constructive models and leaders on campus, were more responsive to a values seminar, and tended to have psychological problems. If they left Simon's Rock before obtaining a degree, they tended to transfer to other colleges. The preconformist students were less serious and ambitious about their academic work, tended to be unrealistic about factors which contribute to life and career success, were catalysts for disruptive activities, were more subject to disciplinary action, and were apt to leave the college as dropouts or to be suspended. These observations pertain to samples of varying numbers, ranging from 266 freshmen to 14 graduating seniors.

Snarey and J. Blasi (1978) studied 42 subjects in 4 cohorts of residents on an Israeli kibbutz founded in the 1930s: a group of the original founders; senior members who joined after World War II; older sabras, most of whom were married; and a group of young sabras, many unmarried. Tests were given in Hebrew and translated into English for scoring. The founder cohort was significantly higher in mean ego level than the others. In giving their motivation for joining the kibbutz, five out of six of those at the Conformist or Self-Protective Stage gave nonideological reasons, 7 out of 18 at the Conscientious-Conformist level gave nonideological reasons, whereas all 9 at the Conscientious Stage or higher gave ideological reasons. Of the 13 orientations in Morris's Philosophy of Life Test, 2 were significantly and positively associated with ego development—Helpfulness and Unselfish Concern for Others, and Active Solution of Problems by Scientific Advances.

Strodtbeck, Ronchi, and Hansell (1976) provide a different kind of external validation. In connection with a federal program called the Neighborhood Youth Corps, inner-city youths from poor families were given a variety of jobs enabling them to earn about \$35 per week in order to keep them in high school. About half the students in the program were hardworking and school oriented, whereas the others were multi-problem and retarded in reading skills. As an experiment, many of the students were assigned to tutor younger children; these assignments were not limited to those academically able or motivated. SCT scores, by schools, ranged from a mean between Impulsive and Self-Protective to a mean at the Conformist Stage. Students of relatively high ego level tended to have fewer days' absence from school the previous year and to have a slightly higher sense of efficacy as measured by Rotter's I-E Scale. The higher the ego level within the range covered by this study, the greater the gain in efficacy scores for tutors in the 4- to 5-month program, with no gain for those low in ego level. But for those given custodial jobs, the lower the ego level, the greater the gain in efficacy, with no gain for those of rel-

atively high ego level. Thus, ego level appeared in this study as a moderator variable mediating the relation of work assignment and gain in efficacy.

Distortions and Confounded Variables

Extraneous variables correlating with a test score may be distortions of measurement, in which case the lower the correlation, the more valid the test. They may also be variables inextricably confounded with the continuum at issue, in which case no valid test can achieve a low correlation with the variable. Indeed, a fairly high correlation could be interpreted as evidence for validity. The problem is that there is no easy way to separate the two cases. In either case, however, tests must demonstrate incremental validity (Sechrest, 1963), that is, they must predict better than can be done from age, sex, socioeconomic status, and school grade or highest educational level reached.

In projective tests the chief distortion to be contended with is verbal productivity, which has the same effect that response bias does in objective tests. Although verbosity is a distortion of measurement, number of words per response is not a direct measure of it. Complex thoughts referring to interpersonal relations, psychological causation, or development of the sense of self (all indications of high ego level) require more words to set forth than simple thought referring to talking, having fun, or being happy (all indications of lower ego levels). Number of words per response thus reflects both verbosity and an element of complexity that is a legitimate index of ego level. Therefore, a moderately high correlation of SCT with number of words is not, *ipso facto*, evidence against construct validity.

For their Sealed Sample, Loevinger and Wessler (1970) found that the median correlation between item rating and total protocol rating was .44, whereas the median correlation between the item word count and total protocol rating was .35. The sum of item word counts was more reliable than the sum of item ratings—coefficient

alpha of .95 for the former, .91 for the latter. However, correlation with the total protocol rating for the word count sum was .58; for the item rating sum, .93. This circumstance, higher reliability and lower correlation with the criterion, indicates that there is a systematic source of distortion (Loevinger, 1954), which is also intuitively the case for word count as opposed to item rating by the manual.

Average differences in SCT scores for different socioeconomic groups have been reported by Lasker (1977, 1978) for Curaçao and by Kusatsu (1977) for Japan. They are apparent also in the studies of Redmore and Loevinger (1979), though the studies were not directed to that problem. Putting together results from fairly comparable samples from different social strata in different studies, a mean difference favoring the higher social groups is invariably found. In addition, the rate of growth seems to level off before the end of high school, at least in some inner city black samples (Coor, 1970), whereas for some upper class groups, growth seems to continue into the college years (Loevinger, 1978).

The correlation of the SCT with intelligence has been reported in a number of studies (see also Hauser, 1976). Blasi (1971) correlated the SCT at sixth grade with IQ as measured by the Lorge-Thorndike Intelligence Test, obtained from school records. He used the SCT, given prior to his experiment but scored later, and included a few students who did not take part in the role-playing. For the 55 boys, the correlation was .46; for the 62 girls, the correlation was .49. Coor (1970), using a similar inner-city sample, found a correlation of .27 between IQ and the SCT in a group of 108 seventh graders, about equally divided between boys and girls. (The higher values that Redmore and Loevinger give for these studies are based only on retested subjects, not on the entire sample.) For 34 boys in an integrated suburban school, very heterogeneous as to economic and intellectual level, the correlation between the Henmon-Nelson test in seventh grade and the SCT in eighth grade was .60. In Bonneville's (1978) study, the SCT

correlated .37 with Lorge-Thorndike IQ for 77 sixth-grade girls and .36 for 84 eighth-grade girls from regular classes; when the girls in the gifted classes were included, the correlations were .59 for the sixth grade and .44 for the eighth grade. Hoppe (1972) found a nonsignificant correlation of .14 between ego level and IQ for his prep school sample of 107 boys. When correlations were computed for the classes separately, values ranged from -.13 to .33 (Redmore & Loevinger, 1979). The tests were the Junior Scholastic Ability Test and the Canadian Academic Aptitude Test.

Vetter (1978) used analysis of variance to test the relation between verbal ability and the SCT for 103 boys and 95 girls between 13 and 16 years of age in German schools; there was no significant relation for the sexes separately, but a relation significant at the .05 level of significance for the sexes combined. Sheridan (1974) reported that there was no relation between SCT and IQ in her retarded group of late adolescents.

Schenberg (1973) found a correlation of .43 between the SCT and the Ohio Psychological Test in his group of 85 students in a high school psychology course. For his junior college sample of 48 students, there was a nonsignificant correlation of .17 between the SCT and the Cooperative School and College Ability Test (SCAT) verbal score, and a significant negative correlation of $-.33$ with the SCAT quantitative score. Farrell (1974) found no significant correlation between the SCT and the College Board Scholastic Aptitude Test (SAT) verbal score in a sample of 46 freshmen at Worcester Polytechnic Institute. Rock (1975) found a correlation of .30 between the SCT and the WAIS information subtest. That compares with a correlation of .52 between the SCT and the presumably less reliable measure of self-insight in the same college women. Candee (1974), studying 76 students who had participated in leftist activities at the University of Chicago in 1969 and 1970, found a correlation between the SCT and the SAT math score of .11; the correlation with the SAT verbal score was $-.27$. The higher correlation of SCT

with a math than with a verbal score is so unusual as to suggest either a misprint, an error in computation, or unusual sampling circumstances. These results show the extreme variability in the relation between ego level and intelligence, even in somewhat similar samples.

Discussion

The fairly widespread use of the SCT, including its translation into a number of other languages, shows that many people have accepted a presumption of its validity. That can only be on the basis of face validity (that is, substantive or content validity), since only a small portion of the other evidence in this review has been widely available. The evidence for the communicability and coherence of the conception of ego development is strong. The statistical homogeneity of the test and the impossibility of breaking the items into meaningful or statistically supportable subscales are also firmly established.

As in any projective test, there are many sources of unreliability, including rater differences, variability in subject performance, and (perhaps most difficult) deliberate noncooperation on the part of some subjects under some circumstances. The faking studies show the test is not immune to such effects. These sources of unreliability put an upper limit on the validity of the test. No important decision on an individual case should rest on an uncorroborated score on this one test, a statement that probably applies also to other personality tests. The same unreliability puts an upper limit on the coefficient of validity when the SCT is used for research purposes and leads to underestimating the extent of correlations of other variables with ego development.

The several kinds of evidence for sequentiality are generally supportive for construct validity, but not decisive for sequentiality. Unreliability makes a rigorous test of sequentiality impossible. The evidences for sequentiality (Table 3) are as follows: The profile or distribution of item ratings for protocols at adjacent stages are more

alike than at nonadjacent stages. Mean ego level increases with age for cross-sectional samples during adolescence. Mean ego level increases with age for longitudinal studies during adolescence. Test-retest correlations are significantly positive for longitudinal studies, even over spans of 6 years. Attempts to raise ego level experimentally in a few weeks have not succeeded, but experiments that have lasted 6 to 9 months have had statistically significant success. Finally, people can lower their score more reliably and decisively than they can raise it.

A major problem in evaluating the evidence with respect to the correlation of the SCT with other measures is the looseness of the logic defining theoretical expectations. There do not seem to be any important studies with statistically significant negative results, that is, results that are in the opposite direction of the hypothesis. Some studies failed to find expected results, and many more found results in the expected direction, but of smaller size than anticipated. The question to be asked is, How high are correlations between the SCT and other tests expected to be? There are many circumstances that can attenuate a relation, including the unreliability of the SCT, unreliability or invalidity of the measure of the other variable, narrow range of either or both variables, and various errors in research technique.

As Table 5 shows, the evidence for the external component of construct validity cuts across techniques, including studies based on interviews, objective tests, projective tests, and behavioral measures. Tests of other developmental stage theories are interpreted here as evidence for convergent validity. For example, correlations with Kohlberg's test are of the same order of magnitude as correlations with measures of ego level applied to other behavior samples. Kohlberg, Marcia, Perry, and Heath might prefer to interpret the moderate correlations between the SCT and their tests as evidence for divergent validity.

This review has excluded some kinds of research with the SCT, such as study of race and

sex differences, because they have no clear-cut implications for construct validity. Correlations of the SCT with IQ, socioeconomic status, and word count are germane, but ambiguous in import. Granted the moderately high correlation between intelligence and ego level in some samples, how shall it be interpreted? The conclusion most unfavorable to the SCT is that ego level is nothing but intelligence, poorly measured; but this conclusion requires that much data be ignored, including particularly the evidence for content validity. Or it may be concluded that the measure of ego level is, at least in some samples, unduly contaminated by IQ as an extraneous variable. Or it may be concluded that intelligence and ego level, though quite different aspects of a person, are for various reasons highly correlated in available groups. Finally, it may be concluded that in some cases the measure of IQ may be contaminated by ego level. That is not so farfetched as it may sound. Schools have long had to contend with sullen and uncooperative adolescents who seemed, but were not actually, stupid.

Probably there is some truth to all of the last three interpretations. At least it seems safe to conclude that whatever the SCT measures, it cannot be completely dismissed as nothing but IQ, socioeconomic class differences, or verbosity, considered singly. As a fallback position, even if it be maintained that the SCT variable can be completely accounted for with IQ, socioeconomic status, and verbosity together, the conception supplies some interesting slants on those scoring low and high on the variable. The stage descriptions could not have been deduced from the correlates.

For ego development as a construct and the SCT as a measure to be considered seriously, there must be some correlations and some group differences in other variables that can be accounted for better by ego development than by intelligence alone (Sechrest, 1963). Thus, special attention must be paid to the few studies that permit a relevant comparison. Bonneville's study does not show a significant difference be-

tween the prediction of behavior from IQ and from SCT, suggesting that at sixth grade, ego level—or at least SCT—is closely tied to IQ, hence difficult to study independently. Blasi, studying the same age range, found somewhat better evidence in favor of the SCT as partly independent of IQ. Frank and Quinlan's delinquents remained significantly lower in ego level when intelligence was covaried out. Rock, studying college students, found the SCT far more predictive than the WAIS subtest he used. Several studies of prep school and college students have not found high correlations with intelligence.

The SCT must be validated, not just as a whole, but in each part of the scale. Evidence of validity pertaining to the preconformist stages is shown by overrepresentation of preconformists in all studies of delinquents and by the relation of specific deviant behaviors, such as fighting, to low ego levels. Evidence for the validity of scores in the Conformist range is the curvilinear relation between measures of conformity and the SCT. Evidence for validity of scores in the post-conformist range consists not so much in specific behaviors as in correlations with attitudes, philosophy of life, insight into self, ability to communicate empathy, and the like.

Among the most interesting studies have been those that used unique methodologies. Strodtbeck, Ronchi, and Hansell's use of the SCT as a moderator variable is an example. Another is Lasker's correlation of ego level with use of specific categories of achievement themes, as defined by previous research not related to ego level.

It seems useful to summarize the methodological strengths and weaknesses of the studies cited here. The requirement that the raters be blind to the subgroup of the protocol (pretest vs. posttest, experimental group vs. control group, and age or grade group) appears to be generally observed. Most studies appear to have used adequately trained raters, though in many cases only one rater was used for at least some of the protocols. (Where a percentage of

protocols was scored by a second rater to provide an interrater reliability estimate, the study has been recorded in Tables 3 and 5 as using one rater.) A few studies have mailed out the test or asked the subjects to take it home and mail it back. This procedure appears to be too risky in the light of the work required to score the test adequately, though there may be rare exceptions where the subjects are highly motivated to participate.

Sampling is a problem in most studies. Any setting is likely to yield persons within a narrow range of ego levels. Most college groups, for example, are predominantly at the Conscientious-Conformist and Conscientious levels. A sample that has few persons beyond those two levels cannot yield high correlations with other variables, given the modest test reliability. High school groups from a single grade are likely to be even more homogeneous. Extending the range of ego levels by drawing additional groups from other sources introduces a new problem, for the various samples will invariably differ in ways other than ego level. Correlations between variables in such a combined sample may be due to extraneous factors. Achieving matched groups through a wide range of ego levels is difficult or impossible. Some samples, such as those of Atkins, Bonneville, Brinkerhoff, Farrell, Hauser, Lambert, and Rock, were selected on the basis of the SCT so as to overrepresent extreme scores. Doing so results in an extreme departure from the statistical assumption of random sampling.

A problem pervading the studies as a group is their hit-and-run character, deriving from the fact that many have been doctoral research. Studies with minor methodological problems should be repeated with variations to substantiate the results. Doctoral candidates generally prefer to break new ground, hence the scatter-shot pattern.

Where inconsistencies have been evident in thesis data, an attempt has been made to eliminate them by consulting original data or the author; all errors of reporting may not have been identified, however.

In summary, the Sentence Completion Test of Ego Development is adequately validated for research use, but is neither so valid nor so reliable that it can be used as a clinical instrument without confirming data. The construct of ego development has a good deal of support from data, but it lacks definite evidence for sequentiality, and it does not seem that the obstacles to providing conclusive evidence on that point can be readily overcome.

References

- Adams, G. R., & Shea, J. A. The relationship between identity status, locus of control, and ego development. *Journal of Youth and Adolescence*, 1979, 8, 81–89.
- Atkins, S. R. *Experiencing and ego development*. Unpublished doctoral dissertation, University of Chicago, 1976.
- Blasi, A. A developmental approach to responsibility training (Doctoral dissertation, Washington University, 1971). *Dissertation Abstracts International*, 1971, 32, 1233B. (University Microfilms No. 71–19, 807)
- Blasi, A. Personal responsibility and ego development. In R. deCharms, *Enhancing motivation: Change in the classroom*. New York: Irvington, 1976.
- Bonneville, L. P. *The relation of role playing and personal characteristics to ego development*. Unpublished doctoral dissertation, Washington University, 1978.
- Brinkerhoff, R. S. *Ego development in adolescent girls*. Unpublished doctoral dissertation, University of Chicago, 1971.
- Candee, D. Ego developmental aspects of new left ideology. *Journal of Personality and Social Psychology*, 1974, 30, 620–630.
- Christie, R., & Geis, F. L. *Studies in Machiavellianism*. New York: Academic Press, 1970.
- Coor, I. F. The effects of grade level and motivation training on ego development (Doctoral dissertation, Washington University, 1970). *Dissertation Abstracts International*, 1970, 31, 332A. (University Microfilms No. 70–26, 854)
- Cox, L. N. Ego development and helping behavior (Doctoral dissertation, Washington University, 1973). *Dissertation Abstracts International*, 1973, 34, 2764A. (University Microfilms No. 73–24, 880)
- Cox, N. Prior help, ego development, and helping behavior. *Child Development*, 1974, 45, 594–603.
- Cresswell, D. L., & Lacks, P. B. *A comparison of female delinquents with nondelinquents on parental identification, self-concept, ego identity, and morality*. Unpublished manuscript, Washington University, no date.
- Cronbach, L. J., & Meehl, P. E. Construct validity in psychological tests. *Psychological Bulletin*, 1955, 52, 281–302.
- Ebel, R. L. Obtaining and reporting evidence on content validity. *Educational and Psychological Measurement*, 1956, 16, 269–282.
- Erickson, V. L. Psychological growth for women: A cognitive-developmental curriculum intervention. *Counseling and Values*, 1974, 18, 102–116.
- Erickson, V. L. Deliberate psychological education for women: From Iphigenia to Antigone. *Counselor Education and Supervision*, 1975, 14, 297–309.
- Exum, H. A. *Cross-age and peer teaching: A deliberate psychological education curriculum for junior college students*. Unpublished doctoral dissertation, University of Minnesota, 1977.
- Farrell, G. E. P. *The relation of ego development to intellectual and ethical development* (Doctoral dissertation, Washington University, 1974). *Dissertation Abstracts International*, 1975, 35, 4648B. (University Microfilms No. 75–6592)
- Feffer, M., & Jahelka, M. Implications of the decentering concept for the structuring of projective content. *Journal of Consulting and Clinical Psychology*, 1968, 32, 434–441.
- Fischer, M. E. *Manipulativeness and interpersonal style by ego development stages*. Unpublished master's thesis, University of Chicago, 1973.
- Frank, S., & Quinlan, D. M. Ego development and female delinquency: A cognitive-developmental approach. *Journal of Abnormal Psychology*, 1976, 85, 505–510.
- Goldberger, N. *Breaking the educational lockstep: The Simon's Rock experience*. Great Barrington, MA: Simon's Rock Early College, 1977.
- Goldsmith, J. C. *Ego development, Machiavellianism, and interpersonal orientations*. Unpublished master's paper, University of Chicago, 1971.
- Harakal, C. M. Ego maturity and interpersonal style: A multivariate study of Loewinger's theory (Doctoral dissertation, Catholic University, 1971). *Dissertation Abstracts International*, 1971, 32, 1190B. (University Microfilms No. 71–19, 421)
- Hauser, S. T. Loewinger's model and measure of ego development: A critical review. *Psychological Bulletin*, 1976, 83, 928–955.
- Hauser, S. T. Ego development and interpersonal style in adolescence. *Journal of Youth and Adolescence*, 1978, 7, 333–352.

- Heath, D. H. *Growing up in college*. San Francisco: Jossey-Bass, 1968.
- Holt, R. R. *Item manuals for men*. Unpublished manuscript, New York University, undated.
- Hopkins, L. B. *Construction and initial validation of a test of ego identity status for females*. Unpublished doctoral dissertation, Temple University, 1977.
- Hoppe, C. F. Ego development and conformity behaviors (Doctoral dissertation, Washington University, 1972). *Dissertation Abstracts International*, 1972, 33, 6060B. (University Microfilms No. 73-13, 707)
- Hoppe, C. F., & Loevinger, J. Ego development and conformity: A construct validity study of the Washington University Sentence Completion Test. *Journal of Personality Assessment*, 1977, 41, 497-504.
- Hunt, D. E., & Halverson, C. *Manual for scoring sentence completion responses for adolescents*. Unpublished manuscript, Syracuse University, 1964.
- Hurt, B. L. Psychological education for college students: A cognitive developmental curriculum (Doctoral dissertation, University of Minnesota, 1974). *Dissertation Abstracts International*, 1975, 35, 5119A. (University Microfilms No. 75-02, 107)
- Kinnane, J., Lorr, M., & Suziedelis, A. *The influence of interpersonal situations on response modes*. Washington, DC: Catholic University Press, 1969.
- Klein, M. H., Mathieu, P. L., Gendlin, E. T., & Kiesler, D. J. *The Experiencing Scale: A research and training manual* (Vol. 1). Madison: Wisconsin Psychiatric Institute, 1969.
- Kohlberg, L. Development of moral character and moral ideology. In M. L. Hoffman & L. W. Hoffman (Eds.), *Review of child development research* (Vol. 1). New York: Russell Sage Foundation, 1964.
- Kohlberg, L. Stage and sequence: The cognitive-developmental approach to socialization. In D. A. Goslin (Ed.), *Handbook of socialization theory and research*. Chicago: Rand McNally, 1969.
- Kohlberg, L. Continuities and discontinuities in childhood and adult moral development revisited. In P. B. Baltes & K. W. Schaie (Eds.), *Life-span developmental psychology: Personality and socialization*. New York: Academic Press, 1973.
- Kusatsu, O. Ego development and socio-cultural process in Japan. I. *Keizagaku-Kiyo (Journal of Economics)*, 1977, 3, 41-109.
- Lamb, N. J. The relation of role-taking to ego development (Doctoral dissertation, Washington University, 1971). *Dissertation Abstracts International*, 1971, 32, 2401B. (University Microfilms No. 71-27, 331)
- Lambert, H. V. *A comparison of Jane Loevinger's theory of ego development and Lawrence Kohlberg's theory of moral development*. Unpublished doctoral dissertation, University of Chicago, 1972.
- Lasker, H. M. *Interim summative evaluation report: An initial assessment of the Shell/Humanas OD program*. Cambridge, MA: Harvard University, 1977.
- Lasker, H. M. *Ego development and motivation: A cross-cultural cognitive-developmental analysis of an achievement*. Unpublished doctoral dissertation, University of Chicago, 1978.
- Limoges, J. *French translation of the Sentence Completion Test*. Unpublished manuscript, Université de Sherbrooke, Sherbrooke, Quebec, Canada, 1978.
- Loevinger, J. Effect of distortions of measurement on item selection. *Educational and Psychological Measurement*, 1954, 3, 441-448.
- Loevinger, J. Objective tests as instruments of psychological theory. *Psychological Reports*, 1957, 3, 635-694.
- Loevinger, J. Measuring personality patterns of women. *Genetic Psychology Monographs*, 1962, 65, 53-136.
- Loevinger, J. The meaning and measurement of ego development. *American Psychologist*, 1966, 21, 195-206.
- Loevinger, J. *Ego development: Conceptions and theories*. San Francisco: Jossey-Bass, 1976.
- Loevinger, J. *Personal growth in an engineering institute in transition: Interim report*. Unpublished manuscript, Washington University, 1978.
- Loevinger, J. *Scientific ways in the study of ego development*. Worcester, MA: Clark University Press, 1979.
- Loevinger, J. *Sentence Completion for Girls, Form 2-77 and Sentence Completion for Boys, Form 2-77*, undated.
- Loevinger, J., & Nettles, E. J. *Ego development in young single women*. Unpublished manuscript, Washington University, 1963.
- Loevinger, J., & Wessler, R. *Measuring ego development 1. Construction and use of a sentence completion test*. San Francisco: Jossey-Bass, 1970.
- Loevinger, J., Wessler, R., & Redmore, C. *Measuring ego development 2. Scoring manual for women and girls*. San Francisco: Jossey-Bass, 1970.
- Lorr, M., & Youniss, R. *Interpersonal style inventory*. Washington, DC: Catholic University, 1969.
- Love, C. *Item manuals for men*. Unpublished manuscript, undated.
- Luborsky, L. Self-interpretation of the TAT as a clinical technique. *Journal of Personality Assess-*

- ment and Projective Techniques, 1953, 17, 217-223.
- Lucas, R. H. Validation of a test of ego development by means of a standardized interview (Doctoral dissertation, Washington University, 1971). *Dissertation Abstracts International*, 1971, 32, 2204B. (University Microfilms No. 71-27, 335)
- Mikel, E. *Preliminary research studies of character development among imprisoned offenders*. Unpublished manuscript, Washington University, 1974.
- Mosher, R. L., & Sprinthall, N. A. Psychological education: A means to promote personal development during adolescence. *Counseling Psychologist*, 1971, 2, 3-82.
- Perry, W. G., Jr. *Forms of intellectual and ethical development in the college years*. New York: Holt, Rinehart, & Winston, 1970.
- Powitzky, R. J. Ego level and types of federal offenses (Doctoral dissertation, University of Texas Health Science Center at Dallas, 1975). *Dissertation Abstracts International*, 1976, 36, 5768B. (University Microfilms No. 76-11, 838)
- Redmore, C. The relationship of the achievement motive and the affiliation motive to ego development (Doctoral dissertation, Washington University, 1969). *Dissertation Abstracts International*, 1969, 30, 1365B. (University Microfilms No. 69-15, 220)
- Redmore, C. Susceptibility to faking of a sentence completion test of ego development. *Journal of Personality Assessment*, 1976, 40, 607-616.
- Redmore, C. D., & Loevinger, J. Ego development in adolescence: Longitudinal studies. *Journal of Youth and Adolescence*, 1979, 8, 1-20.
- Redmore, C., Loevinger, J., & Tamashiro, R. T. *Measuring ego development: Scoring manual for men and boys*. Unpublished manuscript, Washington University, 1978.
- Redmore, C., & Waldman, K. Reliability of a sentence completion measure of ego development. *Journal of Personality Assessment*, 1975, 39, 236-243.
- Rest, J. R. The hierarchical nature of stages of moral judgment. *Journal of Personality*, 1973, 41, 86-109.
- Rest, J., Turiel, E., & Kohlberg, L. Level of moral judgment as a determinant of preference and comprehension of moral judgments made by others. *Journal of Personality*, 1969, 37, 225-252.
- Rock, M. Self-reflection and ego development (Doctoral dissertation, New York University, 1975). *Dissertation Abstracts International*, 1975, 36, 3066B. (University Microfilms No. 75-28, 583)
- Schain, W.S. Psychodynamic factors affecting women's occupational choice: Parent-child relations, expressed needs and level of ego development (Doctoral dissertation, George Washington University, 1974). *Dissertation Abstracts International*, 1974, 35, 1991A. (University Microfilms No. 74-23, 499)
- Schenberg, R. G. The relation of time perspective and self-actualization to ego development (Doctoral dissertation, Washington University, 1973). *Dissertation Abstracts International*, 1974, 34, 4641B. (University Microfilms No. 74-07, 064)
- Sechrest, L. Incremental validity: A recommendation. *Educational and Psychological Measurement*, 1963, 23, 153-158.
- Sheridan, S. J. W. Level of moral reasoning and ego development as factors in predicted vocational success with the mentally retarded (Doctoral dissertation, University of Houston, 1974). *Dissertation Abstracts International*, 1975, 35, 4290A. (University Microfilms No. 75-1023)
- Shumate, W. L., Jr. The relation of authoritarianism and locus of control to ego development (Doctoral dissertation, Washington University, 1969). *Dissertation Abstracts International*, 1970, 30, 3395B. (University Microfilms No. 69-22, 559)
- Snarey, J. R., & Blasi, J. *Ego development among kibbutz founders and sabras*. Cambridge, MA: Institute for Cooperative Community, 1978.
- Strodtbeck, F. L., Ronchi, D., & Hansell, S. Tutoring and psychological growth. In V. L. Allen (Ed.), *Children as teachers: Theory and research on tutoring*. New York: Academic Press, 1976.
- Sullivan, E. V., McCullough, G., & Stager, M. A developmental study of the relationship between conceptual, ego, and moral development. *Child Development*, 1970, 41, 399-411.
- Sullivan, P. J. A curriculum for stimulating moral reasoning and ego development in adolescents (Doctoral dissertation, Boston University, 1975). *Dissertation Abstracts International*, 1975, 36, 1320A. (University Microfilms No. 75-20, 971)
- Turiel, E. An experimental test of the sequentiality of developmental stages in the child's moral development. *Journal of Personality and Social Psychology*, 1966, 3, 611-618.
- Vetter, M. *Dimensionen des Selbstkonzeptes und Ich-Entwicklung*. Unpublished master's thesis, Johannes-Gutenberg Universität, Mainz, Germany, 1978.
- Zielinski, C. E. Stage of ego development as a correlate of ability in discrimination and communication of empathic understanding (Doctoral dissertation, University of Houston, 1973). *Dissertation Abstracts International*, 1973, 34, 1635A. (University Microfilms No. 73-22, 941)

Acknowledgments

Preparation of this paper was supported by Research Scientist Award MH-00657 from the National Institute of Mental Health, Public Health Service. Some of the work reported here was supported by Research Grant MH-05115 from the National Institute

of Mental Health, Public Health Service, and by a grant from the Spencer Foundation.

Author's Address

Send requests for reprints or further information to Jane Loevinger, Social Science Institute, Washington University, St. Louis, MO 63130.