AN EVALUATION OF AN INSTRUMENT DESIGNED TO MEASURE THE CONSTRUCT "SELF-CONCEPT OF (ACADEMIC) ABILITY"

By

Ann Carlson Paterson

AN ABSTRACT OF A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Sociology and Anthropology

1966

ABSTRACT

AN EVALUATION OF AN INSTRUMENT DESIGNED TO MEASURE THE CONSTRUCT "SELF-CONCEPT OF (ACADEMIC) ABILITY"

By Ann Carlson Paterson

The "Self-Concept of Ability (SCA) Scale" was designed and used at Michigan State University to study the relationship between self-definitions of academic ability and actual school achievement. The self-report instrument consists of eight multiple-choice items with five response alternatives. The instrument evaluation is based on results obtained in 1960-61 from 513 male and 537 female 7th grade students in the four junior high schools of a Midwestern community. The criterion of achievement was grade point average (GPA). In addition to the SCA Scale, four specific-subject scales were constructed directly parallel to the SCA Scale except that their content was restricted to either arithmetic, English, social studies or science. These specific-subject scales were not the focus for evaluation, but were designed for study of specific-subject variations in the self-concept and achievement relationship.

The instrument evaluation focused on the establishment of construct validity although other conventional validity and reliability indicies were examined. To this end, four

hypotheses were theoretically derived from the symbolic interactionist framework and empirically tested. Major conclusions from the hypothesis testing follow:

1. The SCA Scale scores are significantly and positively correlated with GPA ($r = .57$ for each sex). When the effect of IQ is partialled out the SCA-GPA correlations remain significant ($r_{12.3} = .42$ and .39 for males and females). Even for a group of 110 "Over-" and "under-achievers" where there is a known negative correlation between IQ and GPA, the correlation between SCA and GPA is .40 for males and females combined.

2. A combination of high GPA and low SCA is significantly less likely to occur than a combination of high SCA and low GPA. This suggests that a high-self-concept of ability is a necessary but not sufficient basis for high achievement.

3. The specific-subject scales are positively and significantly correlated with achievement in parallel subjects. The mean scores of the general SCA Scale are higher (more positive) than the means of any of the specific-subject scale scores but closest to the specific-subject score in that subject in which the student has his highest achievement. This suggests that general self-definitions of ability are more heavily influenced by areas of strength than by areas of weakness. Among students with "nonuniform" achievement patterns, the specific-subject scales were in general

significantly better predictors of achievement in the parallel subject than was the general SCA Scale. Specific-subject variations in the self-concept and achievement relationship were found for males and females and are examined in some detail.

4. The general SCA Scale is a better predictor of achievement in a specific subject than is any specific-subject scale other than the one in the parallel subject. The general SCA Scale is also a better predictor of general achievement than is any specific-subject scale.

Other evidence for reliability and validity of the SCA Scale follows:

1. The SCA Scale was judged to have content validity.

2. In predicting end-of-year GPA, the SCA Scale scores have only slightly lower beta weights than do IQ test scores ($R_{1.23}$ = .69 and .72 for males and females). A cross-validation of the two-variable prediction equation provides evidence that the equation accounts for real and stable variance.

3. The stability reliability coefficients of the SCA Scale for a 12-month interval are .75 and .77 for males and females. Such figures are difficult to interpret as theory suggests self-conceptions of ability may change.

4. Guttman scalogram analysis, the Hoyt internal consistency procedure, factor analysis and individual item analysis all affirm the basic homogeneity of the SCA Scale.

Reproducibility coefficients for the Guttman analysis were above .95; the Hoyt coefficients were .82 and .84 for males and females. The centroid factor analysis of the SCA Scale items gave loadings ranging from .53 to .72 on the first factor interpreted as self-concept of ability. A weak second factor having item loadings of .06 to .40, was interpreted as the time dimension which had been postulated by logical analysis of the Scale. Individual item analysis showed no item-total score correlation below .59; only two were below .65.

AN EVALUATION OF AN INSTRUMENT DESIGNED TO MEASURE THE CONSTRUCT "SELF-CONCEPT OF (ACADEMIC) ABILITY"

By

Ann Carlson Paterson

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Sociology and Anthropology

1966

ACKNOWLEDGMENTS

The writer wishes to express her thanks to Dr. Wilbur B. Brookover and the many members of the Michigan State University Research Staff who have contributed to the completion of this research. The funds for this research were contributed in part by grants from the Cooperative Research Branch of the U. S. Office of Education.

TABLE OF CONTENTS

LIST OF TABLES

TABLE Page

LIST OF FIGURES

LIST OF APPENDICES

CHAPTER I

DESCRIPTION OF THE PROBLEM

Introduction

The need for the creation and evaluation of instruments designed to measure limited or specific self constructs is clearly stated in Ruth C. Wylie's book The Self Concept: A Critical Survey of Pertinent Research Literature published in 1961.[1] After pulling together much of the research literature on self concept and taking a hard look at the methodology of the various studies, Dr. Wylie concludes that "the total accumulation of substantive findings is disappointing, especially in proportion to the great amount of effort which obviously has been expended."[2] The reasons for this state of affairs, Dr. Wylie says, are four:[3]

1. The lack of proper scientific characteristics of the theories themselves;
2. The inevitable difficulties encountered in formulating relevant, well-controlled research in a new area;

---

[1]Ruth C. Wylie, The Self Concept: A Critical Survey of Pertinent Research Literature (Lincoln: University of Nebraska Press, 1961).

[2]Ibid., p. 317. [3]Ibid., p. 323.

3. The understandable fact that individual researches in a new area are not part of a planned research program and therefore cannot be easily synthesized; and
4. Avoidable methodological flaws.

Among the suggestions made by Wylie for implementing more interpretable research was that limited theoretical self constructs be studied by means of more limited and well-analyzed measuring instruments which are shown to have behavioral correlates. Wylie argues that it is only by the careful compendium of data at the molecular level that we shall have an adequate base for making conclusions about the more global self constructs.

This dissertation is an evaluation of an instrument designed from a molecular analysis of a self construct which, it is hoped, avoids many of the limitations noted by Wylie. The construct, "self-concept of ability" was theoretically derived from the symbolic interactionist framework of G. H. Mead and C. H. Cooley. A self-report instrument, the Self-Concept of Ability (SCA) Scale, was devised to tap the construct, and theoretically-derived hypotheses were tested to evaluate the construct validity of the instrument. Conventional reliability and validity checks were also made on the instrument so that it might be more completely understood. The relevant behavioral correlate to the "self-concept of ability" construct was academic achievement as indexed by grade point averages

(GPA). The relationship of the construct to the behavioral criterion was examined in some detail.

The theoretical framework for the "self-concept of ability" construct will be outlined in the next section of Chapter I followed by a more detailed discussion of the theoretical background for the four hypotheses tested in the establishment of construct validity. While the present research focuses on construct validity, the broader issues of validity and reliability are also considered following the classification scheme set forth by the American Psychological Association.[4] Although some controversy exists as to the adequacy of the APA approach, it has been generally accepted and extensively used and thus offers more continuity with other research than alternative classification systems. The relevant constructs recommended by the American Psychological Association for evaluation of psychological tests provide the general outline for Chapter I as follows:

Section 1: Validity

Construct Validity (Hypothesis Testing)
Content Validity
Predictive Validity

---

[4]American Psychological Association, American Educational Research Association, and National Council on Measurements Used in Education, Joint Committee, "Technical Recommendations for Psychological Tests and Diagnostic Techniques," Psychological Bulletin, 51, No. 2, Part 2, Special Supplement (March, 1954). Since this writing the 1966 revision of the "Technical Recommendations" has become available.

Section 2: Reliability

Stability Reliability
Internal Consistency Reliability

Chapter II will review literature in two areas: first literature pertinent to the theoretical derivation of the "self-concept of ability" construct; and secondly, literature on the establishment of construct validity, a frontier area in measurement.

Chapters III and IV will present the research methodology and findings respectively, following the same order of presentation as found in Chapter I. Chapter III will also include a detailed discussion of the development of the measuring instrument. Chapter V will consist of a summary of findings and some general conclusions.

The present research is part of a larger research program being carried out at Michigan State University under the direction of Dr. Wilbur Brookover. This research program, dealing with the relationship of self-conceptions of ability to classroom achievement, has been sponsored by the U. S. Office of Education. The writer served as Assistant Project Director on the first phase of the research program which is reported in Project Report #845,[5] and aided in the concepualization of the second phase of

---

[5]Wilbur B. Brookover, Ann Paterson, and Shailer Thomas, Self-Concept of Ability and School Achievement. Final Report of Cooperative Research Project No. 845 (East Lansing, Michigan: Office of Research and Publications, College of Education, Michigan State University, 1962).

the research reported in Project Report #1636.[6] These project reports are discussed in some detail in Chapter II.

The research reported in this dissertation is an evaluation of the Self-Concept of Ability Scale--an original measuring instrument designed for use in the Michigan State research program on self-concept and achievement. The research program did not include provision for extensive instrument evaluation so the present research does not approach an optimal evaluation program, but represents "best possible" within a somewhat restrictive context. The data reported in this dissertation consist of original analyses plus relevant data from Project Reports #845 and #1636. The concern here is with giving a comprehensive evaluation of the Self-Concept of Ability Scale for the benefit of future users of the instrument. The analysis of original measuring instruments is rarely done in sociological research (or in self-concept research) resulting in the loss of valuable data on the relation of the instrument to its theoretical context. The present research hopes to avoid at least this gap in understanding.

---

[6]Wilbur B. Brookover, and others, Self-Concept of Ability and School Achievement, II. Final Report on Cooperative Research Project No. 1636 (East Lansing, Michigan: Bureau of Educational Research Services, College of Education, Michigan State University, October 1965).

General Theoretical Framework for the Research

The theoretical background for the Michigan State research program on self-concept of ability and achievement derives from the symbolic interactionist theories of George Herbert Mead and Charles Cooley.[7] Essentially these theories hold that man's mind and self are a reflection of the society in which he lives. Man acquires his self-consciousness and his basic ideas and values in the process of interacting with "significant" people in his life who either explicitly or implicitly encourage certain kinds of behavior and discourage other kinds. "Significant others" in one's life are those persons on whom one is dependent for emotional gratification. In the process of interacting with these "significant others" an individual takes on the ideas, beliefs and evaluations of those "others" and uses them as a frame of reference for judging his own behavior. Such a "looking glass" process results in a self-evaluation or what is often called the "phenomenol self-concept" by current theorists. In order that one's self-evaluation may be positive the evaluations by significant others must be perceived as positive (though they may actually not be), and such positive evaluations are in turn the result of

---

[7]Charles H. Cooley, Social Organization (New York: Charles Scribners Sons, 1909); George H. Mead, Mind, Self and Society (Chicago: University of Chicago Press, 1934).

engaging in positively-regarded behavioral activities. Thus we come to behave as we are expected to by the significant others in our life, and we come to "see ourselves as others see us."

Applying this general framework to classroom learning, Brookover argued that achievement behavior is also subject to the same interactionist laws.[8] In this case, one learns (academically) what he is expected to learn by significant others; and perhaps more importantly, one's conception of his ability to learn is a learned evaluation also the product of significant interaction. More concretely, an individual learns the "new math" if he is expected to learn it by teachers, parents, etc. But the individual also learns a self-evaluation about his ability to do the new math--e.g. "It will be easy to learn," "I don't think I'll be able to learn it," etc. The amount of learning that takes place becomes functionally limited by the individual's self-definition of his ability to learn.

The "self-concept of ability" construct thus refers to those self-definitions held by individuals about their ability to learn. (In the present case, the concern is with ability to learn school work, but any kind of learning situation would be relevant.) The "self-concept of ability" is presumed to be only one aspect of a more

---

[8] Wilbur B. Brookover, "A Social Psychological Conception of Classroom Learning," School and Society, 87 (February 28, 1959), pp. 84-87.

global "self-concept" and is assumed to both influence and be influenced by the larger self. Similarly, "self-concept of ability" is presumed, in turn, to be influenced by and to influence more specific academic self-conceptions such as the ability to do the new math, English, social studies, etc.

From the interactionist theories one postulates that self-conceptions will be dynamic and thus have behavioral correlates. This does not mean that there will be a one-to-one relationship between self-definitions and behavior, but that there will be important communalities between the two. Further, it is assumed that self-conceptions are phenomonological--known to the individual--and that he can therefore report on his self to others if he wishes.

A major theoretical assumption of the Michigan State University research program is that self-concept of ability is a "functionally limiting" factor in grade achievement. This statement needs further clarification. The statement does not mean that there are no biological limits which operate in the learning process. Rather, the assumption is made that no one has ever approached the biological limits of learning and that we are all capable of learning many times what we actually learn. If this is so, then these functional limits to learning are of crucial importance for they are capable of change. Self-concept of ability is presumed to be such a functional limit; others may exist.

One might readily assume that the theory states that we cannot learn anything unless someone specifically expects us to learn it. Such a point of view would ignore the exploratory learning of children--e.g. putting everything in the mouth--which is certainly not systematically encouraged by parents. There are clearly many areas of potential learning where behavioral or attitudinal expectations by significant others simply do not exist because of ignorance, rate of social change, or simple indifference. Further research will be needed to specify the dynamics of spontaneous learning.

It should also be noted that we do not necessarily learn everything we feel we are capable of learning. A student may feel perfectly capable of doing advanced mathematics, but may decide this is not important to his career plans, or would interfere with his social schedule. A self-conception that one is able to learn becomes a necessary condition for learning, but is not sufficient to guarantee learning (see Hypothesis 2). There must still be some kind of desire or motivation to learn specific content.

When the theory is applied to school learning with actual achievement as a behavioral correlate, many would argue that a self-concept of (academic) ability scale would tap no more than our memory of past achievement experience, and thus any findings would be hopelessly circular. Past experience is admittedly a powerful force

and can act as a validating experience for our self-conceptions. For this reason we normally find a positive correlation between past and present achievement and would expect that both would correlate positively with self-concept of ability indicies. The same thing would hold for IQ test scores which are generally considered a sample of past achievement behavior--achievement in this case seen as broader than just class-room experience. Again a positive correlation between IQ test scores and self-concept of ability indicies would be anticipated. (For a more detailed discussion of IQ scores in the context of the present theory see Hypothesis 1.)

That past and present achievement, IQ scores, and self-concept of ability scores are typically positively correlated follows from the assumption that self-conceptions are dynamic and thus will have behavioral correlates. This is not, however, to say that the three constructs are reducible to one. Partialling out the effect of IQ or past achievement should not reduce the correlation of present achievement and self-concept of ability to zero. More importantly where there is a discrepancy between past performance or IQ scores and present achievement, it is hypothesized that the self-concept of ability construct can explain the difference. The discussion of "over-" and "under-achievement" under Hypothesis 1-B is a pertinent illustration. Viewing self-concept of ability as a functionally limiting factor in achievement does not obviate

the importance of past experience, or invalidate the traditional relationship of IQ to achievement. Rather the self-concept of ability construct helps explain why the traditional IQ-GPA relationship holds in general, and furthermore why it does not hold in particular situations. The theoretical base for understanding intelligence and achievement is thus broadened.

Other theoretical issues will be considered as required especially in the section on construct validity. While understanding of the theory is essential to the present research it should be remembered that the purpose of the research is not to test the theory, but an analysis of a measuring instrument. The theory, for such purposes, is presumed to be true.

The Instrument Evaluation

Section 1: Validity

Validity has to do with the question of whether we are measuring what we are supposed to be measuring. Implied in the question is the purpose for which the measuring device was created and the actual use made of the instrument. We may invalidly use for one purpose an instrument which is valid only for another. In the present case the question is whether we can meaningfully measure an aspect of self-conception which is predictive of academic achievement. The American Psychological Association outlines four types of validity--concurrent, construct,

content and predictive.[9] These are in effect four different ways of trying to determine if we have, in fact, measured what we claim to be measuring.

The first type of validity listed, concurrent validity, requires asking whether a given test correlates with an outside criterion in a manner comparable to that of other tests purporting to measure the same phenomenon. Thus, for example, a group IQ test might be compared with an individual IQ test in its ability to predict achievement. In the present case, there is no other known instrument which measures self-concept of (academic) ability, thus making this approach to validity impossible.

Construct Validity.--The determination of construct validity requires the analysis of test scores in terms of the related theory. To test the validity of a construct (in this case self-concept of ability) is to predict empirically from the theory in which the construct is imbedded. Assuming the theory is "true," what should be the practical consequences of that fact, and are our test results consistent with these predictions? If it can be shown experimentally that theoretically derived predictions are supported, then the validity of the construct as measured by the particular instrument is likewise supported. If the reader prefers the terminology commonly used in the philosophy of science, one can conceive of the deductive

---

[9]American Psychological Association, American Educational Research Association, and National Council on Measurement Used in Education, Joint Committee, op. cit.

procedure of predicting consequences from theory and the inductive process of stating that if certain consequences are true then the hypotheses which generated them may be true. For a complete discussion of both of these processes and the problems associated with the conclusions, the reader is referred to Braithwaite.[10] The establishment of construct validity requires theory building and hypothesis testing. A statement of the theoretical background for the construct has already been given; more detailed theoretical background will be given as required by individual hypotheses. It should be noted that the four major hypotheses of this study do not form an integrated unit, but rather constitute quite different approaches to the basic construct under study. It is hoped that the very divergence of focus will provide more complete understanding of the construct and instrument. Also, it should be pointed out that tests of construct validity are tests of the validity of the instrument under conditions of assumed theoretical accuracy. The same hypotheses could be tested for the purpose of testing the validity of the theory. Thus construct validation and theory validation are inextricably bound and can be separated only conceptually. Particularly in exploratory research the instrument and the theory move hand in hand aiding in mutual refinement and clarification. If the reader is not always sure whether the theory or the instrument is being examined, it is because the two processes are not empirically distinct.

---

[10]Richard B. Braithwaite, Scientific Explanation (Cambridge: Cambridge University Press, 1953).

Hypothesis 1: The self-images that junior high students hold of their ability are significantly and positively correlated with school achievement. 1-A: This relationship will be true if IQ (as measured by standard tests) is held constant. 1-B: This relationship will hold true even for a subgroup of "over-achievers" and "under-achievers" where there is a negative correlation between IQ and grade achievement.

If self-concept of (academic) ability is a functionally limiting factor in school achievement, then these two factors should be significantly and positively correlated. This follows from the theoretical assumption that self-conceptions are dynamic and thus influence behavior. Further it is assumed that people are motivated to see themselves and their behavior in a consistent fashion. Thus a student with a high self-concept of ability is motivated to achieve at a level consistent with this self-definition to "prove" to himself that he ought to maintain this high self-definition. At the other end of the line, the individual who feels he is not able is most unlikely to be motivated to expend energy to prove to himself that he is not what he believe he is. Initial self-definitions of ability, it has been suggested, are a result of internalizing perceived images of ability coming from significant others. Behavior becomes consistent with these images setting up a self-perpetuating cycle which is vicious for the low achiever although beneficial for the more successful. The above is a very general statement of the relation of self-concept of ability to

achievement; specific exceptions to the general rule will be suggested below.

The relationship of self-concept of ability is intelligence is by no means clear. Perceptual psychology and interactionist social psychology in combination allow at least one consistent explanation of the relationship. Perceptual psychologists Combs and Snygg state that intelligence is "a function of the factors which control the rightness, extent and availability of perceptions in the perceptual field."[11] Further, they state that intelligence is behavior which "effectively and efficiently satisfies the needs of the individual and his society."[12] Implicit in these two statements is the assumption that not all possible perceptions are going to be relevant to effective societal living. Perceptions are then selected by the individual. Extrapolating from the interactionist framework, one would say that the selection process is facilitated by the internalization of expectations for intelligent behavior which are held by significant others. Therefore, we behave as intelligently as we are expected to behave, or our perceptions are as right, numerous and available as they are expected to be. Put still another way, we behave as intelligently as is necessary to confirm the internalized expectations others have for us.

---

[11]Arthur W. Combs, and Donald Snygg, Individual Behavior (New York: Harper and Brothers, 1959), p. 214.

[12]Ibid., p. 213.

In regard to the more conventional definition of intelligence as that which is measured by intelligence tests, it would be argued that behavior on such a test is correlated with self-image to the extent that the individual feels that such behavior (answering test questions) is relevant to his internalized definitions of self. For the average individual one would expect a positive correlation between self-concept of ability and IQ. In the first place there is a general societal acceptance of IQ tests as accurate indicators of intelligence. Thus any student motivated to appear intelligent will be motivated to do as well as possible on such a test. Secondly, there is by now a rather general agreement that significant portions of all IQ tests are samples of past achievement. Since past experience is typically, although not inevitably, congruent with present behavior, there would also be an expected positive correlation between IQ and grade achievement.

While IQ and self-concept of ability would both normally be expected to correlate with GPA, it is expected that self-concept of ability would be more highly correlated with GPA than with IQ. This is largely because of the difference between the situations of taking an IQ test and taking a course exam. Self-concept theory assumes congruence between self-definitions and behavior--i.e. that an individual will expend energy to behave consistently with his self-definitions. On an IQ test it is difficult

to know how to expend energy that will make a difference in performance outcome other than being alert, positively motivated, and doing one's best. The expectations for the test are not clear; the individual is in fact systematically prevented from knowing in advance what he will be expected to know on the IQ test. In the classroom situation there is a clearer set of performance expectations either stated by the teacher or learned from past experience in the course, and there is the opportunity to expend energy--to study, review, memorize--to guarantee greater congruence between self-definition and behavior. Formal testing and classroom testing provide different problems for the self, the former situation probably the more frustrating. The distinction in the test situations is confirmed by Sarason and Mandler's findings that classroom and Scholastic Aptitude Test situations produce different types of anxiety.[13] They found that the SAT situation produced non-task-relevant anxiety due to the novelty of the situation, while task-relevant anxiety was produced in the classroom situation where "there is time for previously learned anxiety-reducing task-relevant responses to become operative."[14]

---

[13]Seymour B. Sarason and George Mandler, "Some Correlates of Text Anxiety," Journal of Abnormal and Social Psychology, 47 (October 1952), pp. 810-17.

[14]Ibid., p. 814.

While IQ and GPA are typically correlated by virtue of both measuring achievement, we know that there are many cases of so-called "over"- and "under-achievement." In reality, of course, these are incorrectly predicted cases. Any adequate theory must explain such typical exceptions as they relate to self-conceptions of ability. To retain theoretical consistency one must continue to assume that in these cases the basic axiom of self-concept as a dynamic force still operates. Therefore, typically the "over-achiever" would have a relatively positive definition of his ability to achieve and the "under-achiever" a relatively deflated image of his ability. In the situation where the individual has the opportunity to expend (or not expend) energy to effect the outcome, it is assumed he will behave consistently with the expectations for his behavior communicated by significant others.

Assume, for example, that the over-achiever is sensitive to the expectations of the classroom and perceives significant others to evaluate his performance and ability at a high level. He behaves accordingly. The IQ test situation, we have noted, lacks clear expectations other than motivation and attitudinal factors--it cannot be studied for. So there results a discrepancy between IQ performance and classroom performance. Over-achievers may simply represent a group of students who tend not to know things they have not been specifically instructed to study, or perhaps, those who really work to please specific

individuals. One might hypothesize that the over-achiever deliberately fails to try on the IQ test for fear of having to prove to himself that his self-concept of ability is inflated. This seems most unlikely. Most "over-achievers" would probably, in the absense of prior test score knowledge, assume that they would do well and might be surprised to find evidence to the contrary. Such evidence might have a depressing effect on the self, but more than likely the test score results would be ignored in the face of experience which proves high achievement possible. This would be especially likely if significant others continued to hold high level expectations.

In the case of under-achievers, it would follow that they perceive significant others to hold low images of their ability, and so they are not motivated to expend energy to study and earn success. In view of the high IQ score, the under-achiever's position is puzzling. It might be expected that the significant others would be prone to believe in the IQ scores and thus have high level expectations. It is, of course, possible that significant others do, in fact, hold high expectations for achievement, but these are not adequately communicated to the student. Or the individual might have a positive conception of his ability to learn but the significant others do not value achievement and so offer no rewards in this realm. Such a student puts his energies in another direction (sports, dates, etc.), and gets poor grades from lack of studying.

Much of the literature on under-achievement points to the existence of emotional problems in many of these students. It is possible to conceive of an under-achiever who has a relatively high image of his ability and yet who deliberately does poorly in school to "punish" significant others who have failed him. Or fear of failure may pose such a threat that an individual is unwilling to try to achieve at a high level. Poor achievement may also function an an unconscious attention-getting mechanism. The present theoretical framework has not been systematically articulated with self concept theory related to emotional problems, but presumably there would be no inconsistency. The present research assumes an emotionally "normal" student in its test of a molecular self-construct. The inability of the theory to adequately explain many cases of under-achievement attests to the limitations of molecular analysis and the long-range necessity of molar synthesis.

Hypothesis 2: A high self-concept of ability is a necessary but not sufficient cause for high achievement.

The present theory assumes that self-concept of ability is a functionally limiting factor in academic achievement. Thus it must follow that a high (positive) self-concept of ability is a prerequisite for consistently high achievement. The converse is not true--one does not necessarily have to get grades if he has a positive image of his ability, although this would normally be expected. It is conceivable that a student might have high confidence in

his intellectual skills and yet lack time, skills or interest in obtaining high grades. Coleman, for example, has suggested there are subcultural norms against high achievement in many high schools.[15] To be consistent with the theory, it would be expected that such individuals would have to validate their intellectual self-definitions from time to time to maintain them. Such a student might work hard in one subject ignoring the others, or perhaps hit the books for the final exams after "goofing off" the rest of the semester.

The test of the hypothesis that a positive self-concept of ability is necessary but not sufficient for high achievement involves an immediate problem that the present research cannot handle. It has been established that individuals vary in their tendency to belittle or inflate their skills according to their own personality needs. A self-report instrument, as a public affirmation of self, would be particularly vulnerable to such variation. Ideally some kind of scale would be administered to allow control of this variable, but this is not possible in the present case. The present research must then assume the veracity of the self-reported definitions of ability and look for a trend supporting the hypothesis. A strict test of the hypothesis requires expectation of zero cases

---

[15]James S. Coleman et al., The Adolescent Society (Glencoe, Illinois: Free Press, 1961).

of high achievement and low self-concept of ability; realistically there are insufficient controls to expect other than trend data.

The test of a trend in the direction of the hypothesis will involve looking at individuals who deviate from the expected self-concept and GPA relation. There should be significantly fewer individuals who are high in GPA and low in self-concept of ability than there are individuals low in GPA and high in self-concept of ability. Chi Square analysis can be used to test this assumption. An immediate problem involved in such a test is that we have no way of estimating the likelihood of having low grades if one has a high self-concept of ability. This is clearly unexpected behavior in the context of general theory, although it is still presumed more likely than high grades and low self-concept. The actual frequency of such behavior may prove a more interesting finding than the test of the hypothesis.

While Hypothesis 2 is stated as a test of a cause and effect relationship, it should be noted that, strictly speaking, this is not true. As Wylie has pointed out, if we assume that a self-concept variable is an antecedent to some behavior, "we are reduced to response-response correlational designs where both responses are obtained from the subject in the same study. In such a case, although plausible cause-effect inferences may be made, we can never claim to have demonstrated a cause-effect

relationship unequivocally."[16] In the present case, we assume that a given self-definition of ability acts as a stimulus leading to a particular response which we denote as achievement behavior. In fact, however, we cannot observe the stimulus but must rely on a report (i.e. another response) made by the subject about the stimulus. Thus we have a response-response design and can conclude with certainty only that the two responses are correlated. At the theory level, cause and effect are appropriately designated and carefully designed experiments will produce a cumulative body of results which is "certain enough."

Hypothesis 3: The self-images that junior high students hold of their ability in specific subjects areas are significantly and positively correlated with school achievement in the parallel subject.
3-A: These relationships will hold true even if IQ (as measured by standard tests) is held constant.

If self-concept of (academic) ability is a functionally limiting factor in general achievement, what is its relationship to achievement in specific subjects? The fact of differential subject area achievement is obvious; the possibility of differential self-images is not. The present theory postulates that such differences in achievement exist because there are in fact differential images of ability in different subject areas. The source of these

---

[16]Wylie, op. cit., p. 19.

differential images is again the internalized expectations of significant others. The specificity of the subject area self-conceptions is undoubtedly cultural. Our society compartmentalizes knowledge into certain subjects and makes evaluations as to the importance of learning these subjects. A similar set of arbitrary and specific self-images is evident in the term "athlete." While most young men probably have a general image of their skills as an "athlete," they are likely to also hold specific self-conceptions of their skills as football players, baseball players, basketball players, etc. It is assumed one's self-image as an athlete is somehow a composite of these various specific-sport images; in analagous fashion general self-concept of ability is presumed to be some kind of composite of specific-subject self-conceptions. The exact nature of the relation of specific to general images is not specified but will be explored.

The test of Hypothesis 3 will have a dual function. On the one hand it will allow examination of whether specific subject self-conceptions of ability operate analagously to the more general self-concept of ability. Is, for example, the relationship of IQ analagous? Secondly, examination of the specific subject images of ability can serve to establish that the specific and general images of ability are not reducible to each other, and, assuming they are distinct constructs, how they are, in fact, related. Any construct is theoretically capable of infinite

refinement or generalization, but the scores on a given measuring device have a much narrower referent, the boundaries of which need to be specified so that research findings can be meaningfully interpreted. For example, are scores on the general SCA Scale equivalent to an average of the specific subject SCA scales? Can the general scale predict the specific subject scale scores? Hypothesis 3 will aid in answering such questions. Unfortunately, putting a boundary on the generalization end of the construct (as tapped by a given measuring device) is not possible in the present case. Ideally, one would like to know how self-conceptions of academic ability relate to self-conceptions of ability in other areas of life, or to more comprehensive definitions of "total self." Other scales could be devised to examine this problem.

Hypothesis 4: The Specific Subject Self-Concept of Ability Scales will be significantly better predictors of achievement in the parallel subject than will be the more general Self-Concept of Ability Scale. 4-A: The general Self-Concept of Ability Scale will be a significantly better predictor of achievement in a specific subject than will any Specific Self-Concept of Ability Scale other than the one in the corresponding subject. 4-B: The general Self-Concept of Ability Scale will be a significantly better predictor of general achievement (GPA) than will any Specific Subject Self-Concept of Ability Scale.

For convenience, the above hypotheses are stated in terms of the measuring scales used in their test; the specific

nature of these scales is discussed in Chapter III. The general constructs to which these scales refer should be clear from the discussion of the third hypotheses.

The tests under Hypothesis 4 are further attempts to establish a partial "boundary" for the general self-concept of ability construct. At the same time they are used to establish that the specific subject self-concepts of ability are distinctive and cannot be used interchangeably with the more general construct. If the specific subject constructs are psychologically meaningful they should be better predictors of specific subject achievement than the general construct. This, of course, can be empirically tested only if differential achievement patterns are actually found. Research has shown that differential achievement may be related to personality factors. For example, "stereopath" students are more likely to have uniform achievement patterns than "non-stereopath" students.[17] Such personality controls are again not possible in the present research, but control for differential achievement patterns may be required.

Hypotheses 4-A and 4-B are further attempts at validation of the general self-concept of ability construct by a process of eliminating alternative hypotheses. Thus the scale measuring the general construct should correlate more highly with general than specific achievement, and

---

[17]George Stern, M. I. Stein, and B. S. Bloom, Methods of Personality Assessment (New York: Free Press of Glencoe, 1956).

the specific subject scales should be more highly related to parallel specific subject achievement than to general achievement.

Content Validity.--The establishment of content validity requires that one show how the content of the test or inventory samples the subject matter about which conclusions are to be drawn. In the present case, how adequately do the items in the Self-Concept of Ability Scale sample the construct self-concept of ability? There is rarely quantitative evidence of content validity; rather, the criteria for item selection become the focus for establishing content validity. How comprehensively is the construct dimension sampled, and what evidence is there that all items belong to the same dimension?

The logical analysis of item content may also be used to offer supportive evidence of content validity. Close inspection of the items in the Self-Concept of Ability Scale suggests that the Scale may contain subsets of items. Such a finding would be evidence against the unidimensionality of the scale and would suggest the need for analysis of the subdimensions of the self-concept of ability construct. Are there, for example, certain aspects of the construct which are more highly correlated with the total score or criterion measure? Examination of the internal consistency of the measuring instrument will offer evidence on such issues.

Predictive Validity.--The establishment of predictive validity requires prediction from a test of some outcome. The actual outcome is then measured at some future time and the results compared with the predicted outcome. A simple index of predictive validity in the present case is the correlation between scores on the Self-Concept of Ability Scale (administered in the Fall of 1960) with the grades obtained in January 1961 or the following June. More appropriately, however, predictive validity requires "cross-validation"--using the prediction equation derived from analysis of one sample to predict the scores for a new sample. This constitutes evidence that the results obtained in the initial analysis were not peculiar to that sample, but will generalize across populations. In the present case, the calculated prediction equation employed a combination of IQ and Self-Concept of Ability Scores to allow comparison of their relative effectiveness in predicting achievement. The relationship of IQ to self-concept of ability has already been discussed.

Section 2: Reliability

While validity measures ask if the instrument measures what it is supposed to measure, reliability is concerned with the consistency of the measuring instrument over time, compared with other instruments, or internally. It is essential to know the consistency or accuracy of a test, for this becomes a limiting factor in the establishment

of validity. (The correlation between a test and an independent criterion can never be higher than the square root of the reliability of the test.)

As in the case of validity, there are many kinds of reliability which ask somewhat different questions and thus give us different dimensions of consistency. The American Psychological Association test recommendation booklet distinguishes three types: stability reliability, internal consistency reliability, and equivalence reliability.[18] Equivalence reliability asks whether two forms of a test are equally able to measure a given phenomenon. The Self-Concept of Ability Scale exists in only one form so determination of equivalence reliability is impossible.

Stability Reliability.--Stability reliability asks whether the same score would be received on a test given at a later time such as an hour later, two weeks later, six months later, a year later, etc. It is clear that this is a theory-tied matter for we would often not expect stability. Achievement test scores, for example, are not expected to be stable from the beginning to the end of a semester. In the present case, one must specify whether self-concept of ability would be expected to be stable over time.

General self-concept theory would indicate that, on the whole, self-concept should be relatively stable. If

---

[18] American Psychological Association, American Educational Research Association, and National Council on Measurement Used in Education, Joint Committee, op. cit.

self-concept acts as a gyroscope for the personality, then the individual is motivated to maintain consistency and stability of self-definition. This is not, however, to take the position that self-concept does not change. What is maintained is that the average individual's self-concept will remain relatively more stable than changeable. As self-concept of ability is only one of many self dimensions, it may be that these more narrow self definitions are subject to greater fluctuation over time. Only empirical evidence can establish the probability of stability over time. However, for present purposes, it will be assumed that self-concept of ability is in general stable over time. Therefore, the correlation between two administrations of the Self-Concept of Ability Scale will be a test of the stability reliability of the instrument, not a test of the stability of the self-concept of ability over time. That a dual interpretation will be tempting must not obscure the fact that these are logically different issues. Unfortunately, failure to demonstrate high stability reliability will be difficult to interpret in view of the possible theoretical error.

Internal Consistency Reliability.--An assessment of internal consistency reliability asks about the homogeneity of individual items in the test. Do all the items measure the same thing and to the same degree? Three different approaches to internal consistency reliability will be considered.

The first type of internal consistency reliability, often not considered in test analyses, is Guttman scaling. If the items on the test form a Guttman-type scale with adequate "reproducibility," then the items are assumed to be a sample from a unidimension, and in addition, to form a pattern such that the order of responses is predictable. The items in this case need to be homogeneous with respect to content, but will not be of equal difficulty.

Secondly, internal consistency can be determined by the Hoyt method of analysis of variance. This allows one to determine if the ratio of error variance to individual variance is appropriately small. If all the items are of equal intercorrelation and difficulty (assumptions of the Hoyt method), then, in theory, the only variance in scores comes from the differences in the individuals answering the items. In practice these assumptions are almost never met, but one can still estimate the item, individual, error and total variances and examine the proportion of error variance to individual variance.

Thirdly, a correlation matrix of each item with every other item and the total score will allow inter-item comparisons and influence on the total score. In particular, the possibility of subscales will be examined both theoretically and empirically. Theoretically, subtotals for sets of items which logically appear to ask slightly different questions (see content validity above) will be

compared with each other and the total score for possible variation. Empirically, cluster analysis will be used to show if subscales exist. Cluster analysis uses an inter-item correlation matrix and asks whether certain items are more closely intercorrelated with each other than with other items.

It has already been indicated that internal consistency measures will offer evidence of content validity; in addition they will provide predictive validity data for individual items. These analyses will also have implication for construct validity. In the present case, the absense of subscales would confirm the unidimensionality of the instrument and thus offer evidence that only a single construct needs to be postulated to account for the obtained results.

Limitations of the Present Research

In addition to the limits imposed by the post facto nature of the research, other limitations should be noted. The general framework for the research has required two important theoretical-methodological assumptions, which, if invalid, would certainly shed doubt on the validity of the research. The first assumption is that self-concept of ability (and self-concept in general) is phenomenological--i.e. known to the subject. This assumption allows the possibility of using a self-report instrument rather than having to rely on projectives as is required by the

assumption of a nonphenomenological or unconscious construct. The assumption draws support from Wylie:[19]

> . . . there is as yet no proff that one can predict behavior as well, let alone better, with unconscious-self-concept measures than with conscious-self-concept measures. The state of validation of unconscious-self-concept measures is even more parlous than is the state of validation of conscious-self-concept measures. Therefore the burden of proof is presently on the person favoring the addition of the unconscious self concept to the variables from which we try to predict behavior. Although it seems quite plausible that phenomonological theories could become more predictive by the addition of constructs concerning the nonphenomenal self, our point here is that this has not been demonstrated with the indices we now have.

The second assumption is that there is an important degree of correspondence between the phenomenal self-concept of ability, and responses made by subjects on the self-report instrument. Combs[20] has cast doubt on this assumption at least for very young children. One of the functions of the present research will be to examine this assumption. If a theory operating under the assumption of "true" self-concept development correctly predicts behavior obtained by the use of a self-report instrument, then there is reason to believe that the "true" and reported self-definitions are highly correlated. That any

---

[19] Wylie, op. cit., pp. 319-20.

[20] Arthur W. Combs, Daniel W. Soper, and Clifford C. Courson, "The Measurement of Self-Concept and Self Report," Educational and Psychological Measurement, 23, No. 3 (Autumn, 1963), pp. 493-500.

self-report instrument would fail to tap self-concept in its full complexity is obvious. But the opposite position that what people tell us has no correspondence to our "true" feelings is equally untenable and will often lead us to overlook valuable data precisely because it is so easy to get.

Although the present research has focused on testing four hypotheses which are relevant for the establishment of construct validity, it should not be presumed that such correlational studies constitute the preferred or only methods of construct validation. Chapter II, Section 2, discusses four approaches to construct validation and indication is given of how the specific hypotheses tested illustrate these approaches. It will be clear in this discussion how far the present research falls short of an optinial evaluation program. Probably the most serious deficiency in the present research is the inability to show that the obtained results are not dependent on a particular methodological procedure. Ideally, as will be pointed out, one should set up an intercorrelation matrix which would show different methods designed to measure the same construct (e.g. questionnaire, interview, projective) and these same methods used in the measurement of different constructs. Correlations should be higher among different measures of the same construct than among the same method indexing different constructs. The present research does

not allow such comprehensive analysis as only one method has been used to index self-concept of ability.

Some of the best evidence for the validity of the self-concept of ability construct comes not from this dissertation but from a series of experiments done concurrently with the present research and reported in the report of Project #1636. This material is summarized in Chapter II where its relevance to the problem of construct validation is indicated. By its very nature, the process of construct validation involves the continual accumulation of supportive data and the concommitant refinement of theory and methodology. The evidence reported in this dissertation is relevant only for the present stage of methodological and theoretical development. One would certainly hope that the construct "self-concept of (academic) ability" proves of sufficient theoretical importance that the present review of research evidence in support of the construct is rapidly outdated.

CHAPTER II

RELATED LITERATURE

The literature directly related to the present research is limited both by the nature of the study (evaluation of an instrument) and by the frontier nature of the content area. The evaluation of a measuring instrument involves certain conventions sufficiently well known that they deserve only the most general reference--in this case to the American Psychological Association technical recommendations.[1] However, even among test experts the process of construct validation remains a frontier area not automatically employed by test evaluators and certainly not obvious in the application of its methodology to the evaluation of a particular construct. Therefore, the methodology of construct validation will be discussed separately in Section 2 of this chapter with emphasis on how the present research illustrates the various approaches to construct validation. Section 1 will briefly review the theoretical sources used in the derivation of the self-concept of ability construct, and empirical studies employing the construct.

---

[1]American Psychological Association, American Educational Research Association, and National Council on Measurement Used in Education, Joint Committee, op. cit.

Section 1: Theoretical Origins and Empirical Test of the Construct

The general theoretical framework for this research has already been identified in Chapter I as deriving from the symbolic interaction psychology of George H. Mead and Charles Cooley.[2] A concise restatement of this theory in symbolic form has been made by Kinch;[3] another recent summary of general symbolic interactionist theory may be found in Rose.[4] The specific application of this framework to classroom learning was made by Brookover in 1959.[5] (An expanded statement of Brookover's position may be found in the revised edition of A Sociology of Education.[6])

In the 1959 article, Brookover states four basic propositions relating the symbolic interactionist framework to the process of classroom learning:[7]

1. Persons learn to behave in the ways that each considers appropriate to himself.

---

[2]Cooley and Mead, op. cit.

[3]John W. Kinch, "A Formalized Theory of the Self-Concept," The American Journal of Sociology, 68 (1963), pp. 481-86.

[4]Arnold M. Rose, "A Systematic Summary of Symbolic Interaction Theory," in Rose (ed.) Human Behavior and Social Processes (Boston: Houghton-Mifflin, 1962), pp. 3-19.

[5]Brookover, 1959, op. cit.

[6]Wilbur B. Brookover and David Gottlieb, A Sociology of Education (2d ed.; New York: American Book Company, 1964).

[7]Brookover, 1959, op. cit.

2. Appropriateness of behavior is defined by each person through the internalization of the expectations of significant others.
3. The functional limits of one's ability to learn are determined by his self-conceptions or self-image as acquired in social interaction.
4. The individual learns what he believes significant others expect him to learn in the classroom and in other situations.

From these propositions it follows that self-concept of ability to learn becomes the immediate determiner of actual achievement. Further, if the quality of interaction with significant others varies, so self-images vary, and thus behavior is changed. Applied to classroom learning, different levels of achievement may be expected from the same individual over time if there are changes in the quality of the interaction with significant others. The actual biological limits of ability to learn are not known, but are assumed by Brookover to be rarely, if ever, approached. The notion of a fixed level of intellectual ability is clearly incompatible with the above propositions and so society is provided with a potentially unlimited resource of intellectual power practically determined by socialization experiences.

As a test of the above propositions, a research program was set up at Michigan State University under the direction of Dr. Brookover and supported by grants from

the U. S. Office of Education. Two project reports have already been published--the reports for Projects #845 and #1636; a third is forthcoming.

Project #845 constituted the exploratory phase of the Michigan State research program. It was designed to determine the feasibility of tapping the self-concept of ability construct with a self-report instrument, and to test three basic hypotheses. The specific hypotheses are as follows:[8]

1. The self-concepts of high achievers among junior high school students with similar levels of intelligence as measured by standard tests vary significantly from the self-concepts of low-achievers.
2. Students' self-concepts of ability in specific school subjects vary both from one subject to the other as well as from their general self-concepts of ability.
3. The expectations of significant others as perceived by junior high school students are positively correlated with the students' self-concepts as learners.

Project #845 also examined other related questions such as: Who are the significant others of 7th grade students?;

---

[8]Brookover, Paterson, and Thomas, op. cit., p. 5.

How is self-concept of ability related to intelligence, sex, and socio-economic status?; How do significant others of 7th graders vary by sex, achievement level and socio-economic status?; and How effectively can measures of self-concept of ability predict school achievement?

Positive support for the Project #845 hypotheses lead to the design of three controlled intervention experiments the results from which are published in the report of Project #1636.[9] This report also contains longitudinal data on associated changes in self-concept of ability, achievement, and perceived images of ability held by significant others.

The three controlled experiments were designed to investigate the feasibility of enhancing the self-concepts of ability of low-achieving junior high students and then observing the long-range effect on academic achievement. Consistent with theory, self-concept enhancement was to be done by modifying the quality of the interaction between the students and significant others. As parents had been found (in Project #845) to be significant others for all students, one experiment involved working with the parents of low-achieving students. A Michigan State University research team worked with the parents over a period of a year in group and individual sessions explaining the basic

---

[9]Brookover and others, op. cit.

propositions of learning and attempting to show the parents how they could communicate more positive expectations for achievement to their children. No contact was made directly with the students. A placebo group of parents was set up which likewise met over the span of a year but focused on the general problems of junior high school students. A control group was identified but not contacted. As in the experimental group, no contact was made directly with the students in the placebo or control groups.

The second experiment introduced an outside "expert" to a group of low-achieving students in a second school. The "expert" discussed in group sessions the basic propositions of the learning theory and attempted to convince the students that they should enhance their own images of their ability and thus their achievement. The placebo group of students for this second experiment met with the same "expert" but the discussion was about the general problems of junior high school. A control group was identified but not contacted. Although the "expert" was not a "significant other" to the students, the purpose of this experiment was to determine if the "expert's" views would have sufficient impact that they would, in fact, be internalized. It was possible that a "significant other" is sufficient but not necessary for inducing change in self-concept of ability.

The third experiment, in a third school, introduced a Michigan State University counselor to a group of low-achieving students with the hope that he would, through

close, personal association, become a "significant other" for these students. The counselor communicated a personal positive conviction about each student's ability which, it was hoped, would be internalized. No placebo group was used in this experiment, but a control group was identified.

Results from the three experiments showed significant increases in self-concept of ability and in achievement for those students whose parents participated in the experimental group of the first experiment. These students also reported significantly higher perceptions of parents' images of their ability. There were no significant changes in the students whose parents were in the placebo group of the first experiment or in the control group. No significant changes in self-concept of ability were found as a result of working directly with the students in the second and third experiments, or in the students identified in control groups for these experiments. The results are taken as support for the interactionist framework. Although it might appear that the third experiment should have had positive results, it should be noted that these students all reported that their parents held low images of their ability. Even if the counselor had indeed become a "significant other" (a conclusion which was not fully demonstrated), his views were still in competition with those of the students' parents and the latter might well be considered to carry more weight. The relevance of the findings from Projects #845 and #1636 for the establishment of construct validity

of the Self-Concept of Ability (SCA) Scale is noted in Section 2.

The body of research from the Michigan State studies (which include several unpublished papers and dissertations) constitutes the only known empirical research employing the construct "self-concept of ability." A few studies do, however, show correlations between various measures of global self-concept and various indicies of achievement.[10] There is even a paucity of empirical evidence supporting the general symbolic interactionist framework although the "Related Research" section of the report of Project #1636 cites several studies which support the relationship of self-definitions to evaluations made by others.[11] Ruth Wylie's survey of self-concept research reviews over 400 empirical studies, yet when considering Mead and Cooley's theories on the relationship of self-concept to social interaction she can only conclude that "this theoretically crucial class of relationships has been inadequately explored."[12]

---

[10]See, for example, G. A. Renzaglia, "Some Correlates of the Self Structure as Measured by an Index of Adjustment and Values" (unpublished Ph.D. thesis, University of Minnesota, 1952); Thelma Adams Reeder, "A Study of Some Relationships Between Level of Self-Concept, Academic Achievement and Classroom Adjustment" (unpublished Ph.D. thesis, North Texas State College, 1955); and J. W. Staines, "Self-Picture as a Factor in the Classroom," British Journal of Educational Psychology, 28 (June 1956), pp. 97-111.

[11]Brookover and others, op. cit., pp. 16-28.

[12]Wylie, op. cit., p. 136.

Section 2: The Methodology of Construct Validation

The process of construct validation, according to the American Psychological Association technical recommendations, involves demonstrating "that certain exploratory constructs account to some degree for performance on the test."[13] The essential process is theory testing, or more explicitly, empirical testing of theoretically-derived hypotheses. The test of the theory and the validating of the construct are thus not empirically distinct processes. Construct validity focuses on the traits or qualities underlying an instrument rather than on the content of the instrument or the nature of the criterion as is true with content or predictive validity. "Construct validity is ordinarily studied when the tester has no definitive criterion measure of the quality with which he is concerned, and must use indirect measures to validate the theory."[14]

There is no single method of construct validation, but rather one engages in a series of "converging operations" which Garner, Hake and Eriksen define as "any set of two or more experimental operations which allow the selection or elimination of alternative hypotheses or concepts which could explain an experimental result. They are called converging operations because they are not perfectly correlated

---

[13]American Psychological Association, American Educational Research Association, and National Council on Measurement Used in Education, Joint Committee, op. cit.

[14]Ibid., p. 14.

and thus can converge on a single concept."[15] It is the systematic integration of multiple sources of evidence which is required because of the vagueness of the construct and the fragmental nature of psychological theory at a given stage of development. Construct validation then serves to clarify and refine theory at the same time as it evaluates measurement devices. Because of the incomplete construct, there can be no single behavioral criterion which will validate a construct; one can only postulate behavior-relevant criterion measures recognizing that they are not behavior-equivalent measures.[16] The correlation between a measuring device and a behavior-relevant criterion does not therefore constitute a validity coefficient, but rather should be interpreted as evidence of a "converging operation."

More specific suggestions for the procedures to be followed in construct validation come from Cronbach and Meehl, Campbell and Fiske, and Wylie.[17] An integration of

---

[15]Wendell R. Garner, Harold W. Hake and Charles W. Eriksen, "Operationalism and the Concept of Perception," Psychological Review, 63 (March 1956), pp. 149-59.

[16]American Psychological Association, American Educational Research Association, and National Council on Measurement Used in Education, Joint Committee, op. cit., p. 15.

[17]Lee J. Cronbach and Paul E. Meehl, "Construct Validity in Psychological Tests," Psychological Bulletin, 52 (July 1955), pp. 281-302; Donald T. Campbell and Donald W. Fiske, "Convergent and Discriminant Validation by the Multitrait-multimethod Matrix," Psychological Bulletin, 56 (March 1959), pp. 81-105; Wylie, op. cit.

these suggestions by Wylie suggests four different approaches to construct validation which are outlined below. Because the relevance of these approaches to the construct under investigation may not be obvious, the various approaches will be illustrated by the present research where possible, or by hypothetical studies.

1. Analysis of Contaminating Variables Approach. Under this approach are included observational and mathematical analyses of variables other than the construct in question which might be influencing the results. Wylie suggests nine such variables: (a) social desirability, (b) content area, (c) known identity of the subject, (d) lack of rapport, (e) instrument form, (f) degree of restrictiveness of subject's responses, (g) set or expectation, (h) response frequency, and (i) scoring or statistical procedures.[18] The relevance of several of these "contaminators" to the measurement of self-concept of ability is clear.

(a) Social desirability undoubtedly influences responses to the SCA Scale items. One item, for example, asks "How do you rate yourself in school ability compared with your close friends?" The response categories are "I am the best," "I am above average," "I am average," "I am below average," and "I am the poorest." The most socially acceptable responses to this item would certainly

---

[18]Wylie, op. cit., pp. 27-36.

be the middle three. It would violate the norm of modesty to declare oneself the best; it would be equally inappropriate to sell oneself short. The democratic ideology of our culture suggests that we should not make invidious distinctions among persons thus further encouraging middle-of-the-road responses. Another item undoubtedly influenced by social desirability asks students what grades they feel they are capable of getting. This question certainly invites a response at least one level above current performance--ours is a society dedicated to the notion of meliorism.

To know that certain response categories are more likely selected because of social desirability does not, of course, automatically invalidate their use. In fact it is inconceivable that any self construct would exist without some loading on social desirability as self-definitions are the product of social interaction. The basic interactionist hypothesis suggests that self-definitions are internalized definitions of desirable (or undesirable) behavior learned in interaction with significant others. It might be possible to control for social desirability in self-definitions by asking subjects to indicate their own or significant others' attitudes toward various topics such as the desirability of good grades or comparing oneself with others. Only if there were considerable agreement among such attitudes would one be justified in using group averages in correcting items for social desirability;

otherwise individual corrections would have to be made. Even if one could control items for social desirability, it would be almost impossible to know if this process would lessen error variance in the measuring instrument, or merely take out part of the "true" variance involved in self-definitions. The measurement of self conceptions is indeed a complex process.

(b) The accuracy of the report of self-definitions is certainly related to the content area receiving attention. Even assuming completely honest responses, subjects will vary considerably in the extent to which they have thought about or are willing to reveal certain self dimensions. Jourard and Lasakow report, for example, that subjects will voluntarily reveal more about their attitudes, opinions, tastes and interests than about their personality or body characteristics.[19] Some students will no doubt have spent a great deal of time thinking about their relative intellectual ability; for others the question has never before been raised. Some students may have considered their math ability but not their literary skills; some may be willing to talk about their conclusions, but others will be reluctant. It should be possible to construct some sort of instrument which would reveal degree of certainly about, or sensativity to various self dimensions.

---

[19]Sidney M. Jourard and Paul Lasakow, "Some Factors in Self-disclosure," Journal of Abnormal and Social Psychology, 56 (January 1958), pp. 91-98.

(c) The preservation of the anonymity of the subject seems to be important in most social-psychological research although the effect of not doing so seems to have received little attention. One would hypothesize that failure to guarantee anonymity would result in selection of more socially desirable responses. In the Michigan State University research students were guaranteed that only the research team would see the questionnaire responses. It might be instructive to know if they would feel differently about information being revealed to their parents, peers, or teachers as against the University staff.

(d) The issue of establishing good rapport is important both in individual testing and in group administration of tests. The Self-Concept of Ability Scale and other instruments were administered in the auditoriums of the various schools involved, and the setting of the school itself could well have influenced the attitudes with which the students approached the questionnaires. Even more important in all probability were the students' attitudes toward the University as a community institution and the impression of University personnel gained from observation of the test administrators who were on the Michigan State University research staff. Whether the staff appeared aloof, sincere, "ivory tower," or "good guys" would certainly affect the seriousness with which the students approached their task.

(e) The form of the instrument may contaminate the measuring process by leading the subject to respond to one end of a response range more frequently than the other, by confusing the student in its directions, by inviting contrast in responses, etc. Each instrument has its own sources of error due to form the consequences of which must be determined empirically. The SCA Scale is known to invite more responses in the positive categories than negative ones. This may be a result of item content, or perhaps an automatic tendency to check in the top half of a range. Reversing the order of responses might show up this bias. The specific subject scales by their format undoubtedly encourage subject matter differentiation when in fact such differentiation may not be psychologically relevant. It is known that females are more likely to make subject matter discriminations than males even though they achieve more uniformly, suggesting that the effect of instrument form needs to be explored separately for each sex at least with the present instrument.

(f) The degree of restrictiveness of response alternatives has been studied particularly in its possible biasing effects on results from Q-sorts and forced-choice items. The rating scale is not immune from bias especially if the response alternatives do not correspond to any psychological reality. What would be the result, for example, of taking the "average" response category used in several items and breaking it down into "pretty

average but a bit above," "average," and "pretty average but a bit below?" Clearly we know very little about our internal self-defining scales.

(g) Set or expectation may bias results by leading the student to respond to questions on the basis of mood or attitude rather then true feeling. Testing in the school setting may invite responses on the SCA Scale which are perceived to correspond to teacher images, or may engender automatic positive or negative attitudes toward the test-taking task. Taking the questionnaire right after receiving report cards would also be likely to influence mental set--in some direction.

(h) Response frequency may bias responses in test-retest situations by allowing the subject to recall a former response rather than rethinking the question in each test situation. Longitudinal studies employing the SCA Scale may introduce this irrelevant response determiner, particularly in older subjects.

(i) Finally, bias in the measuring process which comes from scoring and statistical procedures may be influencing research results. Scores on the SCA Scale are obtained by assigning numerical values to item response alternatives and then summing the item scores. Conventional parametric statistics are then used in analysis of the results. Such a procedure assumes equal-interval data; one is not warranted in assuming, however, that the response alternatives are even remotely "equidistant" in

the minds of the respondents. The issue of the psychological relevance of scores or score ranges is one which will be raised at many points throughout this dissertation.

To establish the psychological relevance of the SCA Scale scores will involve some procedure of determining each individual's definition of the meaning of item response alternatives. One technique often used in psychology is to present descriptions of hypothetical individuals or behaviors and ask the respondent to "label" or characterize the individual or behavior. In the present case one might ask about a student who gets all A's and B's--would such a student be viewed by the respondent as "one of the best," "above average," etc. It is assumed that the respondent would use the same criteria in describing himself. Knowing each respondent's definition of a "good student", etc. provides a basis for assuming that responses across respondents have comparable psychological relevance.

2. Intercorrelation Approach. Among the operations which may "converge" on clarification of a construct are intercorrelation studies. Campbell and Fiske suggest a dual approach of examining the correlations of different methods designed to measure the same trait or construct and comparing these with the same method used to measure different traits or constructs.[20] Their article demonstrates this procedure with a variety of traits culled

---

[20]Campbell and Fiske, op. cit.

from psychological literature. Different methods in the present case might include questionnaires, interviews, projectives, and observational techniques all designed to index self-concept of ability. Different constructs which might be compared would include general self-concept (in the total personality sense), self-concept of ability, self-concept of ability in specific subjects, athletic ability, etc. Campbell and Fiske suggest that ideally correlations using different methods to measure the same construct should exceed (a) correlations using the same method on different constructs, and (b) correlations of scores using different methods on different constructs.

One difficulty in comparing methods is the inability to know if, in fact, the different methods are tapping the same construct. Combs, for example, suggests that observational inferences about self-concept do not, on the average, correlate significantly with self-report statements made by young children.[21] Combs did not, however, demonstrate in his study that the self-report and observational inferences were comparably oriented to the same construct.

The present research offers no opportunity to compare methods as Campbell and Fiske suggest. Although Project #845 included interviews with a subgroup of students who had previously responded to the SCA Scale, no attempt was

---

[21]Combs, Soper, and Courson, op. cit.

made during the interview to arrive at an independent assessment of self-concept of ability.

Data from Project #845 do allow comparison of correlations of the same method used to index different constructs. One scale in Project #845 entitled "Importance of Grades Scale" was designed to determine if self-concept of ability could be distinguished from attitudes towards the importance of doing well in school. The seven items used formed a Guttman Scale with reproducibility only slightly lower than for the SCA Scale. The two scales correlated with each other .46 (males) and .33 (females). More significantly, however, the SCA Scale correlated with the criterion of grade point average .57 for each sex, while the Importance of Grades Scale correlated with GPA only .27 (males) and .14 (females).

The comparison of the general SCA Scale and specific subject scales to predict achievement (see Hypotheses 3 and 4) also controls for method while varying the construct under consideration. Such examples are not strictly in the spirit of the Campbell and Fiske directive as they involve a comparison of correlations with a relevant criterion rather than a comparison with different methods of indexing the same construct. While this ideal procedure for construct validation is not employed, the concern with delineating the unique aspects of the self-concept of ability construct seems clearly within the spirit of construct validation. In particular, Hypotheses 1 and 3 are both

concerned with controlling IQ so it will be quite evident that the construct under investigation is not merely another intelligence measure. Hypotheses 3 and 4 are likewise concerned with establishing that general self-concept of ability is not reducible to self-definitions of ability to do work in the specific subjects encountered in school. The hypotheses thus serve to discriminate the self-concept of ability construct (as measured by a given instrument) from other constructs which might appear to be no different.

3. Internal Analysis Approach. Item analysis and factor analysis of measuring instruments are the essence of this approach. While neither can prove that a given construct has been tapped, Cronbach and Meehl feel that such procedures may throw light on the number of basic processes which must be postulated to account for the responses obtained on the instrument as a whole.[22] In particular one may be aided by knowing if the instrument is in fact unidimensional. Scalability, in the Guttman sense, is one test of unidimensionality; cluster analysis is another. Both procedures were used in analysis of the SCA Scale in addition to other item analyses.

4. Predictable Correlates Approach. In the absense of "ideal" validating criteria, Cronbach and Meehl suggest one may engage in studies which predict from theoretical premises the relationship of a construct to other

---

[22] Cronbach and Meehl, op. cit.

variables.[23] As all the hypotheses in the present research involve the criterion of academic achievement, they all serve to illustrate this approach. Cronbach and Meehl are, however, particularly concenred about studies employing a construct which (a) correctly predicts group differences, and (b) predicts correctly change over time especially after controlled experimental intervention. Hypotheses 1-B and 2 are examples of predicting group differences. In Hypothesis 1-B differences between "over"- and "under-achievers" are predicted from self-concept theory; in Hypothesis 2 high and low self-concept groups are likewise used in the prediction of nontypical achievement patterns.

Prediction over time with controlled experimental intervention is beautifully illustrated by the three experiments reported in Project #1636 which have already been summarized in the first section of this chapter. Results from these experiments clearly illustrate both that change in self-concept of ability is correlated with achievement, and the conditions under which changes in self-concept of ability are likely to occur. Project #1636 also offers another example of longitudinal study but without experimentation. Over a three-year period the associated changes in three variables were examined--perceived images of ability held by significant others, self-concept of ability, and achievement. It was hypothesized that changes in the

---

[23] Ibid.

perceived images of ability held by others would be associated with parallel changes in the self-concepts of ability held by the students themselves. This was strongly confirmed for all time intervals studied. Secondly, it was hypothesized that changes in self-concept of ability would be associated with parallel changes in grade point index. This hypothesis was confirmed for a three-year time interval, but not for one- and two-year intervals. No data were reported on parallel changes in perceived images held by others and changes in achievement. Such analyses would provide even more conclusive support of the interactionist hypothesis than has so far been presented. Such longitudinal studies as those cited also lend strong although not conclusive support for the cause and effect assumptions of the theory.

While the fourth approach to construct validity is certainly the most widely used, Wylie cautions that such studies are no substitute for approaches 1-3 "because the ratio of unknown to known variables does not preclude alternative interpretations."[24] Such hypothesis testing as is described in Projects #845 and #1636 or the present research must make prior assumptions as to the validity of the instrument; the validity of the instrument is in turn supported by such hypothesis testing but can never be separated from this process. Because of the danger of compounding errors, a multiple attack on the problem of instrument validation is required.

---

[24]Wylie, op. cit., p. 26.

CHAPTER III

RESEARCH METHODOLOGY

Development of the SCA Scale

Because the literature revealed no existing instrument designed to tap self-concept as a learner, the research staff of Project #845 was forced to construst their own scale. The staff and panel of consultants worked at the conceptualization of the self dimension to be tapped, and subsequently submitted all the items they could think of consistent with the theoretical deliberations. The resulting list of items was administered in the Spring of 1959 to a classroom of seventh grade students in a neighboring town for whom grades were available. The items were examined to determine whether they did indeed differentiate those with differential achievement. Modification in content and format were made as a result of this preliminary testing, and a list of sixteen items were selected for a formal pretest (see Appendix A).

The pretest of the Self-Concept of Ability (SCA) Scale came as part of an interview with a sample of 49 students who showed differential achievement patterns relative to ability. The base population from which these 49 students were selected consisted of a one-third

random sample (N = 425) of the 1959-60 seventh grade students in Oldtown. School records provided IQ, achievement (GPA) and social class data. Students falling within one standard error of measurement either side of the mean on IQ (two nonverbal scores averaged) and GPA (4th, 5th, and 6th grades averaged) were eliminated from the population resulting in the four cells indicated in Figure 1. The 49 students interviewed came from the four achievement groups

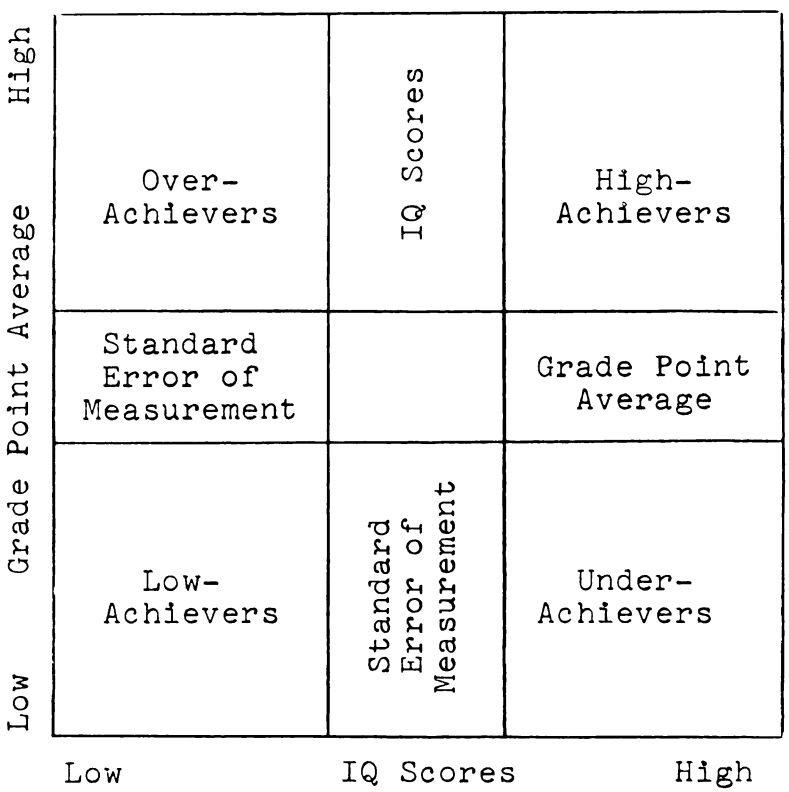

Fig. 1.--Classification of students on the basis of measured intelligence and grade point averages.

indicated and each group represented a range in social class. There were 16 "over-achievers" (nine females and seven males); 15 "under-achievers" (six females and nine males); five "high-achievers" of each sex; and four "low-achievers" of each sex.

The interviews of these students were conducted at Michigan State University, lasted about 1½ hours each, and covered a variety of subjects relevant to Project #845. At the beginning of the interview the subjects were asked to complete the written pretest of the SCA Scale consisting of 16 items.

Analysis of the pretest items consisted of item analysis and Guttman scaling. Items with less than .50 point biserial correlation with the total score were eliminated. The resulting items were subjected to Guttman scalogram analysis following the procedures outlined in Goode and Hatt.[1] Some items in the scale were found to operate in virtually the same manner as other items when scaled. Such duplications were eliminated using item content as the primary criterion for the elimination. The remaining eight items formed a Guttman Scale with a .91 coefficient of reproducibility In these analyses the responses of the males and females were combined.

It should be noted that for the pretest, the average of two nonverbal IQ scores was used in the classification

---

[1]William J. Goode and Paul K. Hatt, Methods in Social Research (New York: McGraw Hill, 1952), Chapter 17.

of students. This was done on the assumption that the nonverbal scores would be more culture-fair. However, correlation analysis of the 4th and 6th nonverbal and the 4th and 6th total IQ scores indicated that the test-retest reliability was significantly higher for the total scores than for the nonverbal. For this reason the total scores were used in the final study for selection purposes.

The results of the pretest had reassured the research staff of the feasibility of tapping self-concept as a learner with a paper and pencil test. Subsequently a full-scale administration of the Scale was done in the Fall of 1960 to all of the 7th graders in the Oldtown public school system. The 7th grade level was selected as the students were presumed mature enough to understand the questions, and yet naive enough to answer them honestly. Junior high students also have different teachers for different subjects and thus would be expected to make the specific subject distinctions which were important for getting at specific subject self-concepts.

Of the 1930 7th graders in Oldtown, data from 1050 were selected for testing the major hypotheses. The 1050 represented all students who were not Negro (the approximately 100 Negroes were separately studied at a later time), who had been in the school system at least two years, and for whom there was complete background data, particularly two IQ tests and some social class data.

Unless otherwise indicated, all data reported in this dissertation come from these 1050 students, 513 males, and 537 females.

Analysis of the SCA items for the Fall 1960 testing again consisted of Guttman scalogram analysis which yielded reproducibility coefficients of .95 and .96 for the 513 males and 537 females respectively. Figure 2 gives the eight items in the order in which they compose the Guttman scale. A line between response categories shows how the items were dichotomized for scaling purposes. The order of items shown here is not as it appeared on the questionnaire administered to the students (see Appendix B).

Guttman scaling provides scores called scale types which correspond to the number of items "right" (in this case, item responses in the top half of the dichotomy). The scale type score range is equal to the number of items --in this case eight. Because of a desire for a larger score range, Guttman scale type scores were compared with scores obtained by conventional summing procedures. Values from 5 to 1 were assigned to response categories "a" through "e" and the resulting values were summed. The potential score range in this procedure is 40 points. For comparison, the two sets of scores were each correlated with GPA, the dependent variable in the study. As the correlations were almost identical, use of the more convenient summative score appeared warranted. Guttman scale scores were not used in any subsequent analyses.

3. Where do you think you would rank in your class in high school?

a* among the best
b above average
c average
d below average
e among the poorest

6. In order to become a doctor, lawyer, or university professor, work beyond four years of college is necessary. How likely do you think it is that you could complete such advanced work?

a very likely
b somewhat likely
c not sure either way
d unlikely
e most unlikely

4. Do you think you have the ability to complete college?

a yes, definitely
b yes, probably
c not sure either way
d probably not
e no

2. How do you rate yourself in school ability compared with those in your class at school?

a I am among the best
b I am above average
c I am average
d I am below average
e I am among the poorest

7. Forget for a moment how others grade your work. In your own opinion how good do you think your work is?

a my work is excellent
b my work is good
c my work is average
d my work is below average
e my work is much below average

8. What kind of grades do you think you are capable of getting?

a mostly A's
b mostly B's
c mostly C's
d mostly D's
e mostly E's

5. Where do you think you would rank in your class in college?

a among the best
b above average
c average
d below average
e among the poorest

1. How do you rate yourself in school ability compared with your close friends?

a I am the best
b I am above average
c I am average
d I am below average
e I am the poorest

*The line below this response and in each of the subsequent items indicates the division of responses used in the scalogram analysis into what, in conventional tests, would be considered "pass" and "fail" responses.

Fig. 2.--Self-Concept of Ability Scale items shown in the order in which they form a Guttman Scale.

In addition to the SCA Scale, there were four additional self-concept of ability scales used in Project #845. Each of these scales had a specific subject reference and thus are referred to as the Specific Subject Self-Concept of Ability Scales (Specific SCA Scales). These scales were not theoretically derived, but were intentionally parallel versions of the SCA Scale. For example, item 1 on the SCA Scale reads:

> How do you rate yourself in school ability compared with your close friends?

The parallel question for the Specific SCA Scales reads:

> How do you rate your ability in (arithmetic, English, social studies or science) compared with your close friends?

For the specific format on these specific subject scales, see Appendix C.

The Specific SCA Scales were checked to see if the items would scale in a fashion parallel to the general SCA Scale. The items did scale in parallel fashion except at the extremes where the two sets of extreme items could not be differentiated. Only one specific subject scale (in arithmetic) was actually analyzed for a reproducibility coefficient and it was above the required level of .90. As the specific subject scales correlated with specific subject achievement in the anticipated fashion, no further analysis of these scales was undertaken. The purpose of devising the specific subject scales was to allow testing Hypotheses 3 and 4, and not to analyze the scale characteristics. Although there is reason to believe that the

specific subject scales are directly parallel to the SCA Scale, it would be unwarranted to assume that the establishment of validity and reliability for the SCA Scale will generalize to the specific subject scales.

Present Analyses of the SCA Scale

Additional analyses of the SCA Scale are the subject of the present research. As previously indicated, the discussion of reliability and validity will follow the classification scheme of the American Psychological Association. The techniques to be employed in establishing validity and reliability are given below in the same order as in Chapter I. For the hypothesis testing discussed under construct validity, the .05 level of significance is used in rejecting the null hypothesis.

Section 1: Validity

Construct Validity.--As noted previously, construct validation requires hypothesis testing of theoretically derived hypotheses which are related to the construct under discussion. The rationale for each of the present hypotheses is discussed in Chapter I and will not be repeated here. However, for convenience, the hypotheses will be restated followed by the methodology used in testing them.

Hypothesis 1: The self-images that junior high students hold of their ability are significantly and positively correlated with school achievement. 1-A: This relationship will be true even if IQ (as measured by standard tests) is held constant. 1-B: This

relationship will hold true even for a subgroup of "over-achievers" and "under-achievers" where there is a negative correlation between IQ and grade achievement.

The test of the general hypothesis will involve examination of the correlations between SCA Scale scores and grade point averages for the samples of 513 males and 537 females. The correlations will be computed separately for each sex and will be examined for direction and significance from zero.

Hypothesis 1-A will be tested by partialling out the effect of IQ test scores from the correlations between SCA scores and GPA, and determining if the partial correlations are still positive and significant from zero.

The test of Hypothesis 1-B will involve repeating the above processes for a subgroup of students consisting of both "over-achievers" and "under-achievers." These students were identified in a parallel fashion to those in the pretest (see Figure 1) and are part of the sample of 1050 used in the test of the hypotheses. The subgroup consisted of 110 students: 22 male and 23 female "under-achievers," and 34 male and 31 female "over-achievers" respectively. In the test of the hypothesis males and females were combined as were the "over-achievers" and "under-achievers."

Hypothesis 2: A high self-concept of ability is a necessary but not sufficient cause for high grade achievement.

The test of Hypothesis 2 used for its sample that subgroup of individuals who fell in cells A and D in Figure 3. One standard error of measurement either side of the mean on

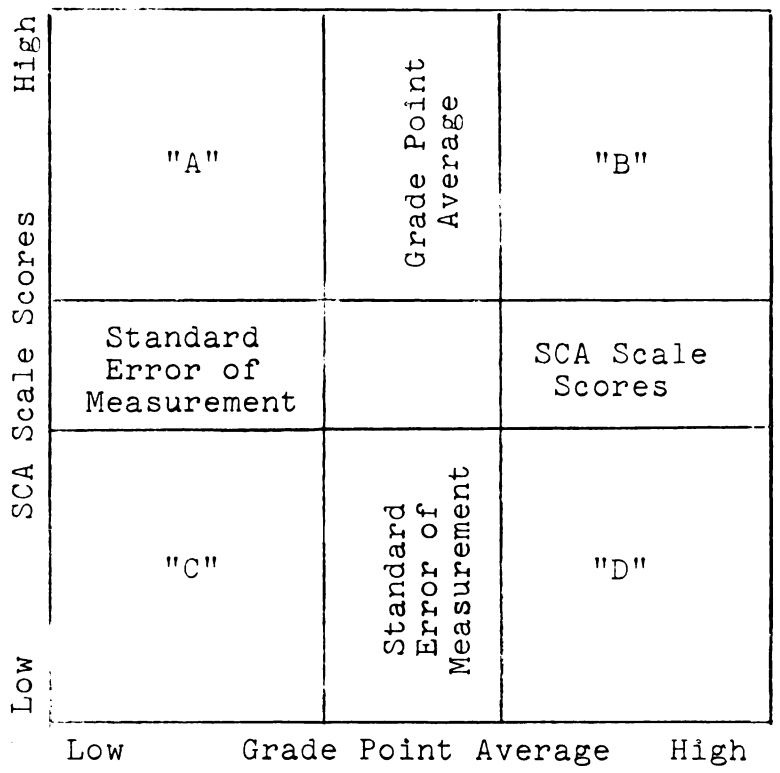

Fig. 3.--Classification of students on the basis of grade point average and SCA scale scores for test of hypothesis 2.

each variable provided the cut-off points for separating the groups. Individuals in cells A and D represent those who are deviant from the predicted self-concept of ability and achievement relationship. Those in cells B and C would be considered "normal" in this relationship and are not involved in the test of the hypothesis.

Stated in other terms the hypothesis says that there should be significantly fewer individuals who are high in GPA but low in SCA (cell D) than there are individuals who are high in SCA and low in GPA (cell A). In theory, as was previously indicated, there should be zero cases in

cell D. Practical operationalizing of the "necessary" part of the hypothesis, however, forces a statement in terms of the probability of occurrence of cases.

The statistical test of Hypothesis 2 will use the two-cell Chi Square test comparing the observed frequencies of cases in cells A and D against the assumption of random distribution of cases that are deviant in the GPA-SCA relationship.

> Hypothesis 3: The self-images that junior high students hold of their ability in specific subject areas are significantly and positively correlated with school achievement in the parallel subject area. 3-A: These relationships will hold true even if IQ (as measured by standard tests) is held constant.

Since Hypotheses 3 and 3-A are designed to indicate that the Specific SCA Scales behave in a fashion parallel to the more general SCA Scale, the test of the hypotheses will follow the same procedures that are used in testing Hypotheses 1 and 1-A. The correlations will be tested for positive direction and significance from zero with, and without the effect of measured IQ partialled out.

> Hypothesis 4: The Specific SCA Scales will be significantly better predictors of achievement in the parallel subject than will the more general SCA Scale. 4-A: The general SCA Scale will be a significantly better predictor of achievement in a specific subject than will any Specific SCA Scale other than the one in the corresponding subject. 4-B: The general SCA Scale will be a significantly better predictor of general achievement (GPA) than will any Specific SCA Scale.

The test of Hypothesis 4 will require determining if the correlations between specific self-concept scores and specific subject grades are significantly greater than the correlations between the general self-concept score and the specific subject grades. The one-tail "t" test for correlated data is the appropriate statistical test for determining significance of difference between correlation coefficients.

Hypothesis 4-A will involve a similar test: the correlations between general self-concept scores and achievement in specific subjects will be examined to see if they are significantly greater than the correlations of specific self-concept scores with grades received in nonparallel subjects. The one-tail "t" test for correlated data will be used.

In corresponding fashion test of Hypothesis 4-B will involve comparison of the correlation of general self-concept and general achievement with the correlations of specific self-concepts and general achievement. The former correlation is hypothesized to be significantly greater than the latter correlations.

Content Validity.--The items for the SCA Scale were selected from a pool of items derived from self-concept theory. Since the pool consisted of all items conceived of by the research staff and panel of consultants, it is assumed that all aspects of the self-concept of (academic) ability dimension have been sampled. This is the essential

methodology of content validity. However, other supportive evidence can be examined. Under the assumption that self-concept of (academic) ability is a single dimension, finding that the items form a Guttman Scale is support that the items are in fact all from the relevant dimension. Further test of the assumption of unidimensionality will be made through an inspection of the content of items from a logical basis. The existence of possible subsets of items was suggested in Chapter I. The examination of the correlations of these apparently different subsets of items with the total score and an outside criterion will allow empirical test of the validity of the logical analysis. The methodology for doing this will be discussed under internal consistency reliability below.

Predictive Validity.--Predictive validity requires the prediction from a test of an outcome which is then compared with the actual outcome measured at a later time. The correlation of SCA Scale scores administered in October 1960 (before the first report cards) with school grades earned in January 1961 is a simple predictive validity indicator.

To allow comparison of the relative ability of SCA Scale scores and IQ scores to predict GPA, a prediction equation may be derived and the relative weights of the two variables determined. The prediction equation takes the form of:

$$X'_1 = B_2X_2 + B_3X_3 + A$$

where X'_1 is the predicted GPA of an individual, and X_2 and X_3 are respectively IQ and SCA Scale raw scores. "A" is a constant. Rewriting the above equation in deviation units gives:

$$x'_1 = b_2x_2 + b_3x_3 + a$$

Since in this equation the value of "a" becomes zero, comparison of the magnitude of b_2 and b_3 will show directly the relative importance of the two variables in predicting x'_1.

To cross-validate the prediction equation derived above, the equation must be used to predict GPA for a group of students not involved in the equation derivation. The cross-validation sample consisted of 50 males and 50 females who were randomly selected. These students were in the same grade and school system as the students involved in the main study, but had originally been eliminated due to incomplete data or race. The cross-validation sample is thus possibly more heterogeneous.

The cross-validation index is the correlation of predicted GPA with actual GPA. In the present case, grades obtained in June of 1961 were predicted from IQ and SCA scores obtained in the Fall of 1960. The cross-validation prediction is then for an additional six months of elapsed time which may result in a lowered correlation although the opposite effect would be expected if the cross-validation

population is actually more heterogeneous (i.e. the range is less restricted).

Section 2: Reliability

Stability Reliability.--The stability reliability index to be used in the present case is the simple correlation of SCA Scale scores obtained in the Fall of 1960 with the scores on the same scale administered in the Fall of 1961. Under the assumption of stability of the self-concept of ability dimension, the correlation will presumably reflect the amount of nonerror variance accounted for by the test over time.

Internal Consistency Reliability.--Internal consistency reliability will be examined from three points of view all of which ask whether the items in the test are consistently related to each other and the total score.

The first approach, of Guttman scaling, asks whether one is warranted in assuming that all of the items are from a single or unidimension, and further whether the items are so related to each other that the order of response is predictable. From knowing only the total score (scale score) one should be able to reproduce the responses to each individual item. To the extent that all individual item responses are "reproducibile" the conclusion of a scalable unidimensional universe is warranted. The Guttman index is a Reproducibility Coefficient (Rep) which is computed from the ratio of error to correct predictions and should not

exceed .10. More conventionally the error ratio is subtracted from 1.00 to give a minimum acceptable level of reproducibility of .90.

The second approach to internal consistency reliability is the Hoyt method of analysis of variance. The ratio of error variance to individual variance is computed. As a rule of thumb, if the error variance is less than 10 per cent of the variance that comes from individuals, the test is considered very good--i.e. reliability equals .90 or better. The Hoyt method may be considered a variation on the more common Kuder-Richardson estimates of reliability; the Hoyt method has the convenience of using raw data.

Examination of the SCA Scale for possible subscales is the third approach to internal consistency reliability. Under "Content Validity" it was noted that the items appear on logical analysis to ask slightly different questions. Whether these are empirically relevant distinctions is of present concern.

Logical analysis first suggested that item 7 (. . . in your own opinion how good do <u>you</u> think your work is?) is really asking for an evaluation of actual performance rather than an evaluation of potential ability. The retention of the item in the scale initially was done on empirical grounds--it had a high correlation with the total score and it scaled. Logic however, suggests its uniqueness in the scale and so it will be considered alone and not used in the item analysis described below.

The remaining items all ask for evaluations of ability, but again, logic suggests slightly different standards for the evaluation. One set of questions has an immediate time reference; the other asks the student to project to the future. Or, taking another approach, the student is asked on the one hand to compare his ability with others, while on the other hand he is asked for an absolute evaluation of his ability. Other bases for analysis could probably be made, but the present research is concerned with only these two. The result is four subsets of items which are identified below with the items composing them. The specific content of the items can be found in Appendix B.

Subset 1: Current Time Reference Items (items 1, 2, 8)

Subset 2: Future Time Reference Items (items 3, 4, 5, 6)

Subset 3: Comparative Evaluation Items (items 1, 2, 3, 5)

Subset 4: Absolute Evaluation Items (items 4, 6, 8)

Subset scores will be compared with each other and with the outside criterion of grades in a correlation matrix to determine if the logical distinctions are psychologically relevant.

Because of the possibility that logical analysis has failed to make the psychologically important distinctions, an empirical analysis of the items will also be done. The appropriate technique would normally be factor analysis,

but the limited number of items makes this difficult. A simplified method with somewhat the same intent is cluster analysis. Cluster analysis differs from factor analysis in that each item is assigned to one cluster or another while factor analysis assigns a portion of the variance from a given item into several factors and indicates where the greatest variance is accounted for. The McQuitty method of cluster analysis will be used.[2]

---

[2]Louis L. McQuitty, "Elementary Linkage Analysis for Isolating Orthogonal and Oblique Types and Typal Relevancies," Educational and Psychological Measurement, 17, No. 2 (Summer 1957), pp. 207-229.

CHAPTER IV

RESULTS AND INTERPRETATION

Section 1: Validity

Construct Validity

Evidence for the construct validity of the Self-Concept of Ability (SCA) Scale is obtained from the test of the theoretically derived hypotheses discussed in Chapters I and III. Results of the tests are given below. For convenience the hypotheses are restated.

> Hypothesis 1: The self-images that junior high students hold of their ability are significantly and positively correlated with school achievement.

The correlation of grade point average (GPA) with self-concept of ability scores was found to be .57 for each sex. (The .95 confidence interval for this correlation is .51-.63. With this size sample any correlation greater than .08 is significantly different from zero.) The correlation is positive and thus in the predicted direction. The hypothesis is supported.

> Hypothesis 1-A: The relationship will be true even if IQ (as measured by standard tests) is held constant.

The correlation between grade point average and self-concept of ability scores remains positive and significantly

different from zero when the effect of IQ is partialled out. This can be seen in Table 1 where the three variables are each held constant to give a more complete picture. The multiple correlation of IQ, SCA and GPA is also given.

TABLE 1.--Correlations between SCA, IQ and GPA with and without the effect of the third variable partialled out for 513 males and 537 females.

Variables Correlated	Correlations		Variable Controlled	Partial Correlations	
	Males	Females		Males	Females
SCA-GPA	.57	.57	IQ	.42	.39
IQ-GPA	.61	.65	SCA	.48	.53
IQ-SCA	.46	.48	GPA	.17	.17

Multiple Correlation of all three Variables: R = .69 (males) and R = .72 (females)

When the effect of the third variable is partialled out, all of the partial correlations drop significantly, but remain positive and significantly different from zero. Most conspicuous however, is the considerably lowered correlation between IQ and SCA when the effect of achievement is partialled out. This is evidence, consistent with theoretical expectations, that IQ and SCA are measuring different things both of which are closely tied to achievement behavior.

One might ask about the "true" correlation between SCA and GPA (or IQ and GPA) under the assumption of

perfectly reliable measures. Correcting for attenuation in both variables increased the GPA-SCA correlation to .69 (males) and .67 (females), values accounting for 48 and 45 per cent of the variance respectively. (The reliabilities used in these calculations were the Hoyt reliability on GPA of .91 and .93 for males and females, and the one-year stability reliability values of .75 and .77 (males and females) for SCA.) Comparable correction for the IQ-GPA correlation (assuming stability reliabilities of .90 for a one-year period on IQ scores) gives corrected correlations of .67 and .71 accounting for 45 and 50 per cent of the variance for males and females respectively. The results of applying such corrections suggest that in Table 1 the greater correlation of IQ with GPA than of SCA with GPA may be attributed to differences in the reliability of the SCA and IQ measures. When corrected for these reliability differences, IQ and SCA each correlate with GPA to about the same degree. It should be recalled that such correlations do not mean causation, and that Table 1 figures are based on the entire sample. When, for example, only the subsample of over-achievers and under-achievers is considered (see Hypothesis 1-B), the correlation of SCA with GPA is much greater than that of IQ with GPA.

Hypothesis 1-B: This relationship will hold true even for a subgroup of "over-achievers" and "under-achievers" where there is a negative correlation between IQ and grade achievement.

The interrelation of the three variables, SCA, IQ, and GPA is shown in Table 2. This is directly parallel to Table 1 above but for a group of students for whom there is a known negative correlation between grades and intelligence scores.

TABLE 2.--Correlation between SCA, IQ and GPA with and without the effect of the third variable partialled out for 110 "over"- and "under-achieving" students (males and female combines).

Variables Correlated	Correlations	Variable Controlled	Partial Correlations
SCA-GPA	.40	IQ	.36
IQ-GPA	-.23	SCA	-.15
IQ-SCA	-.25	GPA	-.18

Multiple Correlation of all three variables: R = .42.

Table 2 shows clearly the positive relationship between self-concept of ability and grades. Partialling out IQ has virtually no effect on the correlation, and addition of IQ as a third variable in the multiple correlation accounts for little more variance than is accounted for by the SCA-GPA correlation alone. It is also noteworthy that even for this special group of students the partial correlation of SCA-GPA is comparable to the partial correlation for the sample as a whole (see Table 1). (It should be noted that the total sample does include the

"over"- and "under-achievers" which may somewhat bias the correlations for the "normal" students.) The hypothesis is supported.

The results of Hypothesis 1-B are similar to those found by Richard Morse in a study of Negro students.[1] For this special group there is also a negative correlation between IQ and GPA but a positive correlation between GPA and SCA. The present research purposely eliminated Negro students; Morse's analysis is for the Negro sample who were in the same grade and school system as those used in the present sample.

Hypothesis 2: A high self-concept of ability is a necessary but not sufficient cause for high grade achievement.

The test of Hypothesis 2 involved looking at those students who had a discrepancy between their SCA score and GPA. It was hypothesized that there would be significantly fewer (in theory, no cases at all) individuals high in GPA but low in SCA (cell D) than individuals high in SCA but low in GPA (cell A). For cell designations see Figure 3.

A two-celled Chi Square test was used to test the null hypothesis that there would be an equal number of cases in cells A and D. The numbers for the males were respectively 23 and 7. This gives a Chi Square value of 8.53

---

[1]Richard J. Morse, "Self-Concept of Ability, Significant Others and School Achievement of Eighth Grade Students: A Comparative Investigation of Negro and Caucasian Students" (unpublished Masters thesis, Michigan State University, 1963).

significant beyond the .01 level of significance. For the females the A and D cell entries were 22 and 11 which gives a Chi Square value of 3.67 significant at the .06 level. While the Chi Square value for females falls short of the .05 level the results are strongly in the direction predicted and will tentatively be taken as supportive of the hypothesis in view of the strong support given by the males. Combining sexes would, of course, yield a significant difference.

Assuming for a moment that the sex differences are real, one might speculate that there is a greater tendency on the part of girls to publicly belittle their own intellectual capacities. This would not be surprising in a society where girls are taught never to "show-up" the boys. Add to this the greater tendency of girls to conform to achievement expectations (i.e. work hard), and the probability of candidates for cell D (high GPA, low SCA) is raised. Further research on the relation of cultural definitions of achievement by sex is needed to clarify this issue.

It is also possible that the noted sex differences are a function of the methodology used in the test of the hypothesis. The standard error of measurement was used to determine cut-off points for the low and high groups on SCA and GPA. There is, however, no way of knowing whether the standard error of measurement corresponds to any psychological reality. The "high" SCA group in the

present test had scores of 30 or more (males) and 31 or more (females) on an instrument with means of about 27 and 28 for the two sexes and standard deviations of about 4. Yet we know that students with "straight A" (4.00) averages have SCA scores ranging from 29-39. How much more variable are the present sample of "high achievers" who are defined by the standard error of measurement procedure as having GPA's of 2.07 or better (males) and 2.43 or better (females)! It is undoubtedly true that the present definition of "high" achievement is psychologically irrelevant. It is probably equally true that "high" self-concept of ability should be redefined.

The importance of the operational definitions can be seen in some reanalysis of the data. Keeping the original GPA cut-off points but cutting on the SCA variable at only one more point extreme gives A and D cell entries of 16 and 1 for the males and 14 and 4 for the females; both Chi Square values are significant.

Looking at the GPA variable, on the other hand, it can be observed that no male with a GPA of B or better falls in a "low" cell; and only one female with a GPA of B+ or above falls "low." The "B or better" boys, and "B+ or better" girls represent 20 per cent of the male and 17 per cent of the female populations. One might hypothesize that psychologically one considers himself a "good" student relative to those in his own sex group, and that it is not an absolute level of achievement but a

relative level which is important here. Other data from Project #845 would suggest that at the junior high age level peers of the same sex are more likely to be "significant others" than are members of the opposite sex. This might change, of course, with age, but it is also possible that one's sex group is always the relevant reference group in matters of achievement. If the Chi Square test is rerun using the top and bottom 20 per cent in GPA for each sex (and the original SCA cut-off points), the A and D cell entries are 8 and 1 for females and 9 and 0 for males. The sex differences observed previously largely disappear and the Chi Square values are significant using the Yates correction.

A third reanalysis of the data simultaneously changing the cut-off points on GPA and SCA reveals an interesting progression. If we use the top and bottom 30 per cent on each variable to designate "high" and "low" groups, the A and D cell entries are 16 and 3 (males) and 16 and 5 (females). Using more extreme 20 per cent cut-off points on each variable gives analagous cell entries of 7 and 0 (males) and 4 and 1 (females). At an even greater extreme of using approximately 10 per cent cut-off points on the two variables, the A and D cell entries drop to 1 and 0 (males) and 0 and 0 (females). The regular progression is consistent with the theory and supports the "necessary" condition of the hypothesis to the effect that there should in theory be no cases in cell D--a condition fulfilled empirically

only for very extreme groups. There is no theoretical probability which can be derived for cell A entries other than that they should be relatively uncommon and yet more likely than cell D entries. (Even in the original test of the hypothesis the total number of A and D entries combined was only 62 or 6 per cent of the total sample.) Clearly the test of the hypothesis requires using groups which are not too extreme or one faces having no cases at all.

It is perhaps clearer that the data support the theory than that they support the hypothesis. Probably most significant is the fact that no data suggest that the hypothesis or theory is incorrect. Taken in this light there appears to be ample evidence that a high self-concept of ability is necessary for high achievement, but not sufficient to guarantee it. To the degree that the operational test of the hypothesis is an adequate test of the theory the data are generally supportive.

Hypothesis 3: The self-concepts that junior high students hold of their ability in specific subjects are significantly and positively correlated with achievement in the parallel subject area.

At the outset it should be noted that there are subject area differences in mean Specific Subject Self-Concept of Ability (Specific SCA) scores for the group as a whole. In general these differences correspond to the differences in mean grade achievement for the various subjects. The means and standard deviations of the Specific SCA scores are given

below with the corresponding mean achievement for the four subjects. These subject area differences, of course, were not predictable and are probably a function of the particular school system and subject matter interest on the part of the students.

Table 3 gives further support to the notion that relative level of achievement rather than absolute level is correlated with self-concept of ability. Girls get about the same average in social studies that boys do in science, yet girls have their lowest Specific SCA score in social studies their worst subject, while boys have their highest

TABLE 3.--Means and standard deviations of SCA and Specific SCA Scale scores compared with means for general achievement and specific subject achievement for 513 males and 537 females.

	Males			Females		
	SCA	S. D.	GPA	SCA	S. D.	GPA
All Subjects	27.35	4.38	2.07	28.25	3.95	2.43
Arithmetic	26.97	6.42	2.10	*27.47	5.68	2.36*
English	*25.45	5.78	2.00*	28.17	4.96	2.61*
Social Studies	*25.63	6.54	1.99*	*26.58	5.78	2.29*
Science	27.18	7.15	2.21*	27.73	5.72	2.49*

*Significantly different from the mean for all subjects using a two-tailed "t" test for correlated data.

Specific SCA score in science, the area in which they have the highest achievement. Note also that the general SCA

mean scores are higher than any of the Specific SCA scores, but closest to the subject in which the student has his highest achievement. This suggests that students generalize from their points of strength, a finding consistent with psychological theories which postulate that people are motivated to seek the most positive self-definitions.

For the test of Hypothesis 3 the specific subject scale scores and achivement are shown in Table 4. All correlations are positive and significantly different from zero. With the exception of English, the specific subject SCA and GPA correlations are comparable in magnitude to the general SCA-GPA correlation. The hypothesis is supported.

TABLE 4.--Correlation between SCA scores and GPA compared with the correlations of Specific SCA scores and achievement in parallel subjects with and without the effect of IQ controlled for 513 males and 537 females.

Variables Correlated	Correlation Without IQ Controlled		Correlation With IQ Controlled	
	Males	Females	Males	Females
SCA and GPA	.57	.57	.42	.39
Arithmetic, SCA & Arithmetic GPA	.59	.54	.47	.39
English SCA & English GPA	.43	.47	.31	.34
Social Studies SCA & Social Studies GPA	.56	.58	.46	.44
Science SCA & Science GPA	.61	.51	.49	.38

Hypothesis 3-A: These relationships will hold true even if IQ as measured by standard tests is held constant.

The effect of partialling out IQ can be seen in Table 4. The Specific SCA and grade achievement correlations remain significant and positive. The effect of partialling out IQ has about the same effect with the specific subject-specific grade correlations as was found for the general SCA-GPA correlation. The hypothesis is supported.

Another way of looking at the same data is to note the increase in variance accounted for when IQ is combined with specific subject self-concept scores to predict achievement. The data are given in Table 5 along with comparative data for general SCA plus IQ to predict GPA. Some subject area differences are noted, especially for females in English, but the increased magnitude of the multiple correlations is as expected and again supports the assertion of comparability between the general SCA and the Specific SCA scales.

TABLE 5.--Correlation between measured intelligence and grades in four school subjects (and GPA) compared with the correlations when Specific SCA scores (and SCA) are added as a third variable, 513 males and 537 females.

Subject Matter Area	Correlation of IQ and GPA Only		Multiple Correlation of IQ, GPA and SCA	
	Males	Females	Males	Females
All Subjects	.61	.65	.69	.72
Arithmetic	.52	.57	.65	.66
English	.46	.57	.53	.63
Social Studies	.56	.58	.68	.68
Science	.56	.59	.69	.66

Hypothesis 4: The specific subject self-concept of ability scores will be significantly better predictors of achievement in parallel subjects than will the more general self-concept of ability score.

The relative ability of general vs specific self-concept of ability scores to predict specific subject achievement can be seen in Table 6. The hypothesis is only partially confirmed, and in the case of English the results are contrary to expectation. Only for males does the hypothesis generally hold up; for the females it is confirmed only in the case of social studies.

The deviant pattern for English has occurred in previous analyses (for example, in Table 5 one may note the much higher correlation of IQ with English grades for females.) The impression is of an unclear self-concept of ability in English especially for males. The explanation probably lies in the noncomparable evaluation systems used in English as contrasted with other subjects, and the fact that 7th graders have never had "English" before, but rather spelling, vocabulary, etc. To the extent that English is graded subjectively, then achievement in the subject will vary more from teacher to teacher than is true of other subjects. This will result in a nonconsensual evaluation of the student's ability in this subject and will be reflected in the student's self-concept score. General self-concept scores would then be expected to predict English achievement as well if not better than the specific self-concept score. More research

TABLE 6.--Correlations comparing the relative ability of general SCA scores and Specific SCA scores to predict specific subject achievement for 513 males and 537 females.

Specific Subjects	General SCA Scores and Specific Grades		Specific SCA Scores and Specific Grades		Probability of Difference	
	Males	Females	Males	Females	Males	Females
Arithmetic	.50	.52	.59	.54	.01	n.s.
English	.44	.52	.43	.47	n.s.	n.s.
Social Studies	.51	.50	.56	.58	.05	.01
Science	.52	.48	.61	.51	.01	n.s.

Note: The one-tail "t" test for correlated data was used to calculate the probability of significant differences.

is needed to show how the student handles such variable evaluations of his ability.

One possible interpretation of the sex differences observed in Table 6 is that they arise from faulty methodology. It was previously noted that test of the hypothesis is impossible unless there is actually differential achievement in the various subjects. It was noted that "stereotypic" students have more uniform achievement patterns than others. It may also be true that females have more uniform achievement patterns than males thus resulting in biased results. If a student gets the same grades in all subjects, our theory would say that the student should have similar images of his ability in all subjects and would likely have a general image of his ability not much different from the specific images. Hypothesis 4 could not be true under such conditions. For greater accuracy the hypothesis should be revised to read:

> Where achievement patterns are nonuniform, the specific self-concept of ability scores will be significantly better predictors of achievement in the parallel subject than will the more general self-concept of ability score.

The revised hypothesis was tested by selecting all students who had a discrepancy of two points or more (only a handful of students had three-point discrepancies) between highest and lowest grades and calling these "nonuniform" achievers. For example, students with grade patterns such as A-B-B-C, C-C-F-F, B-D-D-D and the like would qualify. Another group of students who had completely

uniform grades (straight A's, B's, C's, etc.) were designated as "uniform" achievers. The correlations of specific self-concept scores and specific grades were compared as before with the correlations of general self-concept of ability and specific grades. The results are given in Table 7 separately by sex for "uniform" and "nonuniform" achievers.

At the outset, the table shows that females are significantly more likely to be uniform achievers than nonuniform achievers ($P < .01$ for a two-cell Chi Square test) while males are about equally likely to be uniform or nonuniform achievers. The revised hypothesis is generally supported for among nonuniform achievers the Specific SCA scores are significantly better predictors of specific grade achievement in all subjects for females and in all subjects but social studies for males. Among the uniform achievers, however, there is no case where the Specific SCA scores predict better than the general SCA score, and the general SCA score is a significantly better predictor in all subjects except social studies for females and English for males. These sex and subject matter differences are a mirror image of those found in Table 6 and help clarify those findings. Since the general SCA scores are better predictors of achievement for "uniform" female achievers, and since females are more likely to be "uniform" achievers, it is not surprising that Hypothesis 4 as originally stated was rejected for females (except in social studies).

TABLE 7.--Correlations comparing the prediction of specific subject achievement by Specific SCA scores in the parallel subjects and by the general SCA score for "uniform" and "nonuniform" achievement groups of each sex.

Subject	Specific SCA Scores and Specific Grades		General SCA Scores and Specific Grades		Probability of Difference in r's*	
	Males	Females	Males	Females	Males	Females
"Uniform" Achievers (96 Males and 125 Females)						
Arithmetic	.65	.55	.65	.65	n.s.	.05
English	.46	.52	.65	.65	.05	.05
Social Studies	.55	.61	.65	.65	n.s.	n.s.
Science	.65	.52	.65	.65	n.s.	.05
"Nonuniform" Achievers (112 Males and 80 Females)						
Arithmetic	.56	.45	.32	.17	.01	.01
English	.36	.61	.20	.28	.05	.01
Social Studies	.53	.66	.42	.29	n.s.	.01
Science	.67	.66	.49	.42	.01	.01

*For "uniform" achievers, the two-tail test for correlated data was used; for the "nonuniform" achievers, a one-tail "t" test for correlated data was used as the hypothesis was directional.

Despite the general acceptance of the revised Hypothesis 4, there are still subject area and sex differences to be accounted for. Among females uniform achievers, why is the general SCA score a significantly better predictor of specific subject achievement when this is not true for male uniform achievers? Any why among the female nonuniform achievers are the specific subject scores so much better able than the general score to predict specific subject achievement than is the case with nonuniform male achievers?

One consistent explanation lies in terms of cultural definitons of sex-appropriate behavior. In American society the ideal female is noncareer oriented, generally educated and conforming. Males on the other hand should ideally be individualistic and career oriented. The young male is early made aware of the need for vocational specialization and is encouraged to develop tentative vocational goals--if only to answer the perennial question of "What do you want to be when you grow up?" As school success in subject areas related to particular vocations is one of the few criteria available for validating occupational choice at an early age, young males are no doubt required (by adults) to make tentative plans consistent with areas of academic success. Or, boys may have tentative career goals which motivate them to do well in subjects that are defined (by adults) as vocationally relevant. In either case, the young man is pushed to conceive of himself as a subject matter specialist in a way not required of girls for whom a

general education is the norm. One would speculate that a theory of vocational choice could be bullt up by looking at the degree of specificity of subject area self-images of ability through time. Crystalization of occupational choices should be observable both in terms of the relative magnitude of self-image scores in career-related vs noncareer-related subjects, and the degree to which specific self-images become distinct from the general self-image of ability. Such a theory would also operate for career-oriented females, of course.

This theory is consistent with the data at hand. It explains the greater predictability of specific subject achievement by Specific SCA scores for males as contrasted with females. The exception of English for males may be due either to the problem of noncomparable evaluation techniques used by English teachers, or by failure of the society to specify desirable careers for males which are clearly related to English achievement. That the general SCA predicts specific subject achievement so well for girls is consistent with the general education orientation advocated by society. Social studies violates this general pattern, but it can be seen (Table 3) that social studies is the subject in which females have the lowest achievement. If females are, for one reason or other, motivated toward uniform achievement, then any subject in which they are conspicuously low will be the source of frustration and likely to be focused upon as a distinct area. If this is true, the present findings

on social studies for females are likely to be a function of the particular school system and curriculum for the sample under study. Other samples might show similar patterns for other subjects where achievement is low, or perhaps no subject matter differentiation if achievement were more uniform.

It might be expected that girls would have distinctive images of their ability in English given the traditional societal association of English with females. There is no evidence in this direction again lending weight to the interpretation that females do not conceive of themselves as subject area specialists. English is, however, the subject area in which girls have the highest grades and Specific SCA scores. The average English SCA score is only .08 below the mean General SCA score of 28.25 suggesting that perhaps girls' general self-concepts are heavily influenced by success in English. For males science performs in a comparable way.

The above interpretation is also consistent with the sex differences found among the "nonuniform" achievers. Not only are females less likely numerically to be "nonuniform" achievers than males, but those females who are "nonuniform" in achievement have considerably lower general achievement and SCA scores than females as a whole. This can be seen in Table 8. Male "nonuniform" achievers appear to be a random selection from the total sample; female "nonuniform" achievers are a distinctive group.

TABLE 8.--Mean SCA and GPA for the total sample and for "uniform" and "nonuniform" achievement groups by sex.

	Mean SCA	Mean GPA
Males		
Total Sample (N=513)	27.35	2.07
"Uniform" Achievers (N=96)	27.15	2.08
"Nonuniform" Achievers (N=112)	27.34	2.05
Females		
Total Sample (N=537)	28.25	2.43
"Uniform" Achievers (N=125)	28.64	2.67
"Nonuniform" Achievers (N=80)	27.46*	2.28*

*Significantly lower than the mean for "uniform" achievers using a two-tail "t" test.

It was earlier suggested that females in general developed a distinctive self-concept of ability in social studies (but not in other subjects) because of their lower, and thus "nonuniform" achievement in that subject. Female "nonuniform" achievers similarly should be sensative to their atypical achievement pattern and thus be most likely to develop specific subject self-concepts different from their general SCA. This trend can be seen in Table 7. Males, on the other hand, are not expected to be uniform achievers and thus this trend is not visible to the same extent among them. Males do, of course, develop specific subject self-concepts distinct from the general SCA if they

are "nonuniform" achievers (consistent with revised Hypothesis 4), but these specific self-concepts are less distinctive than those developed by females.

In summary, the revised Hypothesis 4 is generally supported. In distinguishing "uniform" and "nonuniform" achievement groups from the total sample some interesting differences are found. Females are less likely to be "nonuniform" achievers; if they fall in this group they are below the total group in mean GPA and SCA; and "nonuniform" female achievers develop specific subject self-concepts that are more distinctive from the general SCA than is true for males. It is hypothesized that "nonuniform" achievement is perceived as negative by females in contrast with males where it is of apparently no consequence.

Hypothesis 4-A: The general self-concept of ability scale will be a significantly better predictor of achievement in a specific subject than will any specific subject self-concept score other than the one in the corresponding subject.

The results of the test of Hypothesis 4-A are given in Table 9. In no case does a specific self-concept score predict achievement in a nonparallel subject as well as the general self-concept of ability scale. In 17 of the 24 comparisons the general SCA score predicts significantly better; in all cases the results are in the direction predicted. The hypothesis is partially supported.

TABLE 9.--Comparison of the correlations between general SCA scores and specific subject grades with the correlations of specific SCA scores and grades in nonparallel subjects for 513 males and 537 females.

Self-Concept Score	Subject of Grade	Correlations Male	Correlations Female
General	Arithmetic	.50	.52
English	Arithmetic	.37*	.36*
Social Studies	Arithmetic	.39*	.40*
Science	Arithmetic	.44*	.35*
General	English	.44	.52
Arithmetic	English	.35*	.35*
Social Studies	English	.41	.47
Science	English	.37*	.37*
General	Social Studies	.51	.52
Arithmetic	Social Studies	.39*	.38*
English	Social Studies	.48	.43*
Science	Social Studies	.46	.40*
General	Science	.52	.50
Arithmetic	Science	.47	.40*
English	Science	.45*	.41*
Social Studies	Science	.50	.46

*Significantly lower than the correlation between general SCA and the same subject grade using a one-tail "t" test for correlated data.

Hypothesis 4-B: The general self-concept of ability scale will be a significantly better predictor of general achievement than will any specific subject self-concept scale.

Data for the test of Hypothesis 4-B are given in Table 10. In all cases the SCA general scale predicts GPA better than any specific subject scale. The difference is significant in five of the eight comparisons. The hypothesis is partially supported.

TABLE 10.--Comparison of the correlation between general SCA scores and general GPA with the correlations between specific subject SCA scores and general GPA for 513 males and 537 females.

Variables Correlated	Correlations	
	Male	Female
SCA and GPA	.57	.57
Arithmetic SCA and GPA	.52*	.48*
English SCA and GPA	.50*	.47*
Social Studies SCA and GPA	.54	.54
Science SCA and GPA	.55	.46*

*Significantly lower than the correlation of SCA with GPA using the one-tail "t" test for correlated data.

The lower correlations for females of specific self-concept scores and GPA suggest that there is a higher inter-correlation among the specific subject self-concept scales among males. This is substantiated in Table 11. The inter-correlations among specific subject grades, however, are

higher for females (see Table 12) consistent with their being more uniform in achievement patterns. These two tables appear contradictory for one might assume that the pattern in self-concept of ability scores would be directly reflected in the pattern of grade achievement.

TABLE 11.--Intercorrelations among four specific subject self-concept scores for 513 males (below diagonal) and 537 females (above diagonal).

Variables		1	2	3	4
Self-Concept Arithmetic	1	---	.56	.56	.52
Self-Concept English	2	.67	---	.68	.61
Self-Concept Social Studies	3	.64	.81	---	.59
Self-Concept Science	4	.68	.64	.68	---

TABLE 12.--Intercorrelations among grades in four subjects for 513 males (below diagonal) and 537 females (above diagonal).

Variables		1	2	3	4
Grade in Arithmetic	1	---	.67	.69	.71
Grade in English	2	.60	---	.75	.73
Grade in Social Studies	3	.62	.70	---	.73
Grade in Science	4	.65	.62	.68	---

What is found here is another case of unexplained sex differences. The exact nature of the relationship of self-concept of ability to grade achievement is not clear. Turning to Table 6, the second column of figures indicates sex differences in the self-concept and achievement relationship for the four subjects. In science, for example, the self-concept and achievement correlation for females is significantly lower than for males; the other sex differences would not be significant with a two-tail "t" test but certainly suggest variation by subject. When departing from the more general SCA-GPA relationship (where, it will be recalled, the correlation was identical for males and females--.57) separate analysis by sex should be used.

It is possible that some of the specific subject sex differences are a function of the particular school system or age level and would not hold up with other samples. However, it is also possible that females make subject matter distinctions in their self-conceptions of ability that males do not make. They may for example, be prone to indicate contrast between specific subjects based on such factors as the ease of getting the homework done in the subject--factors which do not show up in actual grades. Such distinctions would result in more highly intercorrelated grades, but less highly intercorrelated specific self-concept scores. It is likely that the test format invites subject matter differentiation and thus offers psychological pressure to make distinctions per se. This

pressure would more likely bias results for females who tend to be more uniform in achievement pattern. More research will be needed to clarify such findings.

While tests of Hypothesis 4, 4-A and 4-B do not show significance in all predicted cases, it is probably more relevant to the issue of construct validity that there is no evidence contrary to the Hypotheses. This is also true for Hypothesis 2. This would suggest that the general theoretical foundation for the study is sound but requires refinement at specific points. Clearly the theory needs to take into account existing sex differences. Future research should also investigate the self-concept and achievement relationship for different achievement levels, and possibly for different age levels, social class and religious groups and regional areas as seems warranted. All such inquiry would lead to refinement of the general theory and probably to greater construct validity of the present SCA instrument.

In conclusion, it seems warranted to say that the hypothesis testing reported supports the contention that the SCA instrument has construct validity because it generally operates in a manner consistent with the theoretical base from which it was derived. Other validity data follow in the next sections.

Content Validity

The establishment of content validity in part involved an empirical test of the relative importance of logically derived subscales in predicting the total SCA score and an outside criterion. Also, any empirical evidence that the items are unidimensional can be used to support the contention that the items were indeed derived from a single universe. The fact that the items were originally chosen by the criterion of scalability is itself evidence of content validity. The results from analysis of subscales and other data on unidimensionality are presented below under internal consistency reliability.

Predictive Validity

The prediction equations calculated which used a combination of IQ and SCA scores to predict GPA are given below where X' = predicted 7th June grades; X_2 = SCA score; and X_3 = average of two total IQ scores. For these equations all data are in raw score form.

$$X' = -2.4773 + .0710X_2 + .0244X_3 \text{ (males)}$$

$$X' = -3.0094 + .0743X_2 + .0309X_3 \text{ (females)}$$

As the above equations do not allow comparison of the relative importance of the IQ and SCA variable in predicting achievement, the equations are rewritten in deviation units

below. It will be recalled that the constant becomes "zero" in the three-variable problem.

$$x' = .37x_2 + .44x_3 \text{ (males)}$$

$$x' = .34x_2 + .49x_3 \text{ (females)}$$

The latter set of figures shows that IQ is weighted somewhat more heavily than self-concept of ability when predicting achievement. If it is recalled that IQ is largely a measure of past achievement, this is not surprising.

The raw score prediction equations were used to predict June 7th grades for a random sample of 50 males and 50 females not involved in the study but from the same school population. The predicted grades were correlated with grades actually obtained and gave correlations of .71 for males and .70 for females. These correlations should be compared with the originally determined multiple correlations of SCA and IQ to predict GPA: R = .69 (males) and .72 (females). The assumed greater heterogeneity of the cross-validation population would have the effect of raising the correlation although the additional six-month time period should have the opposite effect. In any case, the predicted correlations are unusually high (in comparison with the multiple correlations) and suggest that the multiple correlations do not capitalize on chance

factors but are accounting for variance coming from stable and real factors.

While the cross-validation was run using a combination of IQ and SCA scores, there are also predictive validity coefficients for IQ and SCA separately. The IQ-GPA correlations were .61 and .65 for males and females; corrected for attenuation in the criterion these figures rise to .64 and .67. (The r_{tt}'s for GPA were computed by the Hoyt method on 35 randomly selected males and females and found to be .91 and .93 for males and females.) The corrected correlations account for 41 and 45 per cent of the variance. By contrast the SCA-GPA correlations of .57 for each sex when similarly corrected are .60 and .59 for males and females accounting for 36 and 35 per cent of the variance respectively. Consistent with the beta weights found above, the IQ variable accounts for more variance than does the self-concept variable. The combination of the two variables accounts for 48 and 52 per cent of the variance as was seen in the multiple correlations.

Further discussion of predictive validity will be found in the discussion of internal consistency reliability where the ability of individual items to predict achievement will be discussed. There is evidence that the SCA scale as a whole does not predict achievement as well as certain subsets or combinations of items. Suggestions of possible modification of the SCA instrument to maximize predictive validity are given.

Section 2: Reliability

Stability Reliability

Stability reliability coefficients for the SCA scale represent a test-retest situation with 12 months of elapsed time. The correlations are .75 and .77 for 446 males and 508 females (some cases were lost between administrations of the test). As previously noted, there is some question as to the validity of assuming stability of the self-concept of ability construct, and thus difficulty in knowing whether to conclude the reliabilities are high or low. Because of the questionable theoretical assumptions involved, it is unwise to use stability reliability figures alone in judging the reliability of the instrument. Internal consistency figures should also be used.

Internal Consistency Reliability

Three different approaches to internal consistency were used. While the use of three approaches may appear redundant, each approach yields different data, and with a new instrument all possible sources of data should be explored. All of the internal consistency measures look at the relationship of individual items to each other and to the total score. The results of the three approaches are given below.

Guttman Scaling.--As the development of the instrument involved selecting items which were, in Guttman's sense "scalable," high internal consistency reliability

as indexed by the Guttman reproducibility coefficient is assured by definition. The high reproducibility coefficients of .95 and .96 (males and females respectively) indicate high consistency in the pattern of responses, but say nothing about the magnitude of the score for a given response. Such findings perhaps have more to say about content validity than about internal consistency reliability in the usual sense. Consistency in pattern of response would be evidence that the items are systematically related to each other and thus derived from a uniform content dimension, or, in Guttman's terminology, have "unidimensionality." Guttman scaling does not permit us to differentiate the relative contributions of different items to the total score; every item by definition has equal weight.

Subsequent checks on scalability for the 8th and 9th year administrations of the SCA scale in connection with Project #1636 have shown reproducibility coefficients of .96 and .97 for males in the 8th and 9th years, and .92 and .93 for females in the same years. These subsequent analyses were, however, computed for a random sample of 35 males and 35 females rather than for the total population as was the case for the 7th year administration. As all computed reproducibility coefficients have been above the accepted .90 level, it is concluded that the self-concept of ability items are scalable and constant over time.

Hoyt Method of Reliability Determination.--The Hoyt method of comparing error variance to individual variance was applied to the entire population. The computed reliability coefficients were .82 and .84 for males and females respectively. These figures are below the desired .90 level but this is not surprising considering how few items are in the SCA Scale. If, for example, the Scale had consisted of 16 items comparable to those actually used, the ratio of individual to error variance would give reliabilities of .90 or better. The obtained reliabilities also compared favorably with internal consistency reliability measures on other self-concept measures. (See, for example, coefficients reported by Wylie for various instruments, pp. 87-98.[2])

Empirical Determination of Subscales.--Empirically, cluster analysis was run to determine if subsets of items were more highly related to each other than to items outside the subset. Using the simple McQuitty technique,[3] items 1, 2, 7, 8 and items 3, 4, 5, 6 formed separate clusters for males; while items 1, 2, 3, 4, 5, 6 and items 7, 8 showed as separate clusters for females. However, the correlations between some of the items in separate clusters were not significantly higher than correlations within a cluster making an interpretation based on a single

---

[2]Wylie, op. cit., pp. 87-98.

[3]McQuitty, op. cit.

cluster more realistic. Specifically, for males the correlation of item 4 with item 6 was .46 (these items are in the same cluster); the correlation of item 4 with item 8 was .45 (and these items were assigned to different clusters by the McQuitty technique). Similar findings occurred with the females. The McQuitty technique assigns items to clusters based only on the highest correlation that a given item has with each other item. The technique does not allow looking at, for example, second highest correlations which may be only .01 lower.

Because of the ambiguity in the results from the cluster analysis, it was decided to run a centroid factor analysis on the items to see if subgroups did, in fact, show up. Unlike cluster analysis which assigns all the variance from a given item to one cluster or other, factor analysis distributes the variance from a given item and allows it to show up in one or more factors. The loadings for each item on the first two factors extracted are given in Table 13. It should be noted that Factor II barely meets the requirements of a nonerror factor using Hymphreys Rule.[4]

Most conspicuous in the table are the very high loadings on all items of Factor I which is presumed to be self-concept of ability. This lends support to the assertion

---

[4]See, Benjamin Fruchter, Introduction to Factor Analysis (New York: D. Van Nostrand Co., 1954), p. 79 for a discussion of "Humphrey's Rule" and other tests for determination of nonerror factors.

that the McQuitty cluster analysis is best interpreted as a single cluster. Further, the factor analysis gives no support for the clusters which were found by the McQuitty system.

TABLE 13.--Loadings on the eight SCA items for the two significant factors extracted by the centroid method for 513 males and 537 females.

Item No.	Factor I		Factor II	
	Males	Females	Males	Females
1	.58	.66	-.40	.16
2	.65	.65	-.29	.20
3	.69	.67	-.09	-.06
4	.66	.72	.37	-.26
5	.71	.70	.10	-.30
6	.55	.53	.29	-.32
7	.68	.61	-.08	.34
8	.60	.64	.09	.25

Of interest is the discovery of a second factor. It is clearly a minor factor compared with the first factor, yet appears to be other than error. The best explanation of this factor is that it is the time dimension which was predicted by one logical analysis of the instrument (see next section). That many of the signs on the loadings are reversed in Factor II for males and females is an artifact of the method--the time dimension was approached from

opposite ends for the two sexes in the process of extraction. The future-oriented items on the scale were 3, 4, 5, 6; the present-oriented items were 1, 2, 7, 8. For the females all loadings are consistent with a time factor interpretation; for the males only the two lowest loadings are inconsistent and they are low enough to be interpreted as error. While the items show considerable similarity in Factor I loadings between males and females, there is clearly less agreement on Factor II loadings. Males and females appear to be more similar on future-oriented items. While such sex differences are not yet understood, it does appear warranted to conclude that in general the SCA scale is basically unidimensional with a minor time factor operating secondarily.

Logical Determination of Subscales.--Logical analysis of the content of the SCA Scale items resulted in consideration of two basic subdimensions. The first was a time dimension and gave subsets 1 and 2 which consisted of current time reference and future time reference items respectively. The other subdimension was based on a comparative vs absolute evaluation of ability and resulted in subsets 3 and 4 respectively. (Item 7 was eliminated from this analysis because of its noncomparable content--from a logical point of view.) The items involved in each subset are given below.

Subset 1: Current Time Reference (Items 1, 2, 8)

Subset 2: Future Time Reference (Items 3, 4, 5, 6)

Subset 3: Comparative Evaluation (Items 1, 2, 3, 5)

Subset 4: Absolute Evaluation (Items 4, 6, 8)

The specific content of items may be seen in Appendix B.

Subset total scores were correlated with each other, with GPA and the total SCA scores to see possible differential relationships. The subset intercorrelations are shown in Table 14; subset correlations with SCA and GPA are shown in Table 15.

TABLE 14.--Correlation matrix showing the intercorrelations of four subsets of items for 506 males (below the diagonal) and 534 females (above the diagonal).

Subset		1	2	3	4
1	Present	---	.61	.84	.69
2	Future	.61	---	.80	.90
3	Comparative	.82	.78	---	.64
4	Absolute	.68	.89	.59	---

TABLE 15.--Correlations of four logically-determined item subsets with GPA and SCA total scores for 506 males and 534 females.

	Subset	GPA		SCA	
		Males	Females	Males	Females
1	Present	.61	.64	.85	.86
2	Future	.42	.41	.93	.91
3	Comparative	.43	.46	.88	.90
4	Absolute	.54	.55	.89	.89

From Table 14 it can be seen that subsets 1 and 2 correlate .61 for each sex; subsets 3 and 4 correlate with each other .59 (males) and .64 (females). The correlations are high enough to indicate considerable communality between subsets, but low enough to suggest these particular subsets are not equivalent. The much higher correlations between the other subsets are in part the result of these other subsets having some of the same items in common. Nevertheless, it is interesting to note the much higher correlations between subsets 2 and 4 than between subsets 2 and 3. The significantly higher correlation between subsets 2 and 4 suggests that future projections of ability are more highly related to absolute rather than comparative evaluations of ability.

The subset correlations with the total SCA scale are shown in Table 15. Again the difference between the present and future subsets is the most conspicuous as would be expected given the results of the factor analysis. Such results, of course, also suggest that the second factor was correctly interpreted. The ability of the various subsets to predict GPA is also shown in Table 15. It is hardly surprising that present-oriented items are significantly better predictors of present achievement than are future-oriented items. A less obvious finding is that absolute evaluations of ability are significantly better predictors of achievement than comparative evaluations. These young

people are apparently not oversensitized to the "curve" which would emphasize relative standing.

Although present-oriented items are more effective for predicting present achievement, it should not be assumed that they will be equally effective in predicting future achievement. One might expect that the future oriented items would be better predictors of long-range achievement. This hypothesis would be worthy of test in subsequent research.

Individual Item Examination

To further understand the nature of the SCA instrument, individual items were examined for their correlations with each other, the total score and GPA. The correlation matrix is given in Table 16.

Consistent with findings from the logical analysis of subsets, the individual items which correlate most highly with GPA are 1, 2, 4, 7, 8. The item content for these items suggests that they all ask questions which are current and meaningful to the students. Item 4, the only future-oriented item is still concerned with a topic that 7th graders have thought about (going to college), unlike, for example item 6 which asks about graduate school. Items 3 and 5 both ask about "rank"--possibly a nonmeaningful term to 7th graders. One would guess these items would do better at the high school level where students are constantly hearing about the "top fifth", etc.

For both sexes, the single best item for predicting GPA is item 8 (What grades are you capable of getting?). That this correlation with GPA is so high suggests that students feel they are actually working to capacity. The correlation is higher for females than males, a finding consistent with the common notion that boys are more likely to "goof off" and not work to capacity.

Examination of item-total score correlations (Table 16) suggests that items 4 and 5 seem to best characterize the SCA scale as a whole. Both questions concern college--ability to complete college, and probable rank in class. These items also have among the highest loadings on Factor I as would be expected. But perhaps more important are the consistently high correlations for all items, again consistent with the findings from the factor analysis. None of the item-total-score correlations are below .59; only two are below .65. This, of course, is further evidence of the basic unidimensionality of the instrument.

While it is clear that all items contribute somewhat comparably to the SCA total score, it is not true that the items operate in parallel fashion when used to predict achievement. If the primary purpose of using the SCA inventory is to predict achievement, items 3, 5, 6, and possibly 1 might well be eliminated when the scale is administered to 7th graders. In fact item 8 alone predicts achievement as well as the full SCA scale. A higher correlation might come from a combination of items 2, 4,

TABLE 16.--Correlation matrix showing intercorrelations of eight SCA items, the SCA total score, and GPA for 506 males (below diagonal) and 534 females (above the diagonal).

Variables		1	2	3	4	5	6	7	8	9	10
1	Item 1	---	.48	.48	.39	.42	.30	.46	.41	.68	.39
2	Item 2	.50	---	.46	.42	.39	.25	.42	.45	.68	.47
3	Item 3	.44	.43	---	.43	.55	.29	.33	.40	.69	.29
4	Item 4	.25	.38	.38	---	.58	.50	.35	.46	.77	.45
5	Item 5	.38	.38	.58	.53	---	.39	.33	.36	.72	.27
6	Item 6	.24	.30	.33	.46	.40	---	.27	.23	.63	.23
7	Item 7	.40	.49	.42	.43	.43	.32	---	.51	.65	.43
8	Item 8	.26	.37	.38	.45	.36	.32	.48	---	.68	.64
9	SCA Total	.59	.67	.70	.74	.74	.66	.71	.67	---	.57
10	GPA	.32	.49	.28	.45	.24	.28	.56	.56	.57	---

7, 8. With different age levels, of course, the ability of individual items to predict achievement will undoubtedly change.

In conclusion, empirical analysis of the SCA Scale suggests that the Scale has moderately high internal consistency reliability. The items appear to be uniformly related to a single factor although a minor second factor was found which was interpreted as a time dimension. Guttman analysis clearly indicated the items were a unidimension; cluster and factor analysis supported this assumption in general. Examination of subsets and individual items in

the prediction of GPA, however, revealed variation in predictive ability by subset or individual items. Suggestions for modification of the Scale to maximize prediction of GPA were made.

CHAPTER V

SUMMARY AND CONCLUSIONS

The present research analyzed the principal measuring instrument employed in the U. S. Office of Education Project #845 and subsequent research. The instrument, "Self-Concept of Ability Scale" (SCA Scale) is a self-report inventory consisting of eight multiple-choice items with five response alternatives. The Scale is designed to tap the construct self-concept of ability, ability in this case being restricted to academic ability. All items were constructed so as to be appropriate for junior high level students who were the population under study in this research. A set of four specific subject scales was also designed directly parallel to the SCA Scale except that their content was restricted to a specific academic subject. The four subjects were arithmetic, English, social studies, and science. The Specific Subject Self-Concept of Ability (Specific SCA) Scales were designed for the study of specific subject variations in the self-concept-achievement relationship and to serve as a "boundary" on one end of the general self-concept of ability construct. The specific subject scales were not the focus of this

analysis but evidence is presented which indicates that in general these scales operate analagously to the more general SCA Scale.

The main focus of the research was the establishment of construct validity. To this end the theoretical base for the research was outlined and four theoretically derived hypotheses were tested by employing the SCA Scale. The conclusions from the hypothesis testing are given below followed by an overall assessment of the SCA Scale.

Summary of Results of the Instrument Evaluation

1. The SCA Scale scores are positively and significantly correlated with grade achievement (GPA). ($r_{12} = .57$) This is true even when IQ (as measured by standard tests) is held constant.
2. Scores on the SCA Scale are not reducible to IQ scores or to grade achievement scores.
3. Scores on the SCA Scale are significantly and positively correlated with grade achievement under conditions of a known negative correlation between IQ scores and grade achievement.
4. For both males and females a combination of high achievement and low SCA Scale scores is significantly less likely to occur than is the combination of high SCA Scale score and low achievement. This is taken as evidence that a high self-concept of ability is a necessary but not

sufficient basis for high achievement. This conclusion is dependent, however, upon the cut-off points used in designating "high" and "low" groups on each variable.

5. Evidence suggests that SCA Scale scores for this age level are geared to achievement relative to one's sex group rather than to an absolute (particular grade point index) level of achievement.
6. The more extreme the cut-off points used on the SCA Scale and GPA the more closely the number of cases of low GPA and high SCA or high GPA and low SCA approaches zero. This is taken as evidence to support the general positive relationship of self-concept of ability and grade achievement.
7. The Specific SCA Scales are positively and significantly correlated with achievement in parallel subjects. This is true even when IQ (as measured by standard tests) is held constant. This is taken as evidence that the Specific SCA Scales are related to achievement in a manner analagous to the relation of SCA Scale scores with general GPA.
8. The general SCA mean scores are higher than any of the Specific SCA scores but the general score is closest to the specific score in that subject in which the student has his highest achievement.

This suggests that general SCA is more heavily influenced by areas of strength than by areas of weakness.

9. Girls are significantly more likely to be "uniform" achievers than "nonuniform" achievers--as defined in the research. Males are about equally likely to be "uniform" or "nonuniform" achievers. This suggests different cultural expectations for achievement by sex.
10. Among "nonuniform" achievers of both sexes, the Specific SCA Scales were, with one exception, significantly better predictors of achievement in the parallel subject than was the general SCA Scale. This tendency was, however, much more conspicious for females than males.
11. Among "uniform" achievers, the general SCA Scale is a significantly better predictor of specific grade achievement than are the Specific SCA Scales for females (in all subjects except social studies), but this is not true for males except in English. This suggests that sex differences are operating.
12. Females who are "nonuniform" achievers have significantly lower mean SCA Scale and GPA scores than female "uniform" achievers. There are no such differences observable for male "uniform"

and "nonuniform" achievers. Again sex differences appear to be operating.

13. A theory is proposed to account for the sex differences observed in 9-12. In general, it is suggested that nonuniform achievement is culturally defined as negative for females while nonuniform achievement is, in fact, encouraged for males.

14. The general SCA Scale is a better predictor of achievement in a specific subject than is any Specific SCA Scale other than the one in the parallel subject. In 17 of the 24 comparisons the differences are significant; in all cases findings are in the predicted direction.

15. The general SCA Scale is a better predictor of general achievement than is any Specific SCA Scale. In five of the eight comparisons the differences are significant; in all cases findings are in the predicted direction. This, and conclusion 14 are evidence that the general SCA Scale is not reducible to any of the Specific SCA Scales.

16. The intercorrelations among the Specific SCA Scale scores are all higher for males than females, but the intercorrelations among specific subject grades earned are higher for females than males. A tentative explanation is offered.

Other indices of validity and reliability were also used in the evaluation of the SCA Scale. The results of these investigations follow.

17. The SCA Scale is considered to have content validity as the method of selection of the items can be considered a comprehensive sample of the construct under consideration.

18. The correlation of SCA scores obtained in the fall of 1960 with grades obtained in January 1961 is .57 for each sex. Corrected for attenuation, the correlations become .60 and .59 (males and females) or account for 36 per cent and 35 per cent of the variance.

19. When Fall 1960 SCA scores and an average of fourth and sixth grade IQ scores are compared in their ability to predict January 1961 achievement, the calculated beta weights for the two variables were .37 (SCA) and .44 (IQ) for the males; and .34 (SCA) and .49 (IQ) for females. IQ is more heavily weighted in the prediction. The multiple correlation of IQ plus SCA to predict achievement gives multiple correlation coefficients of .69 (males) and .72 (females), or, in other words, the combination of variables accounted for 48 per cent and 52 per cent of the variance for male and female achievement.

20. A cross-validation of the prediction equation used a combination of IQ and SCA scores to predict June 1961 achievement (an additional six-months time beyond that used in calculating the prediction equation). The cross-validation population was from the same school system, but included persons who had been eliminated from the main study by virtue of race or incomplete data. The correlations between predicted and obtained GPA for the cross-validation population were .71 (males) and .70 (females). These values, when compared with the multiple correlations (see #19) suggest that the original prediction equation does not capitalize on chance factors but accounts for real and stable variance.

21. The stability reliability coefficients of the SCA Scale for a 12-month interval are .75 and .77 for males and females respectively. The figures are difficult to interpret. Although self-concept of ability is assumed to be more stable than unstable, it is assumed that self-concept of ability can change. It is suggested that stability reliability coefficients should be used in combination with other reliability indicies in view of the difficulty of estimating the expected degree of instability of self-concept of ability for individuals.

22. The SCA Scale was originally constructed so as to form a Guttman Scale which requires reproducibility coefficients above the .90 level. The actual reproducibility coefficients were .95 (males) and .96 (females) for the samples under study. The existence of a Guttman Scale suggests that the items are all from the same universe of content and of unequal difficulty given the specific bases for dichotomizing responses to the items.

23. The Hoyt method of determining internal consistency reliability yielded coefficients of .82 and .84 for the total samples of 513 males and 537 females.

24. The McQuitty technique of cluster analysis of SCA Scale items resulted in two clusters of items for each sex; however, the clusters were not the same for the two sexes. Real doubt is cast on these findings because the correlations of items within clusters were not significantly different from the correlations between items in different clusters. Such results are an artifact of the method which considers only the highest correlation of an item with other items.

25. Because of the ambiguity of the results in the cluster analysis (#24), a centroid factor analysis was run on the SCA items. Two factors were extracted, the second barely meeting the requirements

of a nonerror factor. The first factor was interpreted as the self-concept of ability factor which showed strong factor loadings on each item (the lowest loading was .53, the highest .72). This is taken as evidence of the basic homogeneity of the scale (consistent with findings of Guttman scaling). However a minor second factor was found with loadings ranging from .06 to .40. This factor was interpreted as the time dimension which had been postulated by logical analysis (see #27 below).

26. A logical analysis of item content had suggested that various subsets of items were asking slightly different questions. Correlations between subset and total SCA scores indicated that each subset correlated with the total score at .85 or better, again emphasizing the basic homogeneity of the scale. The subsets dealt with future versus present oriented items; and with comparative versus absolute evaluations of ability.

27. Consistent with the finding of a secondary time dimension in the factor analysis, the subscore from the four future items correlated only .61 with the subscore from the three present time-oriented items. The subscores from the absolute and comparative evaluation items correlated only .59 (males) and .64 (females) with each other.

The latter correlations which are unexpectedly low and not consistent with other results may, however, result from contamination by the time dimension noted above. Evidence for this is found in the significantly higher correlations between future and absolute subsets than between future and comparative subsets. Present-oriented items as a subset are more highly correlated with comparative evaluation although not significantly so.

28. In predicting GPA the present-oriented (and absolute) evaluation subsets were significantly better predictors than their theoretical opposites. It was hypothesized, however, that for future projection of achievement the future-oriented evaluations might be more predictive.

29. Individual item examination was done by means of a correlation matrix comparing the individual items in their relation to the total SCA score and GPA. The individual items which correlate most highly with the total SCA score are items four and five. These two items also have the highest factor loadings on Factor I as would be expected. Again, consistent with the factor analysis, none of the item-total correlations falls below .59; only two are below .65.

30. Individual items varied considerably in their ability to predict GPA. The single best item is item 8 (What grades do you think you are capable of getting?). This one item predicts achievement as well as the full SCA Scale. The poorest predictors of achievement for each sex are items 3, 4, and 6--all future-oriented items.

A General Assessment of the Self-Concept of Ability Scale

The cumulative evidence from research employing the Self-Concept of Ability Scale suggests that it is able to accurately predict theoretically derived relationships suggested by the symbolic interactionist framework. Use of the SCA Scale allows prediction of criterion behavior over time and after controlled-experiment intervention. There is evidence that scores from the Scale are not reducible to IQ scores, past or present achievement, or student attitudes toward the importance of getting good grades. The Scale is able to discriminate general self-attitudes toward ability to achieve from self-attitudes toward ability to achieve in specific school subjects.

Guttman scalogram analysis, factor analysis and individual item analysis all affirm the basic homogeneity of the Scale with respect to content, although a minor time dimension was found in factor analysis which distinguished present-oriented from future items. This time dimension was particularly evident in the prediction of the criterion

of grade point average; GPA (at the next grading period) was more adequately predicted by present-oriented items. It may appear somewhat inconsistent to conclude that a scale has internal homogeneity and yet that the items are not equally able to predict a criterion. The problem is the type of homogeneity one is considering. There is an essential difference between the homogeneity of a compound and an element. All samples of the compound may be like all other samples, and yet the compound is ultimately composed of different elements. By analogy, self-concept of ability may be consistently viewed as homogeneous and yet have a known time dimension operating. For prediction purposes one type of homogeneity may be more relevant than another. Here it is suggested that the criterion of grades earned in the near future is best predicted from the present-oriented "elements" of the Scale, while long-range achievement may be better predicted from the future-oriented "elements." The criterion measure must be considered homogeneous in an equivalent manner.

Despite the general positive results obtained from use of the SCA Scale, there are some obvious gaps in our knowledge about this measuring instrument. These are suggested in some detail in the discussion of construct validity in Chapter II. In particular we do not know the influence of social desirability, instrument form, or response restrictiveness on results using the Scale. An equally serious problem lies in the examination of the

psychological relevance of scoring procedures. Perhaps the most serious deficiency in knowledge is the inability to demonstrate that results from use of the Scale are not method-tied. There is no independent measure of the self-concept of ability construct to allow such evaluation. A related problem is that the results may as likely be criterion-tied. Grade point average is the only criterion so far systematically examined; other achievement indicies should be studied as well.

There is an obvious need to articulate the present research findings with other measures of self-concept--particularly self-conceptions about ability in other than academic areas. This will serve to demonstrate empirically how broad a construct is tapped by the SCA Scale much as analysis of the specific subject scales demonstrated how narrowly results using the Scale should be interpreted.

Until there is empirical evidence as to the probability of change in academic self-concept (change induced by other than experimental intervention), there is no way to determine the adequacy of the stability reliability of the instrument. Further study also should be done to determine if the internal consistency reliability could be increased profitably by modification in the Scale itself. Perhaps elimination of one or more items is called for. Logic suggests that item 7 should be removed from the Scale; empirical evidence supports its inclusion. It may well be that item 7 works because of the level of the

students--junior high students simply do not discriminate between performance and potential for performance. It may also be true that this distinction exists primarily in the minds of the test constructors and is not important to the general public.

Another knowledge gap comes from the general lack of use of the instrument with other age groups. While there is no reason to believe that the SCA Scale could not be satisfactorily modified for use with a variety of age levels, this has not been systematically demonstrated. Also, while there is some evidence that the Scale can be used across different IQ and social class levels, detailed analysis by sub-group has not been done. There is, however, ample evidence to suggest that all analyses employing the Scale should be done separately by sex.

And finally, despite our incomplete understanding of the Self-Concept of Ability Scale, it is encouraging that there is at present no evidence suggesting that the SCA Scale does not adequately index the construct it was designed to tap. It therefore seems warranted to conclude that we have support for the two basic underlying assumptions of the research (1) that the self-concept of ability construct is phenomenological, and (2) that self-concept of ability can be reliably indexed by a self-report instrument.

SELECTED BIBLIOGRAPHY

SELECTED BIBLIOGRAPHY

American Psychological Association, American Educational Research Association, and National Council on Measurements Used in Education, Joint Committee. "Technical Recommendations for Psychological Tests and Diagnostic Techniques," Psychological Bulletin, 51, No. 2, Part 2, Special Supplement (March, 1954).

Braithwaite, Richard B. Scientific Explanation. Cambridge: Cambridge University Press, 1953.

Brookover, Wilbur B. "A Social Psychological Conception of Classroom Learning," School and Society, 87 (February 28, 1959), 84-87.

Brookover, Wilbur B., and Gottlieb, David. A Sociology of Education. Second edition. New York: American Book Company, 1964.

Brookover, Wilbur B., Paterson, Ann, and Thomas, Shailer. Self-Concept of Ability and School Achievement. Final Report of Cooperative Research Project No. 845. East Lansing, Michigan: Office of Research and Publications, College of Education, Michigan State University, December, 1962.

Brookover, Wilbur B., and others. Self-Concept of Ability and School Achievement, II. Final Report on Cooperative Research Project No. 1636. East Lansing, Michigan: Bureau of Educational Research Services, College of Education, Michigan State University, October, 1965.

Campbell, Donald T., and Fiske, Donald W. "Convergent and Discriminant Validation by the Multitrait-multimethod Matrix," Psychological Bulletin, 56 (March, 1959), 81-105.

Coleman, James S., and others. The Adolescent Society. Glencoe, Illinois: The Free Press, 1961.

Combs, Arthur W., and Snygg, Donald. Individual Behavior. New York: Harper and Brothers, 1959.

Combs, Arthur W., Soper, Daniel W., and Courson, Clifford C. "The Measurement of Self-Concept and Self Report," Educational and Psychological Measurement, 23, No. 3 (Autumn, 1963), 493-500.

Cooley, Charles H. Social Organization. New York: Charles Scribners Sons, 1909.

Cronbach, Lee J., and Meehl, Paul E. "Construct Validity in Psychological Tests," Psychological Bulletin, 52 (July, 1955), 281-302.

Fruchter, Benjamin. Introduction to Factor Analysis. New York: D. Van Nostrand Co., 1954.

Garner, Wendell R., Hake, Harold W., and Eriksen, Charles W. "Operationalism and the Concept of Perception," Psychological Review, 63 (March, 1956), 149-59.

Goode, William J., and Hatt, Paul K. Methods in Social Research. New York: McGraw Hill, 1952.

Guilford, J. P. Fundamental Statistics in Psychology and Education. New York: McGraw Hill, 1956.

Jourard, Sidney M., and Lasakow, Paul. "Some Factors in Self-disclosure," Journal of Abnormal and Social Psychology, 56 (January, 1958), 91-98.

Kinch, John W. "A Formalized Theory of the Self-Concept," The American Journal of Sociology, 68 (1963), 481-86.

Lindquist, E. F. (ed.) Educational Measurement. Washington, D. C.: American Council on Education, 1951.

McQuitty, Louis L. "Elementary Linkage Analysis for Isolating Orthogonal and Oblique Types and Typal Relevancies," Educational and Psychological Measurement, 17, No. 2 (Summer, 1957), 207-229.

Mead, George H. Mind, Self, and Society. Chicago: University of Chicago Press, 1934.

Morse, Richard J. "Self-Concept of Ability, Significant Others and School Achievement of Eighth Grade Students: A Comparative Investigation of Negro and Caucasian Students." Unpublished Master's thesis, Michigan State University, 1963.

Reeder, Thelma Adams. "A Study of Some Relationships Between Level of Self-Concept, Academic Achievement and Classroom Adjustment." Unpublished Ph.D. thesis, North Texas State College, 1955.

Renzaglia, G. A. "Some Correlates of the Self Structure as Measured by an Index of Adjustment and Values." Unpublished Ph.D. thesis, University of Minnesota, 1952.

Rose, Arnold M. (ed.) Human Behavior and Social Processes. Boston: Houghton-Mifflin, 1962.

Sarason, S., and Mandler, G. "Some Correlates of Test Anxiety," Journal of Abnormal and Social Psychology, 47 (October, 1952), 810-17.

Staines, J. W. "Self-Picture as a Factor in the Classroom," British Journal of Educational Psychology, 28 (June, 1956), 97-111.

Stern, George, Stein, M. I., and Bloom, B. S. Methods of Personality Assessment. New York: Free Press of Glencoe, 1956.

Wylie, Ruth C. The Self-Concept: A Critical Survey of Pertinent Research Literature. Lincoln, Nebraska: University of Nebraska Press, 1961.

APPENDICES

APPENDIX A

SELF-CONCEPT OF ABILITY SCALE--

PRE-TEST ITEMS

1. How do you rate yourself in school ability compared with your close friends?

   a. I am the best.
   b. I am above average
   c. I am average
   d. I am below average
   e. I am the poorest
   f. don't know

2. How do you rate yourself in school ability compared with your brother(s) and sister(s)?

   a. I am better
   b. I am about the same
   c. I am poorer
   d. I have no brothers or sisters
   e. don't know

3. How do you rate yourself in school ability compared with those in your class at school?

   a. I am among the best
   b. I am above average
   c. I am average
   d. I am below average
   e. I am among the poorest
   f. don't know

4. How do you rate yourself in school ability compared to other young people your age?

   a. I am among the best
   b. I am above average
   c. I am average
   d. I am below average
   e. I am among the poorest
   f. don't know

5. Do you think you have the ability to complete high school?

   a. yes, definitely
   b. yes, probably
   c. don't know
   d. probably not
   e. no

6. Where do you think you would rank in your class in high school?

   a. among the best
   b. above average
   c. average
   d. below average
   e. among the poorest
   f. don't know

7. At the graduation ceremonies in high school they give out awards and prizes for superior school work. How likely do you think it is that you would receive a prize or award?

   a. very likely
   b. somewhat likely
   c. don't know
   d. unlikely
   e. most unlikely

8. Do you think you have the ability to complete college?

   a. yes, definitely
   b. yes, probably
   c. don't know
   d. probably not
   e. no

9. Where do you think you would rank in your class in college?

   a. among the best
   b. above average
   c. average
   d. below average
   e. among the poorest
   f. don't know

10. At the graduation ceremonies in college they give out awards and prizes for superior school work. How likely do you think it is that you would receive a prize or award?

   a. very likely
   b. somewhat likely
   c. don't know
   d. unlikely
   e. most unlikely

11. In order to become a doctor, lawyer, or university professor, work beyond four years of college is necessary. How likely do you think it is that you could complete such advanced work?

    a. very likely
    b. somewhat likely
    c. don't know
    d. unlikely
    e. most unlikely

12. Forget for a moment how others grade your work. In your own opinion how good do you think your work is?

    a. my work is excellent
    b. my work is good
    c. my work is average
    d. my work is below average
    e. my work is much below average
    f. don't know

13. Do you think your school work tends to show

    a. the best work you are capable of doing.
    b. some of the better work you are capable of doing.
    c. average work on your part.
    d. poorer work than you are capable of doing.
    e. much poorer work than you are capable of doing.
    f. don't know

14. How hard do you find you have to work in school?

    a. very hard
    b. hard
    c. average amount
    d. not very hard
    e. not hard at all
    f. don't know

15. How hard do you have to work in school compared with other students in your class?

    a. harder than any of them
    b. harder than most of them
    c. about as hard as most of them
    d. not as hard as most of them
    e. not as hard as any of them
    f. don't know

16. What kind of grades do you think you are capable of getting?

a. mostly A's
b. mostly B's
c. mostly C's
d. mostly D's
e. mostly E's
f. don't know

APPENDIX B

SELF-CONCEPT OF ABILITY SCALE--
AS ADMINISTERED

Circle the letter in front of the statement which best answers each question.

1. How do you rate yourself in school ability compared with your close friends?

   a. I am the best
   b. I am above average
   c. I am average
   d. I am below average
   e. I am the poorest

2. How do you rate yourself in school ability compared with those in your class at school?

   a. I am among the best
   b. I am above average
   c. I am average
   d. I am below average
   e. I am among the poorest

3. Where do you think you would rank in your class in high school?

   a. among the best
   b. above average
   c. average
   d. below average
   e. among the poorest

4. Do you think you have the ability to complete college?

   a. yes, definitely
   b. yes, probably
   c. not sure either way
   d. probably not
   e. no

5. Where do you think you would rank in your class in college?

   a. among the best
   b. above average
   c. average
   d. below average
   e. among the poorest

6. In order to become a doctor, lawyer, or university professor, work beyond four years of college is necessary. How likely do you think it is that you could complete such advanced work?

   a. very likely
   b. somewhat likely
   c. not sure either way
   d. unlikely
   e. most unlikely

7. Forget for a moment how others grade your work. In your own opinion how good do you think your work is?

   a. my work is excellent
   b. my work is good
   c. my work is average
   d. my work is below average
   e. my work is much below average

8. What kind of grades do you think you are capable of getting?

   a. mostly A's
   b. mostly B's
   c. mostly C's
   d. mostly D's
   e. mostly E's

APPENDIX C

SPECIFIC SUBJECT SELF-CONCEPT OF ABILITY SCALES-- AS ADMINISTERED

Now we would like you to again answer some of the same questions, but this time about four different subjects which you are now taking or have taken in the past.

Circle the "X" under the heading which best answers the question. Answer for all four subjects. (You will have one "X" circled on each line.)

1. How do you rate your ability in the following school subjects compared with your close friends?

	I am the poorest	I am below average	I am average	I am above average	I am the best
Arithmetic	X	X	X	X	X
English	X	X	X	X	X
Social Studies	X	X	X	X	X
Science	X	X	X	X	X

2. How do you rate your ability in the following school subjects compared with those in your class at school?

	I am among the poorest	I am below average	I am average	I am above average	I am among the best
Arithmetic	X	X	X	X	X
English	X	X	X	X	X
Social Studies	X	X	X	X	X
Science	X	X	X	X	X

3. Where do you think you would rank in your high school graduating class in the following subjects?

	among the poorest	below average	average	above average	among the best
Arithmetic	X	X	X	X	X
English	X	X	X	X	X
Social Studies	X	X	X	X	X
Science	X	X	X	X	X

4. Do you think you have the ability to do college work in the following subjects?

	no	probably not	not sure either way	yes, probably	yes, definitely
Arithmetic	X	X	X	X	X
English	X	X	X	X	X
Social Studies	X	X	X	X	X
Science	X	X	X	X	X

5. Where do you think you would rank in your college class in the following subjects?

	among the poorest	below average	average	above average	among the best
Arithmetic	X	X	X	X	X
English	X	X	X	X	X
Social Studies	X	X	X	X	X
Science	X	X	X	X	X

6. How likely do you think it is that you could complete advanced work beyond college in the following subjects?

	most unlikely	unlikely	not sure either way	somewhat likely	very likely
Arithmetic	X	X	X	X	X
English	X	X	X	X	X
Social Studies	X	X	X	X	X
Science	X	X	X	X	X

7. Forget for a moment how others grade your work. In your own opinion how good do <u>you</u> think your work is in the following subjects?

	my work is much below average	my work is below average	my work is average	my work is good	my work is excellent
Arithmetic	X	X	X	X	X
English	X	X	X	X	X
Social Studies	X	X	X	X	X
Science	X	X	X	X	X

8. What kind of grades do you think you are capable of getting in the following subjects?

	mostly E's	mostly D's	mostly C's	mostly B's	mostly A's
Arithmetic	X	X	X	X	X
English	X	X	X	X	X
Social Studies	X	X	X	X	X
Science	X	X	X	X	X