Self-Concept: Validation of Construct Interpretations

Richard J. Shavelson
Judith J. Hubner
George C. Stanton

Stanford University

Historically, education goals have tended to fluctuate from emphasis solely on cognitive outcomes to major concern with social and affective ones. The emphasis on achievement and the "cult of efficiency" (Callahan, 1962) early in this century was followed by a shift in the 1930's to the comprehensive high school with its social and affective concerns (cf. the Eight Year Study, Aikin, 1942). Then Sputnik initiated a rapid and dramatic reemphasis on cognitive outcomes (Bruner, 1960) from which the current trend seems to be moving in its emphasis on "humanistic" aspects of education.

The sharp increase in the number of studies on self-concept is one reflection of the reemphasis on noncognitive outcomes of education. (For references to current educational studies, see reviews by Coller, 1971; Purkey, 1970; Yamamoto, 1972; Zirkel, 1971.) Another symptom of this shift has taken the form of increased concern with enhancing the child's self-concept, espe-

The authors wish to express their appreciation to Professors Lee J. Cronbach and Nathaniel L. Gage for their valuable, critical comments on earlier drafts of this paper.

The research reported herein was conducted at the Stanford Center for Research and Development in Teaching, which is supported in part by the National Institute of Education, Department of Health, Education, and Welfare. The opinions expressed in this publication do not necessarily reflect the position, policy, or endorsement of the National Institute of Education. (Contract No. NE-C-00-3-0061.)
cially on the part of Head Start teachers (e.g., Hoepfner, Stern, & Nummedal, 1971). According to Zirkel (1971, p. 211),

It has become increasingly clear in the light of the schools' attempt to serve the disadvantaged that the schools have a fundamental responsibility to enhance the self-concepts of their students [Clark, 1963; Marston, 1968; Tannenbaum, 1967].

This objective has been prescribed and described for virtually all programs for the disadvantaged (Fantini & Weinstein, 1968; Smiley, 1967; Gordon & Wilkerson, Note 1). Thus, improvement of a student’s self-concept seems to be valued as an educational outcome in its own right.

But even if self-concept were not so valued, the construct has potential scientific importance for interpreting achievement outcomes. Most definitions link this construct to achievement, and there is some empirical evidence to support this theoretical linkage (Brookover, Le Pere, Hamachek, Thomas, & Erickson, Note 2; Torshen, Note 3). Self-concept, then, whether used as an outcome itself or as a moderator variable that helps explain achievement outcomes, is a critical variable in education and in educational evaluation and research.

Most self-concept studies examine correlations between a measure of self-concept and measures of other constructs (e.g., Coopersmith, 1967; Sears, Adenubi, Bloch, Hubner, Gamble, & Crist, 1972; Brookover et al., Note 2; Torshen, Note 3; Bledsoe & Garrison, Note 4; Sears, Note 5); differences in mean self-concept scores among different populations of students (e.g., Hishiki, 1969; Soares & Soares, 1969; Zirkel, 1971); and changes in self-concept attributable to some treatment (e.g., Herbart, Gelfand, & Hartman, 1969; Long, Ziller, & Henderson, 1968; Ludwig & Maehr, 1967; Zirkel, 1971, 1972). Taken individually, they often provide important insights into the factors that motivate students in and out of school and into alternative courses of action that may enhance students' self-concepts (e.g., Purkey, 1970; Sears et al., 1972; Yamamoto, 1972; Spaulding, Note 6).

Yet, considered as a body of research, self-concept studies today may be criticized in the same way as Crowne and Stephens (1961) and Wylie (1961) did almost twelve years ago, in that the self-concept interpretations of their measurements may not be valid. First, definitions of self-concept are imprecise and vary from one study to the next. The imprecision makes it extremely difficult to specify (a) the population of self-concept items from which a representative sample would be drawn for the instrument, or (b) the population of subjects for which the measurement techniques and interpretations would be appropriate. And,
as for variability in definitions, a review of such definitions (e.g., Brownfain, 1952; Bruner, 1958; Combs & Soper, 1957; Coopersmith, 1967; Hamachek, 1965; James, 1963; Jersild, 1952; McDonald, 1965; Mischel, 1968; Mote, 1967; Piers & Harris, 1964; Rogers & Dymond, 1954; Sears & Sherman, 1964; Sherif & Cantril, 1947; Snygg & Combs, 1949; Sullivan, 1953; Symonds, 1951; Wylie, 1961; Brookover et al., Note 2; Bledsoe & Garrison, Note 4; Gordon, Note 7; Purkey, Note 8) revealed seventeen different conceptual dimensions on which they could be classified. Some of these dimensions were (a) emphasis on a stable or changing self-concept; (b) methods for changing self-concept—learning/reinforcement, creation of dissonance, or arousal of needs and defenses; (c) determinants of self-concept—situational, phenomenal, or internal; (d) types of evaluation—normative standard, absolute personal standard, or nonevaluative; and (e) dimensionality of self-structure—unidimensional or multidimensional.

A second difficulty in interpreting measures of self-concept arises because data are not readily available on the equivalence of various self-concept measurement instruments. In many cases, researchers develop their own instruments for their own particular problem so that the number of different measurement techniques is increasing almost as rapidly as the number of self-concept studies. Given the imprecision and variability in definitions, we have no reason to believe, a priori, that the measurement techniques are equivalent. The lack of empirically demonstrated equivalence among self-concept measurements makes it impossible to generalize across studies using different instruments. And, for a given instrument, data suggest that generalization of construct interpretations across different populations of subjects may be hazardous (Dyer, 1964; Gordon, Note 7; see also Zirkel's 1971 review of self-concept studies of culturally different students).

Finally, data are not available to test rival counter-interpretations. For example, as with any self-report measure of a personality variable, the self-concept interpretation may be challenged on the grounds that students may: (a) select responses they know to be socially desirable rather than responses that are self-descriptive (cf. Edwards, 1957a), or (b) be unable (cf. Snygg & Combs, 1949) or unwilling (cf. Cronbach, 1970) to report their "private" self-concepts. Crowne and Stephens' (1961) conclusion on this matter is consistent with ours:

While studies of the effect of the social desirability variable on many of the commonly employed tests of self-acceptance have not been done, the results of some ... investigations ... would suggest that self-evaluative tests
are particularly susceptible to criticism on social desirability grounds. A common denominator in research findings on self-acceptance may well be the variable of social desirability. (p. 117)

In summary, then, it appears that self-concept research has addressed itself to substantive problems before problems of definition, measurement, and interpretation have been resolved. Until these problems have been dealt with in a manner made possible by advances in construct validation methodology, the generalizability of self-concept findings will be severely limited, and data on students' self-concepts will continue to be ambiguous.

The importance of the construct and the paucity of studies that examine self-concept interpretations of their measurements make construct validation research vitally needed. Unlike previous methodological critiques (e.g., Crowne & Stephens, 1961; Gordon, C., 1969; Wylie, 1961), our approach is constructive in that we (a) develop a definition of self-concept from existing definitions, (b) review some steps in validating a construct interpretation of a test score, and (c) apply these steps in examining five popularly used self-concept instruments. Perhaps this approach will stimulate further construct validation research.

**Definition of Self-Concept**

Although construct validation can proceed with an informal, intuitive definition, a mature construct definition should be formal and explicit. The ideal would be to define the construct with a network of associations or propositions that relate the construct to: (a) observable properties or quantities of the construct (the within-construct portion of the construct definition; cf. Loevinger's, 1957, "structural component"), and (b) other constructs (the between-construct portion of the construct definition; cf. Loevinger's, 1957, "external component") which are, themselves, related to observables. This network of interrelationships, called a *nomological network* (cf. Cronbach & Meehl, 1955), locates a construct in a conceptual space. The within-portion of the definition specifies the features of the construct and links them to each other and to observable attributes of the person. The between-portion locates the construct in a conceptual space that includes many other constructs related to or independent of the construct. For example, many definitions of self-concept include a multifaceted feature (cf. Coopersmith, 1967; Piers & Harris, 1964; Purkey, 1970; Sears & Sherman, 1964; Brookover, Erickson, & Joiner, Note 9; Gordon, Note 10). The within-construct portion of a definition may identify academic,
social, and physical self-concept facets and their interrelations. The between-construct portion may relate each of these facets to other constructs. Thus, academic self-concept may be more closely related to achievement than physical self-concept.

Currently, in spite of the imprecision and variability of the informal and intuitive self-concept definitions, many self-concept definitions overlap in various ways. By integrating various features that are common to the definitions, and by extending the definitions when necessary, it is possible to construct a working definition of self-concept that is consistent with some current research and can be used to begin to integrate empirical evidence on the validity of self-concept interpretations.

In very broad terms, self-concept is a person’s perception of himself. These perceptions are formed through his experience with his environment, perhaps in the manner suggested by Kelly (1973), and are influenced especially by environmental reinforcements and significant others. We do not claim an entity within a person called “self-concept.” Rather, we claim that the construct is potentially important and useful in explaining and predicting how one acts. One’s perceptions of himself are thought to influence the ways in which he acts, and his acts in turn influence the ways in which he perceives himself. The exact nature and direction of the influence of perceptions and behavior are important parts of the definition, but as yet are unclear and consequently are an important focus of current self-concept studies.

Self-concept is inferred from a person’s responses to situations. The situations and the responses may be physical or symbolic. Since self-concept must be inferred, issues arise as to what are considered admissible observations. In most educational examinations of self-concept, a distinction is made between self-concept and inferred self-concept. Self-concept is restricted to a person’s report of self (cf. Combs, Soper, & Courson, 1963; Parker, 1966). Inferred self-concept is another’s attribution of a person’s self-concept. For the purposes of this paper, we will maintain the distinction between self-concept and inferred self-concept, and focus on the former.

Seven features can be identified as critical to the construct definition. Self-concept may be described as: organized, multifaceted, hierarchical, stable, developmental, evaluative, differentiable. Each of these features is considered below.

An individual’s experiences, in all their great diversity, constitute the data on which he bases his perceptions of himself. To reduce the complexity of these experiences, a person recodes them into simpler forms, or categories (Bruner, 1958). The particular category systems adopted by an individual are to some
extent a reflection of his particular culture. For example, a child's experience may revolve around his family, friends, and school. This may account for these categories in children's descriptive statements about themselves (Jersild, 1952; Sears, 1963). The categories represent a way of organizing experiences and giving them meaning. One feature of self-concept, then, is that it is organized or structured.

A second feature of self-concept is that it is multifaceted; the particular facets reflect the category system adopted by a particular individual and/or shared by groups. At least in the white, middle-class population of students studied by Jersild (1952) and Sears (Note 5) the category system appears to include such areas as the school, social acceptance, physical attractiveness, and ability.

A third feature is that the multifaceted structure of self-concept may be hierarchical on a dimension of generality (Super, 1963; Brookover et al., Note 9). That is, facets of self-concept may form a hierarchy from individual experiences in particular situations at the base of the hierarchy to general self-concept at the apex. One possible representation of this hierarchy is shown in Figure 1. This formulation is in some ways similar to the British psychologists' hierarchical model of intellectual abilities (cf. Vernon, 1950). At the apex of the hierarchy is general self-concept (cf. Spearman's "g"). General self-concept may be divided into two components: academic self-concept and nonacademic self-concept (cf. verbal-educational and practical abilities in the Vernon model). Academic self-concept may be divided into subject-matter areas (cf. specific group factors in the Vernon model) and then into specific areas within a subject matter (cf. specific factors). Nonacademic self-concept may be divided into social and physical self-concepts and then divided into more specific facets in a manner similar to academic self-concept.

If this line of reasoning is pursued to the base of the hierarchy, a conceptualization of self-concept as situation-specific is consistent with our definition. In extremely limited situations (such as those represented by laboratory experiments), alternative interpretations of a person's experience are reduced considerably. At this level, then, an observer's perception of a person's self-concept may correspond with the person's report of his self-concept. Nevertheless, the distinction between self-concept and inferred self-concept is important. The correspondence between observer and self decreases as one moves up the self-concept hierarchy.

A fourth feature of self-concept is that general self-concept is stable. However, as one descends the self-concept hierarchy (Figure 1), self-concept depends increasingly on specific situa-
Figure 1
One possible representation of the hierarchic organization of self-concept.
tions and thus becomes less stable. At the base of the hierarchy, self-concept varies greatly with variation in situations. Furthermore, changes at the lower levels of the hierarchy are probably attenuated by conceptualizations at higher levels, making self-concept resistant to change (cf. Ludwig & Maehr, 1967; see Bem, 1972, for a similar notion from a different perspective). To change general self-concept, many situation-specific instances, inconsistent with general self-concept, would be required. For example, Ludwig and Maehr (1967) showed that success and failure in an athletic task changed subjects' self-concepts of specific physical ability but did not change their general self-concepts.

A fifth feature of self-concept is its developmental aspect (cf. Engle, 1959; Long, Henderson, & Ziller, 1967; Long et al., 1968; Sears, 1964). Infants tend not to differentiate themselves from their environment. As they mature and learn from their increasing store of experiences, differentiation of self from environment begins. The self-concepts of young children are global, undifferentiated, and situation-specific. As children begin to build concepts, as represented by the words I and me, they also begin to build concepts for categorizing events and situations. Young children have not started to coordinate the separate subparts of experience to integrate them within one conceptual self-framework. At different times during development, it seems likely that "as the child grows, different parts of him will become more important to him and different parts of his world will assume changing significance" (Gordon, Note 7, p. 4). With increasing age and experience (especially acquisition of verbal labels), self-concept becomes increasingly differentiated. As the child coordinates and integrates the parts of his self-concept, we can speak of a multifaceted, structured self-concept.

A sixth feature of self-concept is its evaluative character. Not only does the individual develop a description of himself in a particular situation or class of situations, he also forms evaluations of himself in these situations. Evaluations can be made against absolute standards, such as the "ideal," and they can be made against relative standards such as "peers" or perceived evaluations of "significant others." The evaluative dimension can vary in importance for different individuals and also for different situations. This differential weighting of the importance of the various evaluative dimensions probably depends upon the individual's past experience in a particular culture, in a particular society, and so on. As far as we know, the distinction between self-description and self-evaluation has not been clarified either conceptually or empirically. Accordingly, the
terms *self-concept* and *self-esteem* have been used interchangeably in the literature.

A seventh feature of self-concept is that it is *differentiable* from the other constructs with which it is theoretically related. It is beyond the scope of this paper to examine systematically the relationships between self-concept and other constructs, i.e., the between-constructs portion of the nomological network. Nevertheless, it is possible to indicate the direction one could take in specifying how self-concept is differentiable from, and related to, other constructs. For example, self-concept is influenced by specific experiences. Therefore, the more closely self-concept is linked with specific situations, the closer is the relationship between self-concept and behavior in the situation. If one were to focus on the academic side of the hierarchy in Figure 1, one could hypothesize that (a) self-concept of mental ability should be more closely related to academic achievement than to ability in social and physical situations, and (b) self-concept of academic ability in science should be more closely related to achievement in science than to achievement in, say, English or overall grade-point-average (cf. Brookover, Paterson, & Thomas, Note 10).

In a similar manner, the other side of the hierarchy can be explored and relationships between self-concept and other constructs explicated.

**Validation of Self-Concept Interpretations of Test Scores: Methodological Considerations**

Validating a construct interpretation of a test score involves an interplay of construct definition, instrument development, and data collection. In this context, the construct definition sets the boundaries for potential measurement techniques. It operates like a test plan for the development of an instrument. It specifies content areas (e.g., academic, social, and physical self-concept), the type of stimuli (e.g., items referring to self), the observer (e.g., self as observer of self versus “other” as observer of self), the response (e.g., comparison with others or comparison with an absolute criterion), and so on. From this plan, an instrument is developed and data are collected. These data bear on the construct interpretation of the scores.

Interpretations of these data may be considered hypotheses to be challenged over and over with counterhypotheses (Cronbach, 1971; Cronbach & Meehl, 1955). Initial construct validation studies should examine the empirical and logical evidence in support of the within-construct portion of the nomological network; e.g., whether or not measurements of facets such as
academic, social, and physical self-concept warrant separate interpretations. Later studies should examine evidence in support of the between-constructs portion of the nomological network; whether the self-concept facet measurements warrant the interpretations that this construct differs from other constructs, such as intelligence, locus of control, anxiety, and social desirability.

If the empirical evidence is congruent with the construct definitions, test scores are given construct interpretations. If the data are incongruent with the definition, the definition or the instrument, or both, require revision. Note that empirical evidence cannot directly void the construct definition. Rather, it can reflect upon the measurement technique, which, in turn, can reflect upon the definition. If subsequent instrument revisions continue to produce empirical evidence incongruent with the definition, it may be that certain aspects of the construct cannot be measured with existing techniques. In this case, the nature of the warranted interpretations should be specified, and the construct definition should be subjected to critical, logical analyses.

Procedures for investigating rival interpretations fall into three categories: logical, correlational, and experimental.

**Logical Analyses**

The logical analysis examines the consistency between the construct definition and instructions to subjects, instrument format, item content, and scoring procedures. It draws upon the investigator's past experience and upon psychometric considerations.

The function of the logical analysis is to generate counterhypotheses as to the construct interpretation of the test score. For example, Sears and Sherman define self-concept as composed of ten traits (see Sears and Sherman, 1964, Table 5, p. 47); the Sears Self-Concept Inventory (Sears, Note 5) contains items purported to measure each of these traits. A logical analysis of the inventory might lead to a counterhypothesis that items linked, say, to the facets of work habits, school subjects, mental ability, and social relations with teacher do not warrant separate interpretations, but relate rather to a single trait, academic self-concept (cf. Torshen, Note 3).

But it should be noted that "... the logical analysis of content cannot disprove a validity claim. The analysis puts forth a counterhypothesis whose pertinence can be verified only empirically" (Cronbach, 1971, p. 475).
Correlational Techniques

Intercorrelations between facets of a construct—such as correlations between academic, social, and physical self-concept measures—provide evidence on whether the facets deserve to be interpreted separately. Intercorrelations between measures of one construct and other different constructs—such as the correlations between academic and social self-concept and intelligence—provide evidence on whether scores on a construct warrant the interpretation that it is separate from other constructs. In a similar manner, correlations may be used to examine other features of the construct definition such as its stability, developmental character, and hierarchical organization.

A variation on the correlational approach is to identify two populations expected to differ on the construct in question and to determine whether their scores on the construct measure differ (cf. Piers & Harris, 1964; Trowbridge, 1972; Zirkel, 1972). For example, Piers and Harris (1964) compared self-concept scores of public school children and adolescent, institutionalized, retarded ($\overline{X}_{iq} = 69.6$) females. They found, as expected, that public school children earned significantly higher self-concept scores.

Three correlational techniques can be useful in deciding how to interpret test scores. The first, factor analysis, arranges a matrix of correlations into convergences or clusters among tests or among items on a test. If the test operates as its design suggests, items measuring, say, “academic self-concept” should cluster together, and this cluster should be distinct from a cluster of items on, say, “physical self-concept.” When this occurs, we gain some confidence in our facet interpretation of test scores. If unanticipated clusters are found in the test, some revision is called for. In some cases, factor analysis has been used to validate self-concept interpretations of subtest scores (e.g., Gordon, I. J., 1966; Piers & Harris, 1964; Torshen, Note 3; Sears, Note 5) and in some cases it has not (e.g., Coopersmith, 1967).

A second correlational method, the multitrait-multimethod matrix (Campbell & Fiske, 1959), examines patterns of intercorrelations between different traits (e.g., academic, social, and physical self-concept) measured by maximally different methods (say, self-report and peer-report of a student’s self-concept; e.g., Bixler, 1965; Trickett, 1969). If, for example, factor analysis demonstrates that, for one instrument, items group into certain self-concept facets (say, academic self-concept) that are distinct from others (say, physical self-concept), this distinction should be maintained when different methods are used to measure the same traits. A multitrait-multimethod matrix is constructed from
correlations between (a) scores on different traits obtained by the same measurement method, (b) scores on the same trait obtained by different measurement methods, and (c) scores on different traits obtained by different measurement methods” (Campbell & Fiske, 1959, p. 81). (The distinction between factor analysis and the multitrait-multimethod matrix is made for the sake of clarity. For the relation of the multitrait-multimethod matrix to factor analysis and the analysis of variance, see Boruch, Larkin, Wolins, & MacKinney, 1970; Boruch & Wolins, 1970.) In this matrix, “reliability is the agreement between two efforts to measure the same trait through maximally similar methods. Validity is represented in the agreement between two attempts to measure the same trait through maximally different methods” (Campbell & Fiske, 1959, p. 83). Traits are isolated when the following convergent and discriminant validity criteria are satisfied (refer to Table 1; for a critique of these criteria, see Althauser & Heberlein, 1971):

1. Convergent criterion
   a. A validity coefficient should be significantly greater than zero and of practical significance (validity diagonal entries)

2. Discriminant criteria
   a. A validity coefficient should be higher than the correlations obtained between that variable and any other variable having neither trait nor method in common (heterotrait-heteromethod entries)
   b. A validity coefficient should be higher than the correlations among scores on different traits obtained by the same measurement method (heterotrait-monomethod entries)
   c. The same pattern of interrelations among traits should be observed in correlations obtained with the same or different methods

To our knowledge, the multitrait-multimethod matrix has been used to validate self-concept facet interpretations of test scores in only the studies by Dyer (1964), Gordon (1966), and Shulman (Note 11).

So far, the discussion of two correlational techniques has focused primarily on the examination of the within-construct portion of the nomological network. These techniques are applicable to the between-construct portion as well. For example, the multitrait-multimethod matrix has been used to examine the interpretation of a test score as measuring “humor” (Koppel & Sechrest, 1965). Rival hypotheses to this interpretation were that the score measured intelligence or extroversion. Three different
measurement methods (self-ratings, peer ratings, and objective responses) were used to measure humor appreciation, humor creation, intelligence, and extroversion. The results were used to determine the degree to which humor, intelligence, and extroversion could be distinguished as constructs.

Consider the case in which one construct is measured by methods A and B, and a second construct is measured by methods A and C. For example, some theorists argue that self-concept can be measured only by self-report methods (on this issue, see Combs, Soper, & Courson, 1963; Parker, 1966; Sears & Sherman, 1964; Wylie, 1961). But some other constructs, say anxiety, might be measured by self-report, observation ratings, heart rate, and so on. To examine validity claims here, the multitrait-multimethod matrix may not be applicable, but factor...
analysis is. Different measures of the same trait should cluster and perhaps two or more “measurement method” clusters might be found.

Finally, if the network specifies a causal relationship, other correlational techniques such as path analysis might be used to examine causality (see, for example, Blalock, 1964; Crano, Kenny, & Campbell, 1972; Yee & Gage, 1968). Bixler (1965) attempted to use the cross-lagged panel analysis to examine causal effects of teachers’ and peers’ influence on changes in students’ self-concepts. The data suggest that neither teachers’ nor peers’ influence caused changes in students’ self-concepts. (See also Wattenberg & Clifford, 1963; Torshen, Note 3; for attempts to examine causality in correlation studies of self-concept.)

Experimental Techniques

Experiments can test some portion of a nomological network against a counterhypothesis. Such experiments can be used to identify influences to which test scores are sensitive. Although most self-concept experiments are not designed to test counterhypotheses as to the proposed self-concept interpretation of their measurements, the outcomes often bear on construct validity. Thus, experimental studies in which specific treatments are used to change only one aspect of self-concept, say physical self-concept, provide an indirect test of the within-construct portion of the nomological network. For example, Ludwig and Maehr (1967) examined the effect of success and failure in athletic tasks on physical and general self-concept. Subjects were randomly assigned to positive or negative feedback groups or a control group, regardless of athletic ability. In the feedback groups, feedback was either consistent or inconsistent with the subject’s ability. If Ludwig and Maehr’s instruments measured self-concept as specified in their nomological network, scores on measures of physical self-concept would be expected to increase with positive feedback, to decrease with negative feedback that was consistent with ability, to increase with positive, inconsistent information, to decrease initially with negative inconsistent information, and to remain unchanged in the control group. A similar but less distinct pattern of scores would be expected on the general self-concept measure. In general, the results of the study were consistent with these expectations.

Experimental studies in which treatments are designed to change subjects’ test scores examine the construct interpretation against counter hypotheses (Cronbach, 1971). For example, do self-concept scores depend upon the subject’s motivation, upon knowledge of socially desirable responses, or upon strategy
for attacking the task? Parker (1966) examined the influence of students' expectations as to who would see their self-concept test scores on self-concept measurements. Data were collected with a self-report and an inferred self-concept test, first with the expectation of anonymity and then with the expectation that the teacher would see the scores. The expectancy variable did not influence mean scores on the self-report or inferred self-concept measures. But it did influence correlations between self-report and inferred self-concept measurements.

There are, then, a number of techniques—logical, correlational, experimental—for examining the validity of self-concept interpretations of test scores. (Our discussion has used the multifaceted feature of self-concept as an example, but other features may be examined as well.) Each contributes its own kind of evidence to these interpretations. Most self-concept measurements have not been subjected to such examinations of their validity claims. In the following sections, we examine the validity of five widely used instruments. The discussion focuses on definitions and on the validity of the interpretations of the self-concept features (the within-construct portion of the nomological network). This focus seems desirable because, until we are confident about definitions and within-construct interpretations, it is difficult to interpret evidence bearing on the relation of self-concept to other constructs (the between-constructs portion).

Review of Five Major Self-Concept Instruments

Three aspects of construct validity—definition, instrumentation, empirical data—are examined for five commonly used self-concept instruments: (a) the Michigan State Self-Concept of Ability Scale, (b) the Self-Esteem Inventory, (c) the How I See Myself Scale, (d) the Piers-Harris Children's Self-Concept Scale, and (e) the Self-Concept Inventory. These instruments were selected for several reasons. All authors presented some data bearing on construct interpretations of scores from their self-concept instruments. All instruments are being used in educational research, evaluation, or both. All relate, then, in some way to students' educational experiences. All are representative of self-report measures of self-concept.

Michigan State Self-Concept of Ability Scale

The Michigan State Self-Concept of Ability Scale (SCA), a measure of self-concept of academic ability, has been used in research by Brookover (Bilby, Brookover, & Erickson, 1972; Brookover et al., Note 2; Brookover, Erickson, & Joiner, Note 7;
Brookover, Paterson, & Thomas, Note 10) to examine relationships between SCA and measures of antecedent and consequent variables, such as intelligence, school achievement, and perceptions of the evaluations of significant others.

**Construct Definition.** "Self-concept of academic ability refers to behavior in which one indicates to himself [publicly or privately] his ability to achieve in academic tasks as compared with others engaged in the same task [Note 7, p. 8]." Self-concept of ability can be characterized as one of the many concepts of self, hierarchical in nature (academic portion of Figure 1), dependent on a particular role, area of experience, situation, time, etc., relatively stable under similar stimulus situations, and primarily evaluative in nature.

**Description of the Instrument.** The SCA-General Scale (Form A) consists of eight items, selected to differentiate students on achievement (Patterson, Note 12), which form a Guttman scale (reproducibility coefficient = 0.91). The items are self-evaluative questions about academic ability such as "What kind of grades do you think you are capable of getting in the following subjects?" For each of four subjects (mathematics, English, social studies, science), there are five response alternatives ranging from "Mostly A's" to "Mostly F's." The eight items have been divided into two conceptual dimensions, each composed of two logical subsets (no examples were provided): (a) future-oriented and present-oriented items, and (b) comparative and absolute evaluations of ability.

The SCA-Specific Subjects Scale (Form B) was constructed for the same eight items with separate responses requested for science, mathematics, English, and social studies. The mathematics scale had a reproducibility coefficient of 0.90; the other scales were not analyzed (Paterson, Note 12).

**Reliability.** For samples of subjects drawn from grades seven to ten, internal consistency reliability of Form A ranged from 0.82 to 0.92 for boys and from 0.77 to 0.84 for girls (Brookover et al., Note 9; Brookover et al., Note 10). Form B internal consistency coefficients on a tenth-grade sample were above 0.89 for boys and girls (Brookover et al., Note 4).

Test-retest reliability coefficients, reported for one-year intervals, ranged from 0.69 to 0.72 for boys and 0.69 to 0.77 for girls in grade levels 8 to 12 (Brookover, et al., Note 4; Brookover, et al., Note 9). For the subject-matter areas of Form B, they ranged from 0.70 to 0.80 for boys and from 0.63 to 0.77 for girls (Brookover, et al., Note 4).

These data, then, suggest a dependable rank ordering of persons at one point in time which remains fairly stable over one
year. In general, the reliabilities are high enough to permit study of individual differences.

Empirical Evidence in Support of Construct Interpretation. Several analyses have examined the structure or dimensionality of the SCA-Form A. Scalogram analyses have yielded coefficients of reproducibility ranging from 0.86 to 0.97. A factor analysis (Paterson, Note 12) revealed two factors. The first factor was interpreted as self-concept of ability; the second factor, barely distinguishable from error, was interpreted as a time dimension differentiating between four future-oriented and three present-oriented items. These findings, then, support the interpretation of SCA as a unidimensional measure perhaps of self-concept of ability.

Studies by Brookover et al. (Note 2, Note 10) provide data which bear on whether scores on Forms A and B warrant separate interpretation. Correlations between Forms A and B ranged from 0.62 to 0.73 for boys and 0.54 to 0.71 for girls. When these correlations were corrected for attenuation, they ranged from middle 0.70 to middle 0.90. These correlations along with Paterson’s (Note 12, p. 163) report that “The general SCA mean scores are higher than any of the specific SCA scores but the general score is the closest to the specific score in the subject in which the student had his highest achievement” suggest that differential interpretation of Forms A and B should be made quite cautiously.

Data reported by Brookover et al. (Note 4) and Paterson (Note 12) bear on whether separate interpretations of the four subject-matter scales (SCA-Form B) are warranted. For boys, the correlation between scales ranged from 0.63 to 0.88; for girls, the range was 0.52 to 0.68. These correlations, when corrected for attenuation, ranged from the mid-60s to the mid-80s. Even though these intercorrelations are fairly high, differential predictions of grades in corresponding subject matters from the scales is possible. In all but one instance, the correlations between grades and scores on corresponding areas of the Form B scale (ranging from 0.43 to 0.61) were higher than the correlations between grades and scores in different areas.

Interpretation of Scores from the SCA. Scores on the general SCA (Form A) probably warrant an interpretation separate from scores on the subject-matter specific scales (Form B), if made quite cautiously. Likewise, cautious interpretation of the different subject-matter scales seems warranted. However, the high correlations between Form A and the scales of Form B, the high correlations between the scales of Form B, and data on means reported by Paterson (Note 12) suggest that additional validation
work, such as a multitrait-multimethod analysis, is needed to clarify interpretations. Finally, the issue needs to be raised as to whether the SCA is a measure of self-concept, an alternate method of reporting grades, or an alternate method of reporting general mental ability. We suspect the self-concept interpretation will be shown to be warranted, but systematic examination of these and other counterhypotheses is needed.

In regard to our definition of self-concept, the data reported on the SCA provides some support to a multifaceted, hierarchical view of self-concept. Support for a multifaceted self-concept comes from Brookover's definition of academic self-concept as one of many self-concepts and the empirical evidence—internal consistency and reproducibility coefficients, factor analysis—suggesting that the SCA is a unidimensional measure. Thus, one of many possible facets of self-concept may be tapped by the SCA.

Support for a hierarchical view of self-concept comes from data on the relation of scores on SCA Forms A and B. The correlations suggest a strong general feature of academic self-concept that can be differentiated into subject-matter specific self-concepts (see Figure 1).

Self-Esteem Inventory

The Self-Esteem Inventory (SEI: Coopersmith, 1967) has been used as a measure of general self-concept (e.g., Coopersmith, 1967; Dyer, 1964; Smith, I. J., 1973; Epstein & Komorita, 1971). Most of Coopersmith's (1959, 1967) own empirical work examined relationships between SEI scores and measures of other constructs, such as popularity, role in group discussions, conformity and creativity.

Construct Definition. Coopersmith (1967) defined self-esteem as

... the evaluation which the individual makes and customarily maintains with regard to himself: it expresses an attitude of approval or disapproval, and indicates the extent to which the individual believes himself to be capable, significant, successful, and worthy. In short, self-esteem is a personal judgment of worthiness that is expressed in the attitudes the individual holds toward himself. It is a subjective experience which the individual conveys to others by verbal reports and other overt expressive behavior. (pp. 4-5)

Coopersmith elaborated especially on three features of this definition: stability, structure, and process. Stability "centers upon the relatively enduring estimate of general self-esteem
rather than upon more specific and transitory changes in evaluation" (p. 5). Structure focuses on general self-esteem with the acknowledgement that self-esteem may be multifaceted with regard to different areas of experience and according to sex, age, and other role-defining conditions. Process focuses on "a judgmental process in which the individual examines his performance, capabilities, and attributes according to his personal standards and values, and arrives at a decision (evaluation) of his own worthiness."

**Description of the Instrument.** The SEI was developed from a pool of original and reworded items from a scale reported by Rogers and Dymond (1954). Form A contains 58 items; eight items form a "lie" scale, and 50 items form the self-esteem measure. The items are simple, self-descriptive statements such as "I give in very easily." There are two response alternatives for each item: "Like me" or "Unlike me." In addition to a total self-esteem score, four subscale scores have been identified conceptually: General Self (26 items), Social Self-Peers (8 items), Home Parents (8 items), School Academic (8 items).

Form B includes "those 25 items which showed the highest item-total score relationships (Instructions, undated, p. 1)" in an item analysis of Form A. Total scores of Forms A and B correlated 0.86. This level of correlation has been approximated with four different samples (Coopersmith, 1967). Since this was a part-whole correlation, it is spuriously high.

**Reliability.** Internal consistency reliability for the four subscales of the SEI (Dyer, 1964) ranged from 0.28 to 0.82 for boys and girls at third, fifth, seventh, ninth, and eleventh grades. The total score test-retest reliability coefficient was 0.88 over a five-week interval with a sample of 30 fifth-grade students, and a coefficient of 0.70 was obtained over a three-year period with a sample of 56 public school students. These data suggest that a dependable measure can be obtained with total scores but, by using fewer items on subscales, the consistency of the measurement is greatly reduced. Although individual differences may be examined with total scores, some of the subscales appear too unreliable for such work.

**Empirical Evidence in Support of Construct Interpretations.** Although *a priori* subscales for the SEI have been proposed, Coopersmith has not examined the validity of separate interpretations of those subscales. He did report that differences in self-appraisals in distinct areas of experience were not found in test scores of 56 children aged 10 to 12. He reasoned that either these distinctions were not made by preadolescents or, if they were, they were made in the context of general self-esteem. This lack of differentiation between subscales would also be expected
from the low internal consistency coefficients and may suggest heterogeneity of items within subscales.

Only one study has systematically examined the subscale interpretation of the SEI. Dyer (1964) conducted a multitrait-multimethod validation study of the SEI using four subscale (trait) scores: Self, Family, Friends, and School. An Attitude Questionnaire (AQ), with a semantic differential format, was developed as an alternative measurement method. The AQ contained the concepts of “Myself,” “With My Family,” “With My Friends,” and “With My Schoolwork.” For each concept, subjects rated themselves on eight bipolar adjectives (Dyer, 1964, p. 29; adjectives were not reported). The two instruments were administered to 500 public school students in Flint, Michigan. Ten separate multitrait-multimethod matrices were constructed for boys and girls from grades three, five, seven, nine, and eleven.

The ten matrices were examined according to the four validity criteria proposed by Campbell and Fiske (1959). The convergent validity coefficients ranged from 0.02 to 0.64. Twelve of the 20 entries for boys and 19 of the 20 entries for girls were statistically significant and thus met the convergent validity criterion. (Alpha is not stated, but correlations significant at \( p < 0.01 \) are indicated.)

When the convergent validity coefficients were compared with heterotrait-heteromethod coefficients, only 22 of 40 validity coefficients met this first discriminant validity criterion. When the convergent validity coefficients were compared with the corresponding heterotrait-monomethod coefficients, only two met this criterion. They were the coefficients for “Family” for eleventh-grade boys and “School” for eleventh-grade girls. Finally, when the patterns of relationships among the traits were examined for each matrix, only the eleventh-grade data showed evidence of a consistent pattern. In sum, about half of the validity coefficients met the convergent validity criteria but could not meet the three more stringent discriminant validity criteria.

Finally, Dyer examined all the correlations between different traits (i.e., entries in the heterotrait-monomethod and the heterotrait-heteromethod triangles) by averaging over four coefficients for each of the six possible pairs of traits computed. These data indicated that the greatest differentiation (low correlations) between the traits was found between Friends and School (for girls and boys) and Family and School (for girls only). The mean correlations ranged from 0.18 to 0.45 across the five grade levels. The least differentiation (high correlations) was found for Self and School (for boys and girls) and Self and Friends (for girls only). Mean correlations ranged from 0.41 to 0.59 across grade levels.
In summarizing his multitrait-multimethod analysis, Dyer stated:

Strong method variance that was both shared, as indicated in the heteromethod triangles, and specific to each instrument, as indicated in the monomethod triangles, contributed to the scores in most of the samples below the eleventh grade, and thus resulted in lack of construct validity. (p. 70)

*Interpretation of Scores from the SEI*. Although separate areas of self-esteem have been identified conceptually, there appears to be no empirical support for their separate interpretation. Data reported by Coopersmith (1967) bearing on the between-portion of the nomological network suggest that interpretation of the SEI separate from other variables (e.g., popularity, conformity) may be warranted. However, a decision as to whether the SEI deserves a general self-concept interpretation must await further studies in which the between-portion of the nomological network is examined and studies in which plausible counter-hypotheses, such as the SEI is a measure of social desirability, are eliminated.

With respect to our definition of self-concept, the developmental (cross-sectional) data reported by Dyer (1964) seem to be supportive. These data hint that self-concept may become more differentiated with increasing age. In particular, some justification can be offered for a separate "Family" trait for eleventh-grade boys and a separate "School" trait for eleventh-grade girls. In the absence of factor analysis data, we cannot tell whether the traits themselves are not differentiable or whether the subscales have been inadequately formed. Finally, the test-retest correlations suggest that general self-concept is fairly stable over a year's time.

*How I See Myself Scale*

Elementary- and secondary-school forms of the *How I See Myself Scale* (HISM) have been developed in a series of research projects by Gordon (Note 7). At present, the scale is most often used to measure changes in certain aspects of self-concept resulting from intervention programs.

*Construct Definition*. Gordon (1966, p. 53; Note 7, p. 34) defined self-concept as the organization of all of the child's biological and environmental experiences as he has interpreted them into one highly organized, highly integrated, multifaceted system. Self-concept, that portion of the self-system of which the child is most aware, is the product of all his transactions at a particular point
in time and determines his behavior. As the child grows, different parts of self-concept and experience change in relative importance.

According to Gordon, since behavior is determined by the self-concept, inferences can be made from behavior to self-concept. Three facets of the self have been identified along with corresponding behaviors from which inferences might be made: self as revealed from self-report, self as inferred from observed behavior, and self as inferred from projective techniques. The relationships between self-report and inferences from overt behavior and self-report and inferences from projective techniques are expected to be small (Gordon, Note 7, see pp. 35-36; see also Combs, Soper, & Courson, 1963).

**Description of the Instrument.** The elementary (ages 3-6 years) and secondary (ages 7-12 years) forms of the HISM are primarily comprised of simple, self-evaluative statements developed from Jersild’s (1952) categories of children’s responses in compositions about themselves. The statements are arranged on five-point, bipolar scales such as “I do well in school work... I do not do well in school work” (elementary form) or “I do very well in school... I do not do well in school” (secondary form).

**Reliability.** Combs and Gordon (Gordon, Note 7) tested 80 slow-learning high school pupils over a two-week interval and obtained retest coefficients for four subscales (Teacher, Appearance, Body-build, Academic Achievement) ranging from 0.62 to 0.82. Yeatts (Note 13) reported total score stability coefficients ranging from 0.78 to 0.89 for third, fifth, eighth, and eleventh graders from a heterogeneous public school population, after an interval of nine days. No stability coefficients for the current subscale scores have been reported for children. Internal consistency coefficients have not been reported.

**Empirical Evidence in Support of Construct Interpretation.** The structure of the items on the HISM has been examined with factor analysis. In data from 960 laboratory school children, a number of factors were identified and remained consistent over grades three, five, eight, and eleven. Yeatts (Note 13), using data from boys and girls in public school grades three through twelve, found that the items loaded on five factors: physical appearance (e.g., “I like the way I look”), interpersonal adequacy (e.g., “my clothes are very nice”), teacher-student (e.g., “teachers like me”), academic adequacy (e.g., “I learn new things easily”), and autonomy (“I enjoy individual projects”). These labels are, however, only suggestive since, for example, items such as “My skin is nice looking” loaded on the factor, academic adequacy. Depending on grade level and sex, additional factors emerged. According to Gordon (Note 7), changes in factors “... fit well with
developmental theory and self-concept theory” (p. 41). There is reasonable consistency in the factors found over grades and sexes, but the composition of items within factors is sometimes curious. Since Gordon’s definition of self-concept gives no clear direction as to which items should cluster with which others, interpretation of scores based on these factors seems hazardous. And, the fact that the same item may be incorporated into several factors only creates additional problems of interpretation.

Gordon (Note 7) reported correlations using three different measurement methods—behavior observation, projective technique, and self-report (HISM)—and a variety of traits. Although these data were supposed to form a multitrait-multimethod matrix, the traits measured with the observational and projective techniques differed from those measured by the self-report method. Also, correlations were not reported for most of the traits which seemed to be related to the HISM factors. Gordon summarized the data as indicating that the correlations, although positive and significant, were generally low. He concluded that HISM measured self-concept somewhat differently from the inference approaches. We found the data uninterpretable.

Interpretation of HISM Scores. From the data available, interpretation of subscale scores—the scores used to report data from the instrument—seems hazardous. Further definitional and empirical work is needed in order to determine how to interpret the subscales. Work also needs to be done in specifying the between-portion of the nomological network and in examining relations between scores on the HISM and scores on other constructs to determine whether self-concept interpretations are warranted.

With regard to our definition of self-concept, it is difficult to draw inferences from the HISM data until further work is done. The following, then, is speculative. First, the definition of self-concept and factor analyses suggest a multifaceted self-concept, although it is not clear what those facets are. Second, some changes in factors with data from children of different ages suggest a developmental aspect to self-concept. However, these data do not support our contention that self-concept becomes more differentiated with age; they support Gordon’s notion of change in emphasis at different ages.

The Way I Feel About Myself (WIFAM)

The Piers-Harris Children’s Self-Concept Scale (The Way I Feel About Myself) was developed as a measure of general self-concept (Piers & Harris, 1964). The authors claim the instrument can be
used diagnostically in clinical and counseling settings and in classrooms for psychological referral, but its primary use has been in research on the development and correlates (e.g., anxiety, intelligence) of self-concept (Piers, Note 14).

Construct Definition. While an explicit definition of self-concept is not provided (Piers & Harris, 1964; Piers, Note 14), we infer from their writings that self-concept is multifaceted, relatively stable—especially as age increases—with distinct developmental characteristics. Finally, Piers (Note 14) distinguishes between self-concept which is reported by the self from inferred self-concept which is inferred by others from the self's (individual's) behavior.

Description of the Instrument. The 80 items, originally developed from Jersild's (1952) categories, are simple descriptive statements, such as "I am a happy person," with a yes/no response. These items were selected to discriminate between students with extremely low and high total scores. A total score or "cluster" scores can be obtained, the latter from items included in each of six clusters identified by factor analysis (Piers & Harris, 1964). The factors are labeled: Behavior, Intellectual and School Status, Physical Appearance and Attributes, Anxiety, Popularity, Happiness and Satisfaction.

Reliability. Internal consistency coefficients indicate stable rank orderings of students on total scores: (a) KR21 coefficients for 95 items used in a standardization study ranged from 0.78 to 0.93 for boys and girls in grades three, six, and ten; (b) corrected split-half coefficients for the total score were 0.90 and 0.87 for grade six and ten students, respectively. These rank orderings on total scores remain fairly stable over a period of several months: retest coefficients of 0.71 to 0.72 were found for the 95-item version and a coefficient of 0.77 was found for the current 80-item version (Piers, Note 14). Reliability coefficients for "cluster scores" have not been reported.

Empirical Evidence in Support of Construct Definition. The structure of the 80 items on the WIFAM has been examined by factor analysis (Piers & Harris, 1964). Ten factors accounted for 42 percent of the total test-score variance (Piers & Harris, 1964). Six factors were judged large enough to be interpretable: (a) Behavior ("I do many bad things"), (b) Intellectual and School Status ("I am good in my school work"), (c) Physical Appearance and Attributes ("I am good looking"), (d) Anxiety ("I cry easily"), (e) Popularity ("People pick on me"), (f) Happiness and Satisfaction ("I am a happy person").

Although these factors suggest a multifaceted self-concept, in the absence of some conceptual structure, we do not know if these observed factors are the ones to be expected. (There is some
correspondence with the facets identified by Sears; see below.) Furthermore, there are insufficient data to determine whether the scores on separate factors warrant separate interpretation. For example, 13 items did not load on any of the factors, whereas 11 items loaded on more than one factor. Neither correlations between subscale scores nor subscale correlations with measures of other constructs are reported.

Piers (Note 14) reported some data that bear on the self-concept interpretation of the total score. For example, Mayer (1965 reported by Piers, Note 14) found a correlation of 0.68 ($p < .05$) between the WIFAM and Lipsitt's Children's Self-Concept Scale for a sample of 98 males and females ranging in age from 12 to 16 years. Piers (Note 14) found correlations between WIFAM and teacher ratings of fourth and sixth graders on self-concept ranged from 0.06 to 0.41. Correlations with peer ratings ranged from 0.26 to 0.49.

Interpretation of WIFAM Scores. There is some evidence to suggest that total scores on the WIFAM warrant self-concept interpretations. However, this statement is based on convergent validity coefficients—correlations of scores on WIFAM with other measures of self-concept. In some cases, measures of other constructs have correlated as highly with WIFAM total scores as have other measures of self-concept. In order to have reasonable confidence in this interpretation, additional conceptual work on the between-portion of the nomological network accompanied with empirical data is required. At present, there are no data available on the interpretability of “cluster” scores.

With respect to our definition of self-concept, data on the WIFAM support a multifaceted view of self-concept, with general self-concept (total scores on WIFAM) a stable characteristic of the individual. Data from the WIFAM bearing on the hierarchical characteristic of self-concept have not been reported. Likewise, data that would bear on the developmental aspect of self-aspect have not been reported.

Self-Concept Inventory (SCI)

The Self-Concept Inventory provides a measure of general and ten facets of self-concept. The original form (Sears, Note 5) has been used as a research tool (e.g., Bixler, 1965; Gelfand, 1962; Trickett, 1969; Torshen, Note 3; Sears, Note 5; Spaulding, Note 6). A revised form is currently being used (Sears et al., Note 8; Sears, Note 15) to evaluate changes in self-concept resulting from intervention programs.

Construct Definition. For Sears, self-concept represents the child's expectation of success in solving problems and carrying
out tasks. The expectancies are learned and can be changed according to principles of learning. Furthermore, self-concept is multifaceted with each facet differing in importance or reward value. Since expectancies are acquired within the context of a facet (e.g., "school subjects"), an individual can predict success or failure of his behavior within that facet. Finally, various aspects of self-concept have properties similar to drive; to protect the self-concept, a person will select or strive toward those behaviors that maintain and enhance self-concept (see Sears & Sherman, 1964, p. 10).

**Description of the Instrument.** The original form of the SCI contains 100 items, ten items for each of ten facets of self-concept. The ten facets were identified from compositions written by sixth graders and the items within each facet are short, descriptive phrases. The facets and example items are: (a) physical ability ("Being built for sports"), (b) mental ability ("Learning things rapidly"), (c) school relations with same sex ("Making friends easily with boys"),(d) school relations with opposite sex ("Making friends easily with girls"), (e) school relations with teacher ("Getting along well with teachers"), (f) work habits ("Getting my school work in on time"), (g) social virtues ("Being willing for others to have their way sometimes"), (h) happy qualities ("Getting a lot of fun out of life"), and (i) school subjects ("Getting good grades in school"). The student responds to each item on three questions: (a) satisfied with self, yes or no; (b) expect improvement, maybe or probably won't; and (c) compared with others, rate yourself.

The revised form of the SCI has 48 items representing nine facets, most of which correspond to the original scale. The revisions were: (a) Mental Ability was divided into Divergent and Convergent Mental Ability; (b) Social Relations with the Opposite Sex was dropped; (c) School Subjects was reduced in size; (d) the "personality" subareas (e.g., Happy Qualities, Attractive Appearance, Social Virtues) were reduced in size; and (e) the satisfaction and improvement responses were dropped; the student only rates himself as compared with others.

**Reliability.** The reliability of both forms of the SCI, in general, appears to be adequate for research on individual differences. For the 100-item scale, Sears (Note 5) reported a corrected split-half coefficient of 0.95 for the total score, self-rating. Similar coefficients for the subareas ranged from 0.79 to 0.93 (Bixler, 1965; Trickett, 1969; Sears, Note 5). Retest coefficients from fall to spring ranged from 0.29 to 0.67 for the total score. For the revised scale, Sears (Note 15) reported internal consistency coefficients of 0.90, 0.88, and 0.92 for the facet scores combined into total, personality, and mental ability scores, respectively. For the
nine separate facets, the coefficients ranged from 0.56 to 0.89. A retest (six-month interval) coefficient of 0.50 was reported for the total score (Sears et al., 1972). The revised form will not be discussed further since its manual is still in preparation.

**Empirical Evidence in Support of Construct Interpretation.** The structure of the SCI has been examined by factor analysis (Torshen, Note 3; Sears, Note 5). In Sears’ analyses, 195 fifth graders were divided into four subgroups based on sex and their composite scores from the Primary Mental Abilities test (superior or average). For the four separate factor analyses, items clustered on one factor: “generalization.” Sears (Note 5) concluded that, although the ten subscales may measure different self-concept referents, “self-concept is a more or less unitary aspect of personality” (p. 81).

Torshen (Note 3), using data from 402 fifth-grade students, found separate but correlated factors corresponding to each of the ten subareas of the SCI. In addition, two other factors, “self-improvement” and “self-rating,” corresponded to two of the three rating scales on the SCI. The third type of rating—satisfaction with self—did not appear as a separate factor since “it was obtainable from the other factors” (Torshen, Note 3, p. 47). Torshen’s results, then, give some support to the interpretation of subarea scores as representing different subareas of self-concept.

The difference in the Sears (Note 5) and Torshen (Note 3) findings might be explained in two ways. One explanation is that Sears divided students into homogeneous subgroups (by sex and aptitude), whereas Torshen designed her study to maximize variability among students. Another explanation might be the differences in factor analytic method used. Sears used the principal components method with varimax rotation (Kaiser, 1958). In contrast, Torshen (Note 3; cf. Jöreskog, 1969) imposed Sears’ ten-subarea structure on the data and tested the goodness of fit.

Although data have not been collected in a multitrait-multimethod matrix, parts of the matrix can be built from data reported by Bixler (1965), Trickett (1969), and Torshen (Note 3). In the Bixler study, heterotrait-heteromethod data are missing. Also, only six of the ten self-concept subareas are represented. In the Torshen study, only the heterotrait-monomethod data are available. In the Trickett study, three methods were used: self-report (SCI), teacher ratings, and peer ratings. However, data are available for only four of the ten traits and most reliability data are missing. In discussing the four criteria for convergent and discriminant validity of the SCI, data from all three studies will be used.
Reliability. With a few exceptions (Bixler data), the square root of the reliability coefficient is the largest entry in the multitrait-multimethod matrix.

Convergent Validity. The validity coefficients are significantly greater than zero (except for two coefficients from Trickett).

Discriminant Validity. Data are available in the heterotrait-heteromethod triangles from only one study (Trickett, 1969). With this criterion, data from the SCI and peer ratings on four facets (happy qualities, school subjects, physical ability, appearance) suggest that three of the four (except happy qualities) can be discriminated. In comparing teacher ratings with the other two methods, the four facets cannot be discriminated. Finally, data from Bixler (1965) and Trickett (1969) indicate that, in most cases, the correlations in the validity diagonal are not higher than corresponding entries in the heterotrait-monomethod triangle. These data, then, suggest that the facets may not be distinguishable from one another and that the type of measurement method employed strongly influences the scores obtained. These findings cast doubt on interpretations of subscale scores as separate facets of self-concept.

Because of missing data, patterns of intercorrelations cannot be examined thoroughly. A pattern of high intercorrelations in the heterotrait-monomethod triangles is found among "academic" traits: (a) School Subjects, (b) Mental Ability, (c) Work Habits, and (d) Social Relations with Teachers. Correlations among other facets do not show additional patterns. For example, a "social" trait using the Social Virtues, Social Relations, and Happy subscales might be hypothesized. But the intercorrelations do not support this hypothesis. Likewise, a "physical" self-concept facet might be hypothesized, but the data do not support this either.

Interpretation of Scores from the SCI. Most of the data presented above suggest that separate interpretation of the ten subscales of facets of the SCI is not warranted. Nevertheless, certain subsets of facets, such as a subset called "academic self-concept," might warrant an interpretation separate from some other subset, such as "nonacademic self-concept." This observation is based on correlational data examined above and data presented by Torshen (Note 3) in her factor analysis and other data analyses. A decision as to whether the total score warrants a self-concept interpretation must await thorough examination of the between-portion of the nomological network. However, preliminary data from Sears (Note 5) and Torshen (Note 3) suggest that such an interpretation may be warranted.

With regard to our definition of self-concept, these data support to some degree, a hierarchical, multifaceted construct. This
observation is supported by data suggesting a strong general self-concept discriminable into, at least, academic and nonacademic facets (see Figure 1). The retest correlations lend some support to a stable, general self-concept; data are not available to evaluate the developmental aspect of the definition.

**Summary and Recommendations for Construct Validation**

The enhancement of each student's self-concept is important in today's educational *Zeitgeist* either as an educational outcome or as a moderator of achievement. Most self-concept studies examine intercorrelations between self-concept and other constructs, or differences in mean self-concept scores between different populations of students, or changes in self-concept due to some treatment. Taken individually, they often provide important insights into factors that motivate students in and out of school and into alternative courses of action that may enhance students' self-concepts. However, considered as a body of research, self-concept studies lack the focus that would result from an agreed-upon definition of self-concept, lack adequate validation of interpretations of self-concept measures, and lack empirical data on the equivalence of the many self-concept measures currently being used.

The methodology of construct validation represents one approach that may help draw together the diverse studies of self-concept. This methodology examines the ingredients of a construct definition and suggests logical and empirical methods for examining the validity of interpretations of self-concept measurements.

Self-concept was defined as an individual's perception of himself and a number of distinctive features were identified: (a) organized, (b) multifaceted, (c) hierarchical, (d) stable (general self-concept)/unstable (situational), (e) developmental, (f) descriptive and evaluative, and (g) differentiable from other constructs.

Data from five instruments, commonly used as measures of self-concept, tentatively support the interpretations of scores according to features of the proposed definition. Data from each of these instruments suggested a general self-concept interpretation of scores. This interpretation needs to be validated against alternative construct interpretations of these total scores. Although the subscores of the various instruments are moderately correlated, data from four of the five instruments suggested that general self-concept might be divided into different dimensions (multifaceted feature). There was not perfect agreement on the clustering of items from different instruments, but an examina-
tion of items within factors indicated sufficient consistency across instruments and subject populations (mainly Caucasian and middle class) to suggest that self-concept scores can be related to four general areas of experience: academic, social, emotional, and physical. In the absence of data comparing one instrument with another, this interpretation is tentative.

There is a paucity of data bearing on the hierarchical interpretation of self-concept scores. The data suggested a hierarchy in which general self-concept (at the apex) could be divided into academic self-concept and nonacademic self-concept (Ludwig & Maehr, 1967; Torshen, Note 3; Brookover et al., Note 9). Academic self-concept, in turn, may be divided into specific subject-matter self-concepts (Torshen, Note 3; Brookover et al., Note 9). The other facets may include, at least, a physical self-concept (Ludwig & Maehr, 1967). However, without additional information interrelating self-concept facets at various levels of the hierarchy, a counterhypothesis, such as self-concept divides into many situation-specific responses all of which are positively correlated due to a verbal report factor, cannot be eliminated. Until the hierarchical feature of self-concept has been examined more adequately, then, it must be considered extremely tentative.

The hierarchical feature led to a hypothesis that general self-concept is stable, whereas situation-specific self-concept is unstable. The data suggested a stable general self-concept. A few studies have examined the hypothesis that self-concept becomes increasingly stable moving from the base of the hierarchy to the apex. Ludwig and Maehr's (1967) data gave tentative support to this interpretation. Data reported by Brookover et al. (Note 9) suggested that self-concept of achievement in academic subjects is just as stable as self-concept of academic ability. Until stability is estimated at each level of the hierarchy, this feature must be considered tentative.

The data reviewed here have not examined the developmental hypotheses directly. Data from Gordon (Note 7) showed that different factors emerged from students at different ages. Some data from Dyer (1964) hinted at increased differentiation with age. But, further work needs to be done before this feature of the definition is adequately tested.

A distinction between descriptive and evaluative aspects of self-concept is not supported empirically. For example, Torshen (Note 5) reported that a self-descriptive dimension of self-concept was subsumed by two evaluative dimensions.

The intent of the review has been to pull together some of the more significant, individual studies of self-concept into a coherent framework whose evidence could be systematically examined. It has showed a need to limit the number of self-concept definitions, to examine the validity of the proposed
constructs, to examine the rival interpretations of the definitions, and to ascertain the equivalence of self-concept instruments, so that data collected with one instrument can be related to data collected with another. Once some degree of standardization of definition, instrumentation, and interpretation has been achieved, the determiners, consequences, and correlates of self-concept, along with their implications for educational practices, can be investigated systematically.

Reference Notes

1. Gordon, E. W., & Wilkerson, D. A. Compensatory education for the disadvantaged: Programs and practices for preschool through college. New York: College Entrance Examination Board, 1966.
2. Brookover, W. B., LePere, J. M., Hamachek, D. E., Thomas, S., & Erickson, E. L. Self-concept of ability and school achievement, II. USOE Cooperative Research Report, Project No. 1636. East Lansing: Michigan State University, 1965.
3. Torshen, K. P. The relation of classroom evaluation to students' self-concepts and mental health. Unpublished doctoral dissertation, University of Chicago, 1969.
4. Bledsoe, J. C., & Garrison, K. C. The self-concepts of elementary school children in relation to their academic achievement, intelligence, interests, and manifest anxiety. USOE Cooperative Research Report, Project No. 1008, Athens: University of Georgia, 1962.
5. Sears, P. S. The effect of classroom conditions on the strength of achievement motive and work output on elementary school children. USOE Cooperative Research Report, Project No. OE-873, Stanford, California: Stanford University, 1965.
6. Spaulding, R. Achievement, creativity, and self-concept correlates of teacher-pupil transactions in elementary schools. USOE Cooperative Research Report, Project No. 1352, Urbana: University of Illinois, 1963.
7. Gordon, I. J. A test manual for the how I see myself scale. Gainesville: Florida Educational Research and Development Council, 1968.
8. Purkey, W. W. The search for self. Gainesville, Florida: Florida Educational Research and Development Council Research Bulletin, 1968, 4 (2), whole.
9. Brookover, W. B., Erickson, E. L., & Joiner, L. M. Self-concept of ability and school achievement, III. USOE Cooperative Research Report, Project No. 2831, East Lansing: Michigan State University, 1967.
10. Brookover, W. B., Paterson A., & Thomas, S. Self-concept of ability and school achievement. USOE Cooperative Research Report, Project No. 845, East Lansing: Michigan State University, 1962.
11. Shulman, L. S. Multiple measurement of self-concept. Paper presented at the annual meeting of the Educational Research Association, Chicago, 1968.
12. Paterson, A. Reliability and validity of self-concept of ability scale. In W. B. Brookover, E. L. Erickson, & L. M. Joiner, Self-concept of ability and school achievement, III, USOE Cooperative Research Report, Project No. 2831, East Lansing: Michigan State University, 1967, 155-172.
13. Yeatts, P. P. Developmental change in the self-concept of children grades 3-12. Gainesville, Florida: Florida Educational Research and Development Council Research Bulletin, 1967, 3 (2), whole.
14. Piers, E. V. Manual for the Piers-Harris, children's self-concept scale. Nashville, Tennessee: Counselor Recordings and Tests, 1969.
15. Sears, P. S. Memorandum with respect to use of the Sears Self-Concept Inventory. Stanford University, 1966. (mimeographed)
References

Aiken, W. M. *Story of the eight-year study*. New York: Harper, 1942.
Althauser, R. P., & Heberlein, T. A. Validity and the multitrait-multimethod matrix. In E. F. Borgetta (Ed.), *Sociological methodology*. San Francisco: Jossey-Bass, 1971.
Bem, D. J. Self-perception theory. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 6). New York: Academic Press, 1972.
Bilby, R. W., Brookover, W. B., & Erickson, E. L. Characterizations of self and student decision making. *Review of Educational Research*, 1972, 42, 505-524.
Bixler, P. A. Changing self-concept in sixth-grade class groups. (Doctoral dissertation, Stanford University, 1965). *Dissertation Abstracts*, 1965, 26 3750. (University Microfilms No. 65-12, 730)
Blalock, H. M. *Causal inferences in nonexperimental research*. Chapel Hill: University of North Carolina Press, 1964.
Boruch, R. F., Larkin, J. D., Wolins, L., & MacKinney, A. C. Alternative methods of analysis: Multitrait-multimethod data. *Educational and Psychological Measurement*, 1970, 30, 833-853.
Boruch, R. F., & Wolins, L. A procedure for estimation of trait, method, and error variance attributable to a measure. *Educational and Psychological Measurement*, 1970, 30, 547-574.
Brownfain, J. J. Stability of the self-concept as a dimension of personality. *Journal of Abnormal and Social Psychology*, 1952, 47, 597-606.
Bruner, J. S. Social psychology and perception. In E. E. Maccoby, T. M. Newcomb, & E. L. Hartley (Eds.), *Readings in social psychology* (3rd ed.). New York: Holt, Rinehart, & Winston, 1958, 85-94.
Bruner, J. S. *The process of education*. Cambridge: Harvard University Press, 1960.
Bruner, J. S. Social psychology and perception. In D. E. Hamacek (Ed.), *The self in growth, teaching, and learning*. Englewood Cliffs, New Jersey: Prentice-Hall, 1965.
Callahan, R. E. *Education and the cult of efficiency: A study of the social forces that have shaped the administration of the public schools*. Chicago: University of Chicago Press, 1962.
Campbell, D. T., & Fiske, D. W. Validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 1959, 56, 81-105.
Clark, K. B. Educational stimulation of racially disadvantaged children. In H. A. Passow (Ed.), *Education in depressed areas*. New York: Bureau of Publications, Teachers College, Columbia University, 1963.
Coller, A. R. *The assessment of "self-concept" in early childhood education*. Urbana, Illinois: ERIC Clearinghouse on Early Childhood Education, 1971. (ERIC Document Reproduction Service No. ED 057 910)
Combs, A. W., & Soper, D. W. The self, its derivative terms, and research. *Journal of Individual Psychology*, 1957, 13, 134-145.
Combs, A. W., Soper, D. W., & Courson, C. C. The measurement of self-concept and self-report. *Educational and Psychological Measurement*, 1963, 23, 493-500.
Coopersmith, S. A method for determining types of self-esteem. *Journal of Abnormal and Social Psychology*, 1959, 59, 87-94.
Coopersmith, S. *The antecedents of self-esteem*. San Francisco: Freeman, 1967.
Crano, W. D., Kenny, D. A., & Campbell, D. T. Does intelligence cause achievement?: A cross-lagged panel analysis. *Journal of Educational Psychology*, 1972, 63, 258-275.
Cronbach, L. J. *Essentials of psychological testing* (2nd & 3rd eds.). New York: Harper and Row, 1960 and 1970.
Cronbach, L. J. Test validation. In R. L. Thorndike (Ed.), *Educational measurement*. Washington, D.C.: American Council on Education, 1971.
Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. *The dependability of behavioral measurements: Theory of generalizability for scores and profiles.* New York: Wiley, 1972.
Cronbach, L. J., & Meehl, P. E. *Construct validity in psychological tests.* *Psychological Bulletin,* 1955, 52, 281-302.
Crowne, D. P., & Stephens, M. W. *Self-acceptance and self-evaluating behavior: A critique of methodology.* *Psychological Bulletin,* 1961, 58, 104-121.
Duval, S., & Wickland, R. A. *The effects of objective self-awareness on attribution of causality.* *Journal of Experimental Social Psychology,* 1973, 9, 17-31.
Dyer, C. O. *Construct validity of self-concept by a multitrait-multimethod analysis.* (Doctoral dissertation, University of Michigan, 1963). *Dissertation Abstracts,* 1964, 25, 8154. (University Microfilms No. 64-8154)
Edwards, A. L. *The social desirability variable in personality assessment and research.* New York: Dryden, 1957. (a)
Edwards, A. L. *Techniques of attitude scale construction.* New York: Appleton-Century-Crofts, 1957. (b)
Engle, M. *The stability of self-concept in adolescence.* *Journal of Abnormal and Social Psychology,* 1959, 58, 211-215.
Epstein, R., & Komorita, S. S. *Self-esteem, success-failure, and locus of control in Negro children.* *Developmental Psychology,* 1971, 4, 2-8.
Fantini, M. D., & Weinstein, G. *The disadvantaged.* New York: Harper & Row, 1968.
Feo, U. G. *New developments in facet design and analysis.* *Psychological Review,* 1965, 72, 262-274.
Gelfand, D. *The influence of self-esteem on rate of verbal conditioning.* *Journal of Abnormal and Social Psychology,* 1962, 65, 259-265.
Gergen, K. *The effects of interaction goals and personalistic feedback on the presentation of self.* *Journal of Personality and Social Psychology,* 1965, 1, 413-424.
Goode, W. J., & Hatt, P. K. *Methods in social research.* New York: McGraw-Hill, 1952.
Gordon, C. *Self-conceptions methodologies.* *The Journal of Nervous and Mental Disease,* 1969, 148, 328-364.
Gordon, I. J. *Studying the child in school.* New York: John Wiley, 1966.
Hamachek, D. E. (Ed.). *The self in growth, teaching, and learning.* Englewood Cliffs, New Jersey: Prentice-Hall, 1965.
Herbert, E., Gelfand, D. M., & Hartman, D. P. *Imitation and self-esteem as determinants of self-critical behavior.* *Child Development,* 1969, 40, 421-430.
Hishiki, P. C. *Self-concepts of 6th-grade girls of Mexican-American descent.* *California Journal of Educational Research,* 1969, 20, 56-62.
Hoepfner, R., Stern, C., & Nummedal, S. G. *CSE-ECRC preschool/kindergarten test evaluations.* Los Angeles: UCLA Graduate School of Education, 1971. (ERIC Document Reproduction Service No. ED 055 123)
James, W. *Psychology.* Greenwich, Conn.: Fawcett, 1963.
Jersild, A. T. *In search of self.* New York: Teachers College Bureau of Publications, 1952.
Jöreskog, K. G. *A general approach to confirmatory maximum likelihood factor analysis.* *Psychometrika,* 1969, 34, 183-202.
Kaiser, H. F. *The varimax criterion for analytic rotation in factor analysis.* *Psychometrika,* 1958, 23, 187-200.
Kelley, H. H. *The process of causal attribution.* *American Psychologist,* 1973, 28, 107-128.
Koppel, M. A., & Sechrest, L. *A multitrait-multimethod matrix analysis of sense of humor.* *Educational and Psychological Measurement,* 1965, 25, 77-85.
Loevinger, J. *Objective tests as instruments of psychological theory.* *Psychological Reports,* 1957, 3, 635-694.
Long, B., Henderson, E. H., & Ziller, R. C. Developmental changes in the self-concept during middle childhood. Merrill-Palmer Quarterly of Behavior and Development, 1967, 13, 201-215.

Long, B., Ziller, R. C, & Henderson, E. H. Developmental changes in the self-concept during adolescence. School Review, 1968, 76, 210-230.

Ludwig, D. J., & Maehr, M. L. Changes in self-concept and stated behavioral preferences. Child Development, 1967, 38, 453-467.

Marston, A. R. Dealing with low self-confidence. Educational Research, 1968, 10, 134-138.

McDonald, F. J. Educational psychology, Belmont, Calif.: Wadsworth, 1965.

Mischel, W. Personality and assessment. New York: John Wiley, 1968.

Mote, F. A study of the home environment in its relationship to children’s self-concepts. (Doctoral dissertation, Stanford University, 1966). Dissertation Abstracts International, 1967, 27, 3319A. (University Microfilms No. 67-4509)

Parker, J. The relationship of self-report to inferred self-concept. Educational and Psychological Measurement, 1966, 26, 691-700.

Piers, E. V., & Harris, D. A. Age and other correlates of self-concept in children. Journal of Educational Psychology, 1964, 55, 91-95.

Purkey, W. W. Self-concept and school achievement. Englewood Cliffs, N.J.: Prentice-Hall, 1970.

Rogers, C. R., & Dymond, R. F. Psychotherapy and personality change. Chicago: University of Chicago Press, 1954.

Sears, P. S. Self-concept in the service of educational goals. California Journal for Instructional Improvement, 1963, 6, 3-12.

Sears, P. S., Adenubi, M., Block, M., Crist, J., Gamble, J., & Hubner, J. Effective reinforcement for achievement behaviors in disadvantaged children: The first year. Technical Report No. 30, Stanford Center for Research and Development in Teaching, 1972. (ERIC Document Reproduction Service No. ED 067 442)

Sears, P. S., & Sherman, V. In pursuit of self-esteem. Belmont, Calif.: Wadsworth, 1964.

Sheriff, M., & Cantril, H. The psychology of ego involvements, social attitudes, and identifications. New York: John Wiley, 1947.

Smiley, M. B. Objectives of educational programs for the educationally retarded and disadvantaged. In P. A. Witty (Ed.), The educationally retarded and disadvantaged. Sixty-Sixth Yearbook of the National Society for the Study of Education, Part I. Chicago: University of Chicago Press, 1967.

Smith, I. D. Impact of computer-assisted instruction on student attitudes. Journal of Educational Psychology, 1973, 64, 366-372.

Snygg, D., & Combs, A. W. Individual behavior. New York: Harper & Row, 1949.

Soares, A. T., & Soares, L. Self-perceptions of culturally disadvantaged children. American Educational Research Journal, 1969, 6, 31-46.

Sullivan, H. S. The interpersonal theory of psychiatry. New York: W. W. Norton, 1953.

Super, D. E. Toward making self-concept theory operational. In D. E. Super, J. P. Jordan, N. Maitlin, & R. Starishevsky (Eds.), Career development: Self-concept theory. New York: College Entrance Examination Board, 1963.

Symonds, P. M. The ego and the self. New York: Appleton-Century-Crofts, 1951.

Tannenbaum, A. J. Social and psychological considerations in the study of the socially disadvantaged. In P. A. Witty (Ed.), The educationally retarded and disadvantaged. Sixty-Sixth Yearbook of the National Society for the Study of Education, Part I. Chicago: Chicago University Press, 1967.

Trickett, H. Stability and predictability of children’s self-concept and perceptions by others: A developmental study. (Doctoral dissertation. Stanford University, 1968). Dissertation Abstracts International, 1969, 29, 2557A. (University Microfilms No. 69-306)
Trowbridge, N. Self-concept and socioeconomic status in elementary school children. *American Educational Research Journal*, 1972, 9, 525-538.
Vernon, P. E. *The structure of human abilities*. London: Methuen, 1950.
Wattenberg, W. W., & Clifford, C. Relation of self-concepts to beginning achievement in reading. *Child Development*, 1964, 35, 461-467.
Wylie, R. *The self-concept: A critical survey of pertinent research literature*. Lincoln: University of Nebraska Press, 1961.
Yamamoto, K. (Ed.). *The child and his image*. Boston: Houghton Mifflin, 1972.
Yee, A. H, & Gage, N. L. Techniques for estimating the source and direction of causal influence in panel data. *Psychological Bulletin*, 1968, 70, 115-126.
Zirkel, P. A. Self-concept and the "disadvantage" of ethnic group membership and mixture. *Review of Educational Research*, 1971, 41, 211-225.
Zirkel, P. A. Enhancing the self-concept of disadvantaged students. *California Journal of Educational Research*, 1972, 23, 125-137.

AUTHORS
RICHARD J. SHAVELSON *Address*: Graduate School of Education, University of California, 405 Hilgard Ave., Los Angeles, CA 90024. *Title*: Associate Professor. *Degrees*: B.A., University of Oregon; M.A., San Jose State College; Ph.D., Stanford University. *Specialization*: Learning and individual differences, research design and measurement.
JUDITH J. HUBNER *Address*: School of Education, Stanford University, Stanford, CA 93405. *Title*: Doctoral Candidate. *Degrees*: B.A., University of California at Berkeley; M.A., San Francisco State College. *Specialization*: Child development.
GEORGE C. STANTON *Address*: School of Education, Stanford University, Stanford, CA 94305. *Title*: Doctoral Candidate. *Degrees*: B.A., Lake Forest College; M.A., Cornell University. *Specialization*: Learning and individual differences.