1973

A Multitrait-Multimethod Factor Analytic Appraisal of Three Measures of Social Intelligence

Patrick. McKian
Loyola University Chicago

Follow this and additional works at: https://ecommons.luc.edu/luc_theses

Recommended Citation
https://ecommons.luc.edu/luc_theses/2710

This Thesis is brought to you for free and open access by the Theses and Dissertations at Loyola eCommons. It has been accepted for inclusion in Master's Theses by an authorized administrator of Loyola eCommons. For more information, please contact ecommons@luc.edu.

This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 License. Copyright © 1973 Patrick. McKian
A MULTITRAIT-MULTIMETHOD FACTOR ANALYTIC APPRAISAL
OF THREE MEASURES OF SOCIAL INTELLIGENCE

by

Patrick McKian

A Thesis Submitted to the Faculty of the Graduate School
of Loyola University of Chicago in Partial Fulfillment
of the Requirements for the Degree of
Master of Arts

November
1973
ACKNOWLEDGMENTS

Special gratitude is respectfully and gratefully offered to Ronald E. Walker, Ph.D. who served as director of this thesis and to Thomas Petzel, Ph.D. who also served on the committee for this thesis. Their encouragement and guidance in this research proved invaluable.

Special thanks is also offered to Emil Posavac, Ph.D. and Frank Slaymaker, Ph.D. whose assistance in data analysis was very helpful.
VITA

Patrick Joseph McKian was born December 6, 1945 in Detroit, Michigan. He is the son of Francis McKian and Anna (McGarry) McKian.

His elementary education was obtained in the Catholic school system of Detroit. His secondary education took place at St. Joseph's preparatory seminary in Edgerton, Wisconsin, where he was graduated in June, 1963.

In September, 1964, he entered the first of two years of college work at St. Joseph's college seminary from which he graduated in June, 1965. In August, 1965 he entered the Redemptorist novitiate and professed vows in August, 1966. His college education continued at Immaculate Conception seminary in Oconomowoc, Wisconsin where he spent one year. His final year was completed at the University of Detroit where he was graduated in December, 1968.

The author entered the graduate program in Clinical Psychology at Loyola University of Chicago in September, 1969. He spent two 500 hour clerkships at Downey V.A. Hospital between October, 1969 and October, 1971. In November, 1971 he began the first of two 1200 hour clinical experiences which combined served as his internship. These were done between November, 1971 and October, 1973 at Westside V.A. Hospital.

Presently, the author is working at the Loyola Child Guidance Center.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>ii</td>
</tr>
<tr>
<td>VITA</td>
<td>iii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>vi</td>
</tr>
</tbody>
</table>

**Chapter**

I. **INTRODUCTION** ................................. 1

II. **REVIEW OF THE RELATED LITERATURE** .......... 5

The Ability to Understand People and  
Its Relationship to Intelligence .......... 5  
Tests of Social Intelligence .......... 18

Six Factor Tests of Social  
Intelligence .......... 18
Cognition of Behavior in Terms of  
Implications .......... 20
Cognition of Behavior in Terms of  
Classes (CBC) .......... 21
Cognition of Behavior in Terms of  
Systems (CBS) .......... 21
Cognition of Behavior in Terms of  
Transformations (CBT) .......... 22
The Chapin Social Insight Test (CHSIT) .......... 25
The Hogan Empathy Scale (HES) .......... 27

Tests of Abstract Intelligence .......... 30

The Quick Word Test (QWT) .......... 30
The Otis Quick-Scoring Mental Ability  
Tests: Gamma, Form FM .......... 32
The Kuhlmann-Anderson Test .......... 33

Hypotheses .......... 35

III. **METHOD** ................................. 37

Subjects .......... 37
Materials .......... 37
Procedure .......... 38
### TABLE OF CONTENTS.--Continued

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>IV. RESULTS</td>
<td>40</td>
</tr>
<tr>
<td>Relationship of Social Intelligence Measures to Abstract Intelligence Measures</td>
<td>40</td>
</tr>
<tr>
<td>Factor Analysis</td>
<td>47</td>
</tr>
<tr>
<td>V. DISCUSSION</td>
<td>49</td>
</tr>
<tr>
<td>Relationship of Social Intelligence Measures to Abstract Intelligence Measures</td>
<td>49</td>
</tr>
<tr>
<td>VI. SUMMARY</td>
<td>56</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>59</td>
</tr>
</tbody>
</table>
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>List of Variables with Their Means and Standard Deviations for the Males,</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td>Females, and Total Sample</td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>Comparison of Males and Females on All Measures</td>
<td>42</td>
</tr>
<tr>
<td>3.</td>
<td>Initial Correlation Matrix</td>
<td>44</td>
</tr>
<tr>
<td>4.</td>
<td>Correlation Matrix for Males</td>
<td>45</td>
</tr>
<tr>
<td>5.</td>
<td>Correlation Matrix for Females</td>
<td>46</td>
</tr>
<tr>
<td>6.</td>
<td>Varimax Rotated Factor Matrix</td>
<td>48</td>
</tr>
</tbody>
</table>
CHAPTER I

INTRODUCTION

In a recent paper, Walker and Foley (1973) present a review of the history of the concept of social intelligence (SI) and various attempts to measure it. Social Intelligence (SI) as originally defined by Thorndike (1920) is "the ability to understand and manage men and women, boys and girls--to act wisely in human relations" (p. 228). Work on social intelligence has branched off into two main areas: one dealing with the development and evaluation of instruments to measure this ability (Gough, 1965; Lindgren & Robinson, 1953; Thorndike & Stein, 1937), and interest and research focused on the ability to judge others, using such methods as ratings of self and others. While interest in the area of social intelligence fluctuated, interest in interpersonal judgments continued consistently and developed into what is now referred to as person perception or interpersonal processes. Walker and Foley point out several reasons for this divergence, not the least of which is the relative failure to construct an adequate test of social intelligence.

It is for this purpose that the present study is being conducted. Though several tests purporting to measure
social intelligence have been developed (e.g., Chapin, 1942; Dymond, 1948, 1949, 1950; Dymond, Hughes, & Raabe, 1952; Kerr & Speroff, 1947; Moss, Hunt, Omwake, & Ronning, 1927; Sargent, 1953) all of these correlated in the moderate to high positive direction with abstract-verbal intelligence. As a result, one does not know whether SI is a distinct and separate ability or merely a part of the more sweeping ability of general intelligence. As Cronbach (1960) commented: "No evidence of validity is yet available which warrants confidence in any present technique for measuring a person's ability to judge others as individuals. . . . After 50 years of intermittent investigation . . . social intelligence remains undefined and unmeasured" (pp. 319-20).

This study hopes in some way to come a little closer to an answer to this question. Part of the problem appears to be that all of the previous tests, with the exception of Wedeck's (1947), have been predominantly verbal in form with the result that abstract-verbal intelligence is naturally going to correlate highly with them. This study will take three measures of SI and three measures of abstract intelligence (AI) and factor analyze the correlations using a varimax rotation.

The three measures of SI chosen for the study are the Six Factor Tests of Social Intelligence (SFTSI; O'Sullivan, Guilford, & deMille, 1965; O'Sullivan & Guilford, 1966), the Chapin Social Insight Test (CHSIT; Chapin, 1942), and the
Hogan Empathy Scale (HES; Hogan, 1969). The SFTSI, following Wedeck's (1947) lead, attempts to use nonverbal material for the assessment of SI. The CHSIT has been available for 30 years, but aside from Gough's reevaluation (1965) no published research is available, even though Gough indicated it is a promising research instrument. Finally, Hogan's test purports to measure empathy through an assessment of a number of items from the Minnesota Multiphasic Personality Inventory (MMPI; McKinley & Hathaway, 1943) and the California Psychological Inventory (CPI; Gough, 1964). No research except his introductory article (1969) has been published but good reliability and validity indices were cited. Although both Chapin's and Hogan's tests rely on verbal presentation of items, they seem to have promise along with the SFTSI, as assessments of social intelligence.

By comparing these three instruments with three measures of general intelligence (Quick Word Test, Otis Quick-scoring Tests of Mental Ability, and the Kuhlmann-Anderson Test, 7th edition, Booklet H) and factor analyzing the correlations, one will be able to obtain the purest measure of SI and determine if in fact, social intelligence is a separate ability.

Before looking at each of these instruments, a review of the major studies in the area of the ability to understand others and their relationship to intelligence will be presented. This will include studies both from the areas of
person perception and social intelligence. Although these developed separately, there seems to be some overlap and borrowing of terms and methodologies. Furthermore, the influence of general (abstract-verbal) intelligence has had a great effect on both areas.
CHAPTER II

REVIEW OF THE RELATED LITERATURE

The Ability to Understand People and Its Relationship to Intelligence

According to Bruner and Taguiri (in Lindzey, 1954) early studies in the judgment of personality concentrated on the accuracy of the perceptual judgment, and grew out of a combination of such interests as validation of test procedures by independent judges, assessment of traits of personality and the establishment of criteria used for determining perceptual accuracy, and concern for what constituted a good judge of personality. They reviewed a set of studies investigating the personality characteristics associated with the ability to judge others accurately or to judge oneself accurately, with consensual agreement providing the criteria of ability. If those able to judge their own traits as others do were found to differ from those able to judge others in agreement with fellow judges, then one might conclude that the two abilities stem from different sources. In one study, Adams (1927) asked eight teams of 10 girls each to rate themselves and each other on 63 traits and found that the good judge of self and the good judge of others differed in personality traits. In general, he found that the good self-
rater tended to be happier, more intelligent, more socially minded and possessed a greater degree of the more "emotional attributes." They were also less irritable, more sympathetic and generous than the good rater of others. On the other hand, the good judge of others was more independent, was gregarious rather than social, and showed relatively little interest in persons. Adams described them as tending toward self-centeredness and regarding others as tools for their own satisfaction. This shrewd ability to measure others or "understand others" as Thorndike (1920, p. 228) would in part define social intelligence is similar to the dynamics of the sociopath. However, the ability to "act wisely" which Thorndike sees as an integral and essential part of social intelligence, often seems to be lacking in sociopathic behavior. It seems unlikely that either Thorndike or Adams had the sociopath in mind when they were discussing social intelligence and the characteristics of the good judge of others, but it does point out one of the possible ramifications and points of departure for further research in the area.

A paradox in Adams' study which Vernon (1933) confirmed is that the judge most interested in others understands himself the best and the judge most interested in himself understands others best. Vernon also discovered that the accuracy of the judgment depends not only on the subject being judged but on the context of the judgment and
the conditions under which a judgment is given as well.

Taft (1950) essentially confirmed the findings but adds that while the good judge of self has the desirable qualities mentioned above, he is also according to independent ratings by psychologists less stable.

In a later article, Taft (1955) reviews much of the research and classifies it under five different headings: (1) the perception of emotional expression in photographs, drawings, models, and moving pictures; (2) rating and ranking traits; (3) personality descriptions; (4) personality matchings; and (5) prediction of behavior and life history data. He found that in making estimates of how others respond to a questionnaire most people tend to "assume similarity" (Wolf & Murray, 1937)--that is, they attribute to others the same response they themselves would make. Following this to its logical conclusion, it would mean that the judge whose own responses are most similar to the judging criterion will be most accurate in judgments based on that criterion. Bruner and Tagiuri (in Lindzey, 1954) came to this same conclusion, reporting positive significant relationships between the judge's ability to predict and his similarity to the group being judged.

Bronfenbrenner, Harding, and Gallwey (in McClelland, Baldwin, Bronfenbrenner, & Strodbeck, 1958) reviewed Taft's article and suggested that the often-cited high positive relationship between accuracy in judging others and self
insight (or accuracy in judging one's self) is based on an artifact. They describe the possible spurious nature of Taft's results saying the judges often tend to rate themselves high on admirable traits and low on reprehensible ones. As a result, those who are actually high on admirable or low on reprehensible traits will tend to be scored higher than others on self insight. Furthermore, since most people tend to expect favorable evaluations, those who actually receive them will turn out to be right and deemed more insightful.

Several investigators have attempted to use independent measures of insight against which to compare degree of accuracy in rating others. Vernon (1933) was one of the earliest investigators of this and found a correlation of .39 between insight independently defined and the consensual accuracy of self rating. However, between insight and the ability to rate others he found no correlation. Dymond (1948, 1949) found that the ability to "empathize" or take the role of the other tends to be related to the person's insight as determined in a clinical interview by whether he shows an understanding of his relation to others as these are revealed by his own TAT stories.

Cottrell and Dymond (1949) studying empathic responses found results similar to Adams (1927) and Vernon (1933). High empathy scorers were described as emotionally expressive, outgoing, warm people who had a strong interest
in others and were capable of establishing rewarding affectional relations with them. Low scorers, on the other hand, were usually rigid, introverted people whose emotional life appeared inhibited, and who were subject at times to poorly controlled outbursts of emotional behavior. In general, they tended to mistrust people, isolate themselves, and were not as well integrated in life situations as those with high empathy scores. Through their efforts, it is possible to argue that the ability to "empathize" or "take the role of the other" is the underlying process in a diverse range of skills in the area of social behavior and development (cf., Feffer & Suchotliff, 1966; Flavell, 1966; Flavell et al., 1968; Weinstein, 1969). Dymond (1950) suggested that the analysis of empathic ability will lead to a better understanding of such phenomena as the more efficient prediction of individual behavior and the reasons groups become or fail to become integrated. She further contends that empathic ability is a prerequisite for success in "helping" occupations such as clinical work, psychiatric work, social work, and so forth.

Others have had corresponding ideas, some antecedent (Kerr & Speroff, 1947) and others subsequent (Bender & Hastorff, 1950, 1953; Sargent, 1953) to Dymond's investigations. Bender and Hastorff (1950, 1953) indicate that what may appear like accuracy ("social sensitivity") in being able to estimate others' attitudes and feelings may be a function
either of a combination of projection by the judge aided by similarity between judge and judged and/or of something producing empathy. They distinguish "raw empathy" scores from "refined empathy" scores—the former being a straightforward accuracy score and the latter being raw accuracy corrected for the contribution of the judge's projection of his own attitudes. Their data shows a generalized tendency for some individuals to project consistently and for others to display empathic ability. They contend that studies of judgments of others must take this projection factor into account in drawing conclusions about accuracy. They concluded that judge-subject similarity is uncorrelated with refined empathy, while highly correlated with raw empathy scores. This conclusion lends support to Bronfenbrenner et al.'s (1958) postulated artifact discussed above, and at the same time presents a method to deal with it.

In many of the studies cited, the criteria employed for the accuracy of judging others was often agreement with others regarding a person's characteristics. Use of this criteria can present problems because results can easily be confounded by such biases in judgment as the halo effect, logical error, and the leniency effect.

Thorndike (1920) defined the halo effect as the tendency to rate subjects on several traits in terms of a general impression of goodness or badness with the result of a spuriously high correlation in ratings. The halo effect
seems to have its most profound effect when the traits to be judged are unclear in behavioral expression, when they are infrequently used by the judge, or when they have moral implications. Furthermore, according to Symonds (1925, 1931) the halo effect seems to increase with prolonged acquaintance. Logical error is a related tendency and can be defined as the tendency to conceive that certain traits go with certain other traits. Accordingly, if a judge rates a person high in aggressiveness, he will more likely rate him high rather than low in energy (Asch, 1946).

The leniency effect consists of the tendency to rate others and oneself high on favorable traits and low on unfavorable ones (cf. Bronfenbrenner et al., in McClelland et al., 1958), and seems to be little more than a special instance of central tendency of judgments. The result is that, lacking full information, a judge often operates on the assumption that people are moderately good and rates accordingly.

All of these biases of judgment can affect ratings of others and unless they are controlled in the experimental design, the studies will be confounded. Effective countermeasures have been undertaken by some of the early investigators (Asch, 1946; Lemann & Solomon, 1952; Symonds, 1925, 1931). However, for the most part, only tentative conclusions can be drawn for the early studies on the ability to understand and judge others. Some of the most reasonable
According to Bruner and Taguiri (in Lindzey, 1954) seem to be the following:

- accuracy in judgment is aided by similarity between judge and judged. This may be a function of resonance between judge and judged, better acquaintance with people similar to oneself, or to some kind of projection which happens to be more accurate when the other person is like oneself.

- accuracy depends on having cues to work with. Traits with little behavioral manifestation are harder to judge and are usually poorly judged.

- judgmental biases (e.g., the halo effect) contribute to much of the error of judgment.

- empathic ability seems related to judging ability and may even be the critical capacity involved.

- finally, relationships between personality variables and judging ability were found. For example, it was found that social adjustment and intelligence often improve judgment.

This relationship between intelligence and accuracy in judging others is of intrinsic interest to our study. Although the evidence is somewhat ambiguous, on the whole it seems to point to a slight positive relationship. Positive correlations were found by Allport and Allport (1921), Adams (1927), Vernon (1933), Bymond (1949), Wedeck (1947), and Taft (1950). Taft divided his conclusions into two
areas: (1) the ability to judge others analytically; and (2) the ability to judge others nonanalytically. The distinction between analytic and nonanalytic judgments appears to be a particularly important one. In analytic judgments, the judge is required to conceptualize, and often to quantify, specific characteristics of the subject in terms of a given frame of reference. This mainly involves the process of inference. In non-analytic judgments, the judge responds in a global fashion, as in matching persons with personality descriptions, and in making predictions of behavior. An empathic process is usually involved in nonanalytic judgments. He reported a .55 as his highest correlation (the use of intellectually homogeneous groups in most of his studies contributes to the reduced correlations obtained), and concluded that intelligence and the ability to judge others analytically were positively correlated. He contended that these results should not be surprising as such modes require a precise understanding of the meaning and application of abstract terms.

However, other workers failed to find positive results (Bender, 1935; Estes, 1937; Travers, 1941, 1943; & Taft, 1950, examining nonanalytic judgments). Looking at this second area of Taft seems to give us a clue in that non-analytic modes of judging tend to manifest lower correlations between intelligence and accuracy of judgment (Kelly & Fiske, 1951; Lindgren & Robinson, 1953; & Travers, 1943). Taft
concludes that accurate nonanalytic judgments of others are more a function of good perceptual and judgmental attitudes than of the use of abstract intelligence. However, if the mode of making such judgments is not within the level of the judge's comprehension, then abstract intelligence may be an aid in good judging ability. In general, it appears that very low intelligence may hinder the accuracy of judgments especially if perceptual and judgmental attitudes are also low. High intelligence, on the other hand, may help but is no guarantee of good performance.

Detachment is another quality which seems related to the ability to judge others accurately. Some studies (Adams, 1927; Taft, 1950; Vernon, 1933) found good judges to be less social and extroverted, and Estes (1937) discovered that among his judges, those who became emotional in the process of making judgments did least well. It appears, as one might expect, that good judges seem to have an abundant capacity for cool-headed evaluation of others.

Whether the ability to understand people is a separate attribute or one mediated by general intelligence, it seems to increase with age (Gates, 1923; Shanley, Walker, & Foley, 1971) and probably as a result of experience. Shanley and his colleagues studied 300 students from grades 6, 9, and 12, administering an Otis IQ test and the SFTSI, and found that social intelligence increased with age. However, no increase with age has been found in adults on various tests of ability.
to judge others (Dymond, 1952; Estes, 1937), suggesting that the ability to understand and/or judge may be a developmental phenomena that levels off at a certain age.

Taft (1955) examining sex differences concluded that the weight of evidence is in favor of no sex differences in the ability to understand others, even though there is a tendency to report slightly superior ability in women. Valentine (1929) studied the intuitive judgment of men and women in four experiments in which the task was to judge the character qualities of children and youth. He found that women, in general, did not turn out to be superior in making accurate judgments, though it was suggested that in ordinary everyday life women pay more attention than men do to signs of character and disposition in those whom they meet, and that this may help support the view that women's intuitions are superior. He also found that women often form intuitive judgments more quickly and that possibly they may tend to rely upon them more and retain them longer, even though they are in fact erroneous (Westcott, 1968).

Examining the common factors in 52 mental tests, Woodrow (1939) used Thurstone's centroid method of intercorrelation. Included were social intelligence tests, Phillip's attention test, and Seashore's test of musical ability along with various measures of general intelligence. After rotation of axes, the most important factors were verbal facility, spatial ability, numerical ability, attention, musical ability, and memory. Social intelligence tests proved
to be mainly tests of verbal factor. A factorial sex difference indicated the superiority of men in the spatial ability tests. No other sex differences were significant.

Thorndike and Stein (1937) factor analyzed the George Washington Social Intelligence Test (GWSIT) and concluded that although the GWSIT may tap slightly some unique field of ability, it measures primarily the ability to understand and work with words—an ability which contributes so heavily in an abstract intelligence test. He concluded that it was doubtful whether any test which is predominantly verbal could measure social intelligence or the ability to understand others.

Bottril (1967) used the GWSIT to study the social intelligence of social psychology students at the Universities of Guelph in Manitoba and St. John's College in Winnipeg. He compared the GWSIT with both the WAIS verbal and the WAIS performance and reported correlations of .61 (p = .01) and .13 respectively (almost significant at .05). Both social psychology students and females in general tended to score higher than others on SI, although when all students were measured with the WAIS no difference was found for any of the groups. Bottril concluded that either the GWSIT was an imperfect instrument for measuring SI or that SI is not a pure factor.

Rothenberg (1970) studied children's social sensitivity and its relationship to interpersonal competence,
intrapersonal comfort and intellectual level. He defined social sensitivity as "the ability to accurately perceive and comprehend the behavior, feelings, and motives of other individuals" (p. 337). Third and fifth graders were asked to describe the feelings and motives of characters portrayed in tape recorded stories. His findings suggest that age, intellectual ability and interpersonal adjustment contributed most to the development of accurate social perception and that no significant effects on social sensitivity were due to sex ordinal position or size of the family. These findings lend further support to Taft's (1955) conclusion cited above.

Wedek (1947), using an individual difference approach, hypothesized that "psychological ability" (SI) was different from g and verbal ability. Following the lead of Gilliland and Burke (1926) and Moss et al. (1927) he developed eight psychological ability tests which used either pictorial or auditory stimuli. A factor analysis of these produced the three factors of g for general intelligence, v for verbal facility, and x for psychological ability. His success in demonstrating social-intelligence factors with tests using nonverbal stimuli provided the basis for O'Sullivan and Guildord's SFTSI (1965, 1966).

Hoepfner and O'Sullivan (1968) using the SFTSI compared SI and IQ, and obtained correlations ranging from .17 to .42 between the six tests and the Henmon-Nelson Tests of Mental Ability. Many of the SI tests consist entirely of
cartoons, photographs, and other non-semantic stimuli. This may explain the success that O'Sullivan et al. (1965) experienced in factor analyzing and separating the social intelligence factors from one another and from factors in abstract-verbal intelligence. On the basis of their results it appears that the high IQ examinees either have a high level of SI or are able to compensate and solve many of the behavioral-cognition problems in the SI tests by utilizing verbal skills. The authors recognize the potential weakness of the test in that verbally gifted people may excel on the SI test simply because of their highly developed and pervasive verbal skills. However, they feel that the chief value of the tests may lie in their capacity to identify those of relatively low IQ who are gifted in social perception, thereby lending support for two separate, though possibly overlapping factors. It is to this test that attention will now be turned.

**Tests of Social Intelligence**

**Six Factor Tests of Social Intelligence** (O'Sullivan, Guilford, & deMille, 1965; O'Sullivan & Guilford, 1966)

This test is derived from Guilford's (1967) Structure of Intellect Model (SOI) in which he postulates 30 distinct social intelligence or behavioral abilities. These are only a fraction of the 126 theoretical intellectual aptitudes that Guildord hypothesizes in this SOI model: the three dimensions
of the SI model specify the content, the operation, and the product of any given intellectual act. Content is divided into semantic, symbolic, figural, and behavioral; operations include cognition, memory, divergent production (generation of variety of output), convergent production (generation of the one correct solution), and evaluation (judging in terms of criteria); the product dimension which includes the results of intellectual processing is comprised of units, classes, relations, systems, transformations, and implications (cf. O'Sullivan, Guilford, & deMille, 1965).

The social intelligence tests are all within the behavioral content area, and are focused on the six factors of behavioral cognition: cognition of behavioral units, classes, relations, systems, transformations, and implications. O'Sullivan et al. (1965) define social intelligence as "behavioral cognition (or) the ability to understand the thoughts, feelings, and intentions of other people . . . especially as this is manifested in discernible, expressional cues" (p. 6). Graphic and photographic stimuli are almost used exclusively, both in the presentation of information and in the response choice. This sparing use of verbal statements and responses reduces the semantic variance which has been a problem in so many other instruments (e.g., GWSIT, Kerr & Speroff Empathy Test).

The comprehension of other people as defined by Guilford and his colleagues does not include comprehension
of the generalized other. Kerr and Speroff's (1947) empathy test is concerned with the latter, but Brofenbrenner and his colleagues (in McClelland et al., 1958) have clearly portrayed the difference between this kind of social sensitivity and that involved in knowing the feelings of a given individual.

Furthermore, the SFTSI deals only with behavioral cognition and not behavioral operation. In other words, it purports only to answer the first part of Thorndike's definition to understand other people and does not delve into the area of assessing wise action. Accordingly, Hendricks, Guilford, and Hoepfner (1969) constructed a test which attempted to measure the six divergent or creative behavioral abilities. So far this seems to be the closest effort to ascertaining wise action. However, in light of the inconclusive and contradictory research on past instruments, the first order of business seems to be determining the most adequate instrument for assessing the first part of Thorndike's definition--understanding others. With this in mind, brief descriptions of the SFTSI follow.

Cognition of Behavior in Terms of Implications

**Cartoon Predictions.**--In this test the task is to choose one of the three alternative cartoons that shows what is most likely to follow a given cartoon that depicts a certain interpersonal situation. The test was found to
measure the cognition of behavioral implications (CBI) with a factor loading of .55.

Cognition of Behavior in Terms of Classes (CBC)

**Expression Grouping.**—Each item of this test consists of a group of three drawings which depict facial expressions, hand gestures, and body postures in various collections. The subject has to show that he has cognized the class by selecting one of four alternative drawings of expressions. The ability involves abstracting common attributes of behavioral information using different expressive stimuli. A factor loading of .59 for cognition of behavioral classes (CBC) is reported.

Cognition of Behavior in Terms of Systems (CBS)

**Missing Cartoons.**—The task here is to choose one of four cartoon panels that best fills a blank in an otherwise complete sequence. Missing Cartoons is a story measure of CBS, with a loading of .52 but also reports factor loadings of .41 on CBU (cognition of behavioral units) and .35 on CBI (cognition of behavioral implications). CBS is the capability of sizing up situations involving interpersonal interaction and is a common social requirement for adequate understanding and potential reaction.

**Missing Pictures.**—This is similar to Missing Cartoons except that photographs are utilized instead of cartoons. In such items there are two possibilities for cognizing
behavioral systems. One is connected with interpretation of each scene, and the other with the story itself in an overall organization of the succession of situations or systems. A factor loading of .58 in CBS is reported.

Cognition of Behavior in Terms of Transformations (CBT)

**Picture Exchange.**--In this task an individual must choose one of three photographs which when substituted for a marked picture of a four picture story will change the story's meaning. The photographs include pictures of children, students, and mature adults. A factor loading of .51 on CBT is reported. CBT pertains to flexibility of interpretation in contrast to rigidity of such interpretation.

**Social Translations.**--This test is unique among the other behavioral-cognition tests in that it employs printed words only. The task is to choose the one of three stated alternative pairs of people between whom a given verbal statement will have a different behavioral meaning, much different from that if spoken between members of another given pair. A factor loading of .51 on CBT is reported, along with a small secondary loading on CBR (cognition of behavioral relations).

Although the research conducted on the SFTSI is relatively sparse, convincing reliability and construct validity estimates based on factor loadings are indicated. O'Sullivan and her colleagues (1966) in the introductory research derived
reliability estimates for each test from intercorrelations of separately timed halves and obtained favorable results (p. 5). Similarly, in an attempt to determine construct validity for the various tests, their factor analysis demonstrated that the behavioral tests measure abilities other than those actually measured by verbal-IQ tests and tests of other intellectual qualities.

More specifically, Spearman-Brown reliability estimates for the six tests ranged from .32 to .85. The two tests with the lowest reliabilities were Missing Pictures (CBS) and Picture Exchange (CBT). However, Missing Cartoons also measures CBS and Social Translations measures CBT and both have much higher reliabilities (.75 to .85 compared to .32 and .48). As a result, Missing Cartoons and Social Translations were combined with Cartoon Predictions and Expression Grouping to form the Social-cognition Composite. This is a composite of the four tests with the highest reliabilities which at the same time include measures of the four factors. This composite was used by Shanley and his colleagues (1971) and will also be used in the present study.

Hoepfner and O'Sullivan (1968) compared the SFTSII with the Henmon-Nelson Tests of Mental Ability using an adolescent population (mean chronological age of about 16.7 years and a mean IQ of 117.7). The level of the IQ-SI correlation resulted in a mean correlation of .34 and is similar to that reported by Hendricks, Guilford, and Hoepfner (1969) between IQ and various measures of creative potential (.32).
Yet despite the similarity of correlation with both creative potential and social intelligence measures, the patterns of their relationships were found to be very different (cf. Hoepfner & O'Sullivan, 1968, p. 343). The authors concluded that the high IQ examinees either: "(1) have a high level of SI, evidence for the general intellectual well-being of what is called the gifted individual or (2) are able to compensate and solve many of the behavioral problems in the SI tests by utilizing verbal skills" (p. 343). It was suggested that if the latter is the case, then their chief value may lie in their capacity to identify those of relatively low IQ who are gifted in social perception.

Tenopyr (1967) reported further evidence for construct validity and concluded that a substantial proportion of the variance of the SFTSI may be attributed to abilities other than those typically associated with intellectual achievement. Shanley, Walker, and Foley (1971), on the other hand, testing the hypothesis that social intelligence increases with age, obtained correlations between IQ and SI which not only were significant but sufficiently high enough to raise questions about the independence of these two types of intelligence. They also found that social intelligence appears to increase with age and that females tend toward superior performance on the SFTSI. Other studies have been done by Suran (1970) who found a slight relationship between SI and effective sensitivity group leadership; and Clark and
Neuringer (1971) who found no significant difference in SI between repressor-sensitizer personality styles.

The Chapin Social Insight Test (CHSIT)

In 1939, Chapin, a sociologist, developed a technique called the Social Participation Scale which he intended to measure the action-oriented phase of SI and to supplement measures which emphasized the understanding component of SI. In 1942, he presented a technique called the Social Insight Test which would tap the understanding component. This is a 25-item test consisting of verbally presented situations with four possible comments from which to choose.

Little research has been done on this instrument primarily because interest in social intelligence was waning at the time due to criticism of available techniques (Thorn-dike & Stern, 1937). Also, as Walker and Foley (1973) point out, Chapin published in sociological journals and psychologists who were the chief protagonists of SI were, for the most part, unaware of his work.

Gough (1965, 1968) has revived interest in the CHSIT and states the "aim of the test is to assess an individual's ability to appraise others, to sense what they feel and think, and to predict what they may say and do" (1968, p. I)—a definition quite similar to that given by Thorndike (1920) and O'Sullivan et al. (1965).

Gough (1968) correlated the CHSIT with eight standard tests of ability and aptitude and obtained correlations
ranging from .24 to .40 with a grand mean of .34. Though the SIT does relate to intellectual ability, the relationship is modest. In the same publication he reports significant relationships between scores on the CHSIT and the rankings of occupational groups and also between the CHSIT and ratings on such variables as leadership, ability to communicate, ability to evaluate ideas, and good judgment (positive correlations of .26, .31, .29, and .27 respectively).

Prior research by Gough (1965) examined 197 graduate students in psychology tested at entry into graduate training. A t test between difference in means of a dropout group vs. a group who either graduated or were continuing in the program was significant beyond the .01 level of probability, and the biserial correlation between the SIT and the dichotomy of dropout group vs. graduate or continue group was .40.

Reliability for the instrument ranges from .75 as obtained by Chapin (1942) in an odd-even reliability check of his original 45-item form, to .78 as determined in a more recent odd-even check on the present 25-item version of the test (Gough, 1968). Because test-retest coefficients have never been computed, no evidence on stability of scores over time is available at present.

Finally, Gough (1968) correlated the SIT and four inventories assessing personality and attitudes: the
California Psychological Inventory--CPI (Gough, 1964); the Minnesota Multiphasic Personality Inventory--MMPI (Hathaway & McKinley, 1943); the Barron-Welsh Art Scale (Barron & Welsh, 1952); and the Study of Values (Allport, Vernon, & Lindzey, 1951). Only the CPI and the MMPI are of intrinsic value to the present study as the third measure of SI used is the Hogan Empathy Scale (1969)--an instrument composed of items from both of these instruments. Correlations with the CPI are essentially zero except on scales for Achievement via Independence and Intellectual efficiency. Gough concludes that the high scorer on the SIT should as a result be to some extent a resourceful and effective individual with a special responsiveness to problems requiring originality and ingenuity for their solution. With the MMPI the highest median coefficient is that of .19 with the Mf (femininity) scale--a scale which ordinarily stresses intellectuality of interests, cultural breadth, and emotional sensitivity. In view of this, it seems likely that the correlation between the CHSIT and the Hogan Empathy Scale may strongly depend on the number of similar items and/or combination of items that appear on both the Hogan Empathy Scale and the above mentioned scales of the CPI and MMPI.

The Hogan Empathy Scale (HES)

Hogan (1969) defines empathy as "the intellectual or imaginative apprehension of another's condition or state of mind" (p. 307) and feels that it is central for understanding a broad range of social phenomena including, in particular,
moral development. Presently, Hogan's introductory article is the only published data on the Empathy Scale. Hogan asked psychologists and non-psychologists to contribute a 50-item Q-sort description of an empathic man, and obtained an intercorrelation corrected for attenuation of .93, suggesting a common conception of the behavioral connotations of empathy.

The criterion of empathy was then determined by having four faculty and research psychologists and three graduate students in psychology at the University of California, Berkeley, describe an empathic man using a full 100-item Q-sort. Next, 2 groups, 100 military officers and a second sample of 45 research scientists and 66 student engineers, were studied by skilled observers and rated on empathy according to the criterion. Other comparisons were made including correlating empathy ratings with the standard scales of the California Psychological Inventory (CPI; Gough, 1964), the MMPI (Hathaway & McKinley, 1943), and the Chapin Social Insight Test. In general, the correlations were positive for the CPI and the Chapin test which stress effective functioning, and negative for the MMPI which has the opposite orientation.

The empathy scale was developed by the standard technique of an item analysis of the responses of high-rated versus low-rated groups to 957 true-false items in the CPI, MMPI, and an IPAR (Institute of Personality Assessment and Research) pool of items. The chi-square or Fisher's exact
statistics were used to evaluate differences, and 64 items (32 scored true, 32 false) were selected for the final scale. Thirty-one are from the CPI, 25 from the MMPI, and the remaining eight items come from items used in IPAR research.

Items for the final scale were retained on the basis of four considerations. First, differences in endorsement frequencies between the high- and low-rated groups must be in the same direction in both samples. Second, 15 items which appeared in two earlier analyses (Gough, 1955; McClelland, 1951) that compared items of the MMPI and CPI against criteria of role-playing ability or the Social Acuity Index (Gough, 1955), also attained significance in the present research and were retained. Third, 17 of the items finally selected failed to attain statistical significance, but were retained on the basis of relevant content. Finally, items were chosen with an eye toward balancing the scale's true and false keying (Hogan, 1969, p. 310).

Hogan reports a number of validity studies and obtained favorable validity coefficients (ranging from .39 to .80). With a sample of 50 college undergraduates, the reliability of the empathy scale estimated by a test-retest correlation after a 2-month interval was .84. Applying the Kuder-Richardson-21 formula to the scores of 100 military officers, Hogan obtained a coefficient of .71.

The empathy scale was correlated with a number of various personality and intellectual measures and showed
several moderately large coefficients. In one study Hogan attempted to determine if the empathy scale reflected the ability to adopt the moral point of view as discussed within the framework of his multidimensional theory of moral development (cf. Hogan, 1967). The important conclusion of this study for our purposes was that, in comparing officers and inmates at a state prison, differences between the groups were highly significant \( p < .001 \), and although the officers were more intelligent than the prisoners, the obtained differences in empathy scores could not be explained entirely in terms of differences in intellectual endowment.

Hogan concluded that

\[ . . . \text{from peer ratings, test correlates, and other validational evidence . . . , high scorers seem to be socially acute and sensitive to nuances in interpersonal behavior . . . (and) . . . low scorers, on the other hand, seem hostile, cold, and insensitive to the feelings of others (p. 315).} \]

**Tests of Abstract Intelligence**

**The Quick Word Test (QWT)**

The Quick Word Test (Borgatta, 1960, 1964) is based on the general agreement of psychometricians that the best single indicator of mental ability is understanding of the meaning of words. Its primary purpose is to provide a substitute measure, or an estimate of performance, for longer and more time-consuming group tests of general ability.

Each form consists of 100 multiple-choice vocabulary items consisting of stem words of five letters followed by
four four-letter alternatives, one of which has the same meaning as the stem word. The items of the test are distributed in such a way that the difficulty level ranges from low to high within each group of five items. This results in all twenty blocks being approximately equal in median difficulty so that the examinee is not likely to be affected by progressive discouragement. Furthermore, it provides a convenient form for scoring since no matter how much an individual does, his score is the percentage of words gotten correctly. If a subject does only one-half of the test, his percentage can be obtained just as conveniently as if he did the entire test.

Reliability measures cited in the manual (Borgatta & Corsini, 1964) range from .90-.93 for split-half techniques, and .88-.89 for alternate forms. Compared to other measures of general intelligence, correlations range from .62 with the Otis Quick-Scoring Test of Mental Ability, Gamma to .84 for the Lorge-Thorndike Intelligence Tests: Level 5. Borgatta and Corsini (1960a) reported a correlation of .78 between the QWT and the Kuhlmann-Anderson (Form not cited). Also, in a study conducted in the same year, Borgatta (1960b) obtained correlations ranging from .75 to .83 between 4 forms of the QWT and the WAIS.

Grotelueschen and a number of colleagues (Grotelueschen & Knox, 1967a; Grotelueschen & Lyons, 1967b; Grotelueschen & McQuarrir, 1970) conducted several studies
comparing the QWT with other instruments and obtained a reliability index of .94 and correlations of .77 and .74 respectively with verbal and total WAIS IQ scores.

Other researchers (Bohrnstedt, Lambert, & Borgatta, 1971; Reubush, 1968) compared the QWT with high school students and concluded that the QWT permits a rapid and reliable estimate of pupils' academic progress for the immediate future.

The Otis Quick-Scoring Mental Ability Tests: Gamma, Form FM

Odd-even reliability coefficients corrected by the Spearman-Brown Formula ranged from .85 to .92 for samples reported in the manual. Investigations between the Otis and Stanford-Binet and the Otis and WISC were conducted (Estes, 1965; Kandel, 1966; Keach, 1966) and results indicated that the Otis often underestimated Stanford-Binet and WISC IQs and should not be relied on for selection of children of superior intelligence. However, these studies employed a different form of the Otis than the Gamma FM, and therefore may not be valid for the Gamma form. Fox (1966) found that Otis IQs were a significant predictor of graduate education grades. And Hayden (1969) and Ahmad (1968) studying students in the Philippines and West Pakistan respectively, demonstrated that the Otis was useful as a successful predictor of academic success and may be helpful in the selection of applicants for higher education.
Welty (in Buros, 1959), reviewing the Otis finds fault with the manual's criterion of judging intelligence by school progress. He also criticizes the use of the term validity as used in the manual and the method of standardization of the new test on the following points: (1) there is no clear statement of the nature of the normative population or of the normative sample; (2) there seems to be a response bias in the Gamma test so that the hypothesis of equal assignment of correct answers to the five alternative responses has to be rejected. In spite of some of the negative criticism, the Otis Gamma was picked for this study because of its extensive use, its availability, and as a further check on its convergent validity with the other measures of intelligence.

The Kuhlmann-Anderson Test, 7th Edition, Booklet H

Pigeon (in Buros, 1965) comments that the reliability of the 7th edition is generally satisfactory with test-retest coefficients ranging from .83 to .92 with as much as two grades between testing. Testing with adjacent forms produced correlations from .77 to .89. With the new edition the number of tests in each booklet have been reduced from 10 to 8—4 verbal tests and 4 quantitative tests. These are timed as eight separate tests. The H booklet is designed for use in the 9-12 grade and above. It contains 16 items most of which are adopted from the corresponding tests of the 7th edition.
edition. The number of items in each test has been increased, resulting in larger standard deviations and hence better discrimination among high and low scorers. About 55 per cent of the content is new for the 7th edition. In his review, Pigeon cites two studies, one comparing the K-A with the Stanford Binet for 9th graders and one comparing WAIS IQs and the K-A with a group of 12th graders. Close agreement between the mean and S.D. for IQs of the S-B and K-A were found, but such was not the case for the K-A and WAIS.

North (1961), studying the results of the 7th edition K-A test for Independent School Pupils in grades 7-12, obtained reliabilities as follows: verbal, .85-.90; quantitative, .87-.93; and total scores and IQs, .91-.96.

Norms and test development are discussed in the Technical Manual (1962), along with evidence for the construct and concurrent validity of the 7th edition. Correlations between Booklet H and other tests are cited for grades 11 and 12 and range from .64 to .83, indicating that similar traits are being measured by these tests. In addition, the Otis Gamma correlated .89 with Booklet 6 for a 10th grade sample and .80 with Booklet H after a two year interval. In another sample cited, the results for a number of 11th and 12th grade students on the WAIS and K-A, booklet H, were compared. The mean IQ for the K-A H-booklet (131.52) reflected a higher ceiling than did the means of the WAIS verbal (127.6) and full-scale IQs (125.6). Also, there was
a wider range and greater variability of K-A IQs as compared with those of the WAIS (K-A range 110-160, S.D. = 11.6; WAIS range 108-141, S.D. = 7.1).

Yamamoto (1965), in an interesting study using the K-A 7th edition test, found that instructional sets affected the resulting intelligence scores. Subjects were from the 4th, 7th, 10th, and 12th grades and were randomly assigned to groups where the test was introduced either as a test of intelligence, a test of achievement, or a routine test (groups 1, 2, and 3 respectively). The results indicated that the mean IQ of the first group (116.3) was significantly higher than group two (112.1) or the routine group (109.1), while there was no significant difference between the latter two (achievement vs. routine). This may indicate that different levels of motivation and/or ego-involvement are aroused by a test purported to assess one's intelligence.

Hypotheses

In light of the preceding review of the literature, the following hypotheses will be investigated:

1. convergent validity among tests of abstract intelligence will be demonstrated.

2. convergent validity among tests of social intelligence will be demonstrated.

3. divergent validity between tests of abstract intelligence and tests of social intelligence will be demonstrated.
4. no difference between male and female will be obtained on measures of social intelligence.

In addition, further investigation of the relationship between social intelligence and general intelligence will be conducted by a factor analysis using a varimax rotation.
CHAPTER III

METHOD

Subjects

The Ss for this study consisted of 31 males and 28 female undergraduate students enrolled at Loyola University of Chicago. All Ss were volunteers from the introductory psychology classes who were fulfilling part of their required hours of experimental participation.

The Ss ranged in age from 18.2 years to 25.3 with a mean age of 18.7 years. For the males, the age range was 18.2 to 24.6 with a mean of 18.7; for the females, the range was 18.3 to 25.3 also with a mean of 18.7 years.

Racially, the sample consisted of 54 Caucasians, 3 Negroes, and 2 others. No scale of socio-economic level was used. However, it seems reasonable to assume that the majority were of middle class socio-economic status.

Materials

All Ss were given three measures of abstract intelligence and three measures of "social intelligence." All were tests that can be administered in groups.

The three tests of abstract intelligence were the Quick Word Test; the Otis Quick-scoring Mental Ability Test,
form Gamma; and, the Kuhlmann-Anderson Test, 7th edition, booklet H.

The measures of social intelligence included the Guilford and O'Sullivan Six Factor Tests of Social Intelligence, the Hogan Empathy Scale, and the Chapin Social Insight Test.

All tests were administered and scored in accordance with the specifications provided in their respective manuals. Total time of testing for each subject was approximately 3 hours, usually ranging from 2-1/2 to 3 hours.

**Procedure**

Each subject was administered all six measures, with measures of AI and SI alternated. Since three measures were timed (OTIS, SFTSI, and K-A), these were always presented first, though not always in the same order. The remaining three measures were also counterbalanced, so that the following four arrangements of test presentation were used:

- A. OTIS, SFTSI, K-A, HES, QWT, CHSIT
- B. OTIS, SFTSI, K-A, CHSIT, QWT, HES
- C. K-A, SFTSI, OTIS, HES, QWT, CHSIT
- D. K-A, SFTSI, OTIS, CHSIT, QWT, HES

Four testing sessions were conducted with a different test arrangement employed in each session. Seventeen Ss (6 male, 11 female) were tested in session 1 with arrangement A; 20 Ss (11 male, 9 female) were tested in session 2 with arrangement B; 10 Ss (6 male, 4 female) were tested in
session 3 with arrangement C; and 12 Ss (8 male, 4 female) were tested in session 4 with arrangement D.

All measures were correlated with each other and the correlations factor analyzed using a varimax rotation.
CHAPTER IV

RESULTS

The means and standard deviations obtained by the males, females, and total sample on the six variables used in this study are presented in Table 1 as a reference point for the following statistical analyses. The mean age for the male and female samples did not differ significantly (both 18.7). Table 2 presents t tests between the male and female samples on all these measures and a significant result ($p < .05$) was found only on the CHSIT, where females scored significantly higher than the males.

Relationship of Social Intelligence Measures to Abstract Intelligence Measures

A major criterion for the construct validity of any measure of social intelligence is that such a measure should not correlate strongly with measures of abstract or verbal intelligence. Table 2 contains a correlation matrix of the total sample ($N = 59$) in which Pearson Product-moment correlations between AI measures and SI measures are presented. Included in this matrix is the variable of sex, where a biserial correlation was used with positive correlations indicating that females scored higher, and negative correla-
<table>
<thead>
<tr>
<th>Variable</th>
<th>Males (N=31)</th>
<th>Females (N=28)</th>
<th>Total (N=59)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chapin Social Insight Test</td>
<td>18.61</td>
<td>21.07</td>
<td>19.78</td>
</tr>
<tr>
<td>Hogan Empathy Scale</td>
<td>37.84</td>
<td>38.50</td>
<td>38.15</td>
</tr>
<tr>
<td>Six Factor Test of Social Intelligence</td>
<td>78.61</td>
<td>80.25</td>
<td>79.39</td>
</tr>
<tr>
<td>Otis Gamma</td>
<td>60.58</td>
<td>58.43</td>
<td>59.56</td>
</tr>
<tr>
<td>Kuhlmann-Anderson</td>
<td>102.61</td>
<td>99.54</td>
<td>101.15</td>
</tr>
<tr>
<td>Quick Word Test</td>
<td>45.84</td>
<td>48.57</td>
<td>47.14</td>
</tr>
</tbody>
</table>
Table 2
Comparison of Males and Females on All Measures

<table>
<thead>
<tr>
<th>Variable</th>
<th>Male</th>
<th>Female</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>CHSIT</td>
<td>18.61</td>
<td>4.74</td>
<td>21.07</td>
</tr>
<tr>
<td>HES</td>
<td>37.84</td>
<td>4.80</td>
<td>38.50</td>
</tr>
<tr>
<td>SFTSI</td>
<td>78.61</td>
<td>12.64</td>
<td>80.25</td>
</tr>
<tr>
<td>OTIS</td>
<td>60.58</td>
<td>12.10</td>
<td>58.43</td>
</tr>
<tr>
<td>K-A</td>
<td>102.61</td>
<td>25.35</td>
<td>99.54</td>
</tr>
<tr>
<td>QWT</td>
<td>45.84</td>
<td>11.83</td>
<td>48.57</td>
</tr>
</tbody>
</table>

Note: For males, N=31; for females, N=28.

*p < .05
tions indicating males scored higher. The highest correlation between any measure and sex is .27 for the CHSIT, and though this only accounts for 7 per cent of the variance, it is the only measure where a significant difference was found between males and females (see Table 2). Tables 3 and 4 present correlation matrices for the male and female samples. As hypothesized the AI measures correlated in the high positive direction among themselves. However, such was not the case for the SI measures, with the exception of correlations obtained between the HES and the SFTSI (.43, .48, and .46 for the female, male, and total samples respectively). Correlations among other measures of SI ranged between .06 and .24.

Not only did the tests of SI not correlate highly among themselves, but contrary to expectations they tended to correlate higher with the tests of AI. There was one exception to this general tendency, and that was the moderately high correlation between the HES and the SFTSI (.43 to .48 across samples).

Furthermore, there were distinct differences between the intercorrelations of AI and SI in the male and female samples. For example, the female sample tended to have lower intercorrelations between the HES and the CHSIT with the three measures of AI (five correlations between .14 and .17 with one equalling .55), while the male sample, on the other hand, had much higher intercorrelations with five scores ranging
Table 3

Initial Correlation Matrix

<table>
<thead>
<tr>
<th>Variable</th>
<th>CHS</th>
<th>HES</th>
<th>SFTSI</th>
<th>OTIS</th>
<th>K-A</th>
<th>QWT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Sex*</td>
<td>.27</td>
<td>.07</td>
<td>.06</td>
<td>-.10</td>
<td>-.06</td>
<td>.10</td>
</tr>
<tr>
<td>2. CHSIT</td>
<td>.18</td>
<td>.21</td>
<td>.27</td>
<td>.26</td>
<td>37</td>
<td></td>
</tr>
<tr>
<td>3. HES</td>
<td>.46</td>
<td>.29</td>
<td>.32</td>
<td>.27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. SFTSI</td>
<td>.68</td>
<td>78</td>
<td>.63</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. OTIS</td>
<td>.85</td>
<td>55</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. K-A</td>
<td>.60</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*For sex a biserial correlation was used with positive correlations indicating females scored higher and negative correlations indicating males scored higher. The remaining correlations are Pearson product-moment correlations.

Note: df = 57; \( r \geq .256, p = .05; r \geq .333, p = .01; r \geq .418, p = .001 \)
Table 4
Correlation Matrix for Males (N=31)

<table>
<thead>
<tr>
<th>Variable</th>
<th>HES</th>
<th>SFTSI</th>
<th>OTIS</th>
<th>K-A</th>
<th>QWT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. CHSIT</td>
<td>.24</td>
<td>.24</td>
<td>.41</td>
<td>.40</td>
<td>.18</td>
</tr>
<tr>
<td>2. HES</td>
<td>.48</td>
<td>.40</td>
<td>.46</td>
<td>.41</td>
<td></td>
</tr>
<tr>
<td>3. SFTSI</td>
<td></td>
<td>.64</td>
<td>.80</td>
<td></td>
<td>.68</td>
</tr>
<tr>
<td>4. OTIS</td>
<td></td>
<td></td>
<td>.86</td>
<td></td>
<td>.66</td>
</tr>
<tr>
<td>5. K-A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.71</td>
</tr>
</tbody>
</table>

Note: df = 29; $r > .355; p = .05; r > .456, p = .01;$  
$r > .562, p = .001$
Table 5
Correlation Matrix for Females (N=28)

<table>
<thead>
<tr>
<th>Variable</th>
<th>HES</th>
<th>SFTSI</th>
<th>OTIS</th>
<th>K-A</th>
<th>QWT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. CHSIT</td>
<td>.06</td>
<td>.16</td>
<td>.16</td>
<td>.14</td>
<td>.55</td>
</tr>
<tr>
<td>2. HES</td>
<td>.43</td>
<td>.17</td>
<td>.16</td>
<td>.14</td>
<td></td>
</tr>
<tr>
<td>3. SFTSI</td>
<td></td>
<td>.77</td>
<td>.76</td>
<td>.59</td>
<td></td>
</tr>
<tr>
<td>4. OTIS</td>
<td></td>
<td></td>
<td>.84</td>
<td>.50</td>
<td></td>
</tr>
<tr>
<td>5. K-A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.54</td>
</tr>
<tr>
<td>6. QWT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: df = 26; \( r > .374, p = .05 \); \( r > .479, p = .01 \);
\( r > .588, p = .001 \)
between .40 and .46, with one equalling .18. In both male and female samples correlations between the SFTSI and all AI measures were significantly high (correlations ranging from .64 to .80) suggesting that the SFTSI is measuring the same ability as the three measures of abstract or verbal intelligence.

**Factor Analysis**

To analyze the data more thoroughly a factor analysis employing a varimax rotation (Nfactors = 2) was conducted on the total sample. Table 6 presents the results of that factor analysis, with .30 being the conventional level of significance. Factor 1 had high loadings on all measures of AI and on the SFTSI. The HES also loaded above .30 (.38) on Factor 1. Factor 2, on the other hand, seems to be loading on a sex variable more than anything else as the females tended to obtain higher scores on both the CHSIT (see Table 2) and the QWT.

The first factor, factor 1, seems to be a verbal factor, and in light of the high loadings and the obtained eigenvalue of 3.07 is clearly interpretable. Factor 2 does not obtain the conventional 1.0 eigenvalue cutoff (.63) and therefore is not interpretable other than as pointing out some sex differences.
Table 6

Varimax Rotated Factor Matrix

<table>
<thead>
<tr>
<th>Factors</th>
<th>Variable</th>
<th>Factor 1</th>
<th>Factor 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1. Sex</td>
<td>-.06</td>
<td>.58</td>
</tr>
<tr>
<td></td>
<td>2. CHSIT</td>
<td>.28</td>
<td>.47</td>
</tr>
<tr>
<td></td>
<td>3. HES</td>
<td>.38</td>
<td>.17</td>
</tr>
<tr>
<td></td>
<td>4. SFTSI</td>
<td>.83</td>
<td>.15</td>
</tr>
<tr>
<td></td>
<td>5. OTIS</td>
<td>.87</td>
<td>-.05</td>
</tr>
<tr>
<td></td>
<td>6. K-A</td>
<td>.94</td>
<td>-.03</td>
</tr>
<tr>
<td></td>
<td>7. QWT</td>
<td>.67</td>
<td>.30</td>
</tr>
</tbody>
</table>
CHAPTER V

DISCUSSION

Relationship of Social Intelligence Measures to Abstract Intelligence Measures

As was hypothesized (hypothesis 1) measures of abstract intelligence correlated highly among themselves with scores ranging from .50 to .86.

However, intercorrelations between measures of social intelligence did not fare very well, as hypothesis 2 predicted they would. This may be due to several reasons not the least of which is the differences between the stated purpose of each test. For example, Gough (1968) states that

the test (CHSIT) is not intended as a measure of empathy --the degree to which one person identifies with another or feels in sympathy with him--nor is it intended to reflect emotional responsiveness, tolerance, or other attributes and dispositions which might be called to mind by the phrase "social insight." These factors might prove to be correlated with scores on the Social Insight Test, but they do not constitute its principal focus. The aim of the test, to repeat, is to assess an individual's ability to appraise others, to sense what they feel and think, and to predict what they may say and do (p. 1).

Hogan (1969), on the other hand, in developing his empathy scale sees empathy "as an everyday manifestation of the disposition to adopt a broad moral perspective . . . (and
defines it) . . . as the intellectual or imaginative apprehension of another's condition or state of mind without actually experiencing that person's feelings" (p. 308). He makes the comment that, although quite interesting, the Chapin test seems more closely tied to the realm of social and interpersonal effectiveness per se, although he feels that some overlap may exist between the HES and the CHSIT. In fact, in his validational research, he correlated empathy ratings with the Chapin test and discovered that correlations were positive but low, suggesting that his empathy ratings define a dimension which includes aspects of social competence. The present study lends support to his findings, also obtaining positive but low intercorrelations between the HES and the CHSIT (.06 for females, .24 for males, and .18 for the total sample).

Hogan also comments on the relationship of his scale to abstract intelligence. He comments that, if, as Mead (1934) suggested, empathic ability underlies social intelligence, then the empathy scale should relate to both measures of social effectiveness and functional intelligence. In support of this he found that the empathy scale achieved its highest correlations with scales of the CPI which measure social and interpersonal adequacy (.34 to .62) and with the Intellectual Efficiency Scale (.52) which is an index of the degree to which a person effectively mobilizes his intellectual resources. Hogan concluded that the relationship
between the empathy scale and intelligence scores is somewhat ambiguous with coefficients ranging from low negative to moderately positive depending on the population and the measure of intelligence used. The present study obtained low to moderately positive correlations between the HES and the three measures of AI with males obtaining much higher correlations than females.

Gough (1968) also validated the CHSIT against measures of AI. He thought that one should expect a measure of social insight to have significant relationships to these other tests because of the common cognitive component, but expressed the "hope that the correlations would not be so high as to suggest that the CHSIT is just a variant of the measures of intellectual ability" (p. 6). He compared the CHSIT and eight other tests, using various samples of male subjects, and obtained correlations ranging from .24 to .40 with a grand mean of .34. He concluded that a relationship between the CHSIT and intellectual ability exists but felt that it was nonetheless modest. As with the HES, the CHSIT correlated positively with the three measures of AI. In the female sample the CHSIT correlated .14, .16, and .55 with the K-A, OTIS, and QWT respectively. The male sample tended to reverse this pattern with the CHSIT correlating .40 with the K-A, .41 with the OTIS, and .18 with the QWT. Not only is a sex difference evident from the reverse pattern of correlations, but as mentioned above, the CHSIT was the only measure in
which a significant t test was obtained at the .05 level partially disconfirming hypothesis 4 which stated that no sex differences between measures of SI would be found.

The third measure of SI, the SFTSI, seems to be in a category by itself. It correlated modestly with the CHSIT (.21 for the total sample) and much higher with the HES (.46 for the total sample), though not high enough to lend support for the hypothesis of convergence between the measures of SI. In the development of the SFTSI, O'Sullivan et al. (1965) defined SI as the ability to understand the thoughts, feelings, and intentions (psychological dispositions) of others. This comprehension of other people does not include comprehension of the generalized-other as some empathy tests are concerned with (Kerr & Speroff, 1947), but rather the kind of social sensitivity that is involved in knowing the feelings of a given individual. This definition has definite similarities to the definition of empathy and social insight mentioned above. Yet intercorrelations between these measures do not reach the level which might be expected from the similarities of their definitions and realm of behavior investigated.

Looking at the other side of the validation process—divergence from other traits, the SFTSI again falls short. In the original validation research O'Sullivan et al. (1965) used several marker tests to distinguish the hypothesized behavioral-cognition dimensions from intellectual factors. One such test was the Verbal Comprehension test in which an
individual is directed to choose the one of five alternative words that has the same meaning as a given word. This loaded .71 on the CMU factor (cognition of semantic units) and is very similar to the Quick Word Test (QWT) used in the present study. O'Sullivan and her colleagues state that Verbal Comprehension is widely regarded as the major component of the traditional concept of general intelligence, and concluded that because it loaded higher than .15 on only one of the 24 behavioral cognition tests, then whatever these behavioral tests measure, it was certainly not general intelligence. However, subsequent research has called into question this contention of O'Sullivan et al. For example, Shanley et al. (1971) found significant correlations between the SFTSI and Otis IQ scores especially on their ninth grade sample where they obtained a correlation of .64 between the SFTSI composite (tests 1, 2, 3, and 6) and the Otis IQ scores. The significant correlations reported by Futterer (1973) between the Terman Concept Mastery Test and four of the SFTSI give further evidence for such questioning.

In the present study, correlations between the SFTSI (composite 1) and the three measures of AI for the total sample were .68 for the OTIS, .78 for the K-A, and .63 for the QWT. There were some differences between the male and female samples (see Tables 3 and 4), but all correlations ranged from .59 to .80 which demonstrates highly significant correlations. In several cases, the SFTSI cor-
related higher with measures of AI than they did among themselves, suggesting that not only is the SFTSI not a very pure measure of SI, but that it may really be measuring general intelligence by means of a different dimension. Gerdeman (1973) found similar results, though others claim that the SFTSI does load significantly on a factor that they designate as social intelligence (Futterer, 1973; Pavlou, 1973).

It certainly seems reasonable that verbal ability would influence performance on the Guilford measures, especially in light of the fact that their whole rationale is based on a cognitive model. And if only a moderate relationship was found, this could easily be explained and understood since predictions of the behavior of others naturally take place in our conscious thought processes which are often carried on in verbal terms. Yet correlations ranging from .59 to .80 cannot be so readily explained, and certainly call into question the contention that the SFTSI is really measuring something other than general or abstract intelligence.

It is apparent that the present data represents a restricted sample which is most applicable to white, middle-class college students with average or higher IQs, and that the obtained results which are contradictory to other studies in this area (e.g., Hoepfner & O'Sullivan, 1968; Pavlou, 1973) may be due to the particular sample investigated.

As mentioned above, the purpose of this study was to further investigate the more promising instruments purported
to measure SI. From the results of this study the following conclusions seem warranted:

1. The three measures of SI, to wit the Chapin Social Insight Test, the Hogan Empathy Scale, and the Six Factor Test of Social Intelligence, do not show strong evidence of measuring the same trait.

2. Intercorrelations and factor analysis show that the SI measures load low to high on a verbal factor and do not lend support for the existence of an independent SI factor--at least, not one that they are able to tap.

3. Finally, slight evidence supporting the contention that females are more socially intelligent than males was found on one measure of SI (CHSIT).

From these conclusions it seems that paper and pencil methods of measuring social intelligence are contaminated by factors of general intelligence, and that other directions of measurement such as development of behavior rating scales must be pursued if the elusive concept of social intelligence is ever to be effectively grasped.
CHAPTER VI

SUMMARY

This study has attempted to further explore the contention that SI is a factor that is substantially independent from AI. Previous tests have consistently correlated moderately to high positive with AI measures.

In this study three measures of SI were chosen which were either relatively new and/or on which little validation research had been conducted. These three measures of SI are the Six Factor Test of Social Intelligence (SFTSI), the Hogan Empathy Scale (HES), and the Chapin Social Insight Test (CHSIT).

These measures were compared with three measures of general intelligence (Quick Word Test, Otis Quick-scoring Tests of Mental Ability, and the Kuhlmann-Anderson Test, 7th edition), and the resulting correlations factor analyzed with a varimax rotation (Nfactors = 2).

The Ss consisted of 31 males and 28 females, all undergraduates of Loyola University. These were all volunteers from an introductory psychology class fulfilling part of a required amount of experimental participation. Each subject was administered all six tests. The following

56
hypotheses were proposed: (1) convergent validity among tests of abstract intelligence would be demonstrated; (2) convergent validity among tests of social intelligence would be demonstrated; (3) divergent validity between tests of AI and tests of SI would be demonstrated; and, (4) no difference between male and female would be obtained on SI measures.

Pearson product-moment correlations confirmed hypothesis 1 where AI measures correlated in the high positive direction among themselves. Hypotheses 2 and 3 were not confirmed. Not only did the SI test not correlate highly among themselves, but contrary to expectations they tended to obtain higher correlations with the tests of AI. There was one exception to this general tendency, and that was the moderately high correlation between the HES and the SFTSI (.43 to .48 across samples). In both the male and female samples correlations between the SFTSI and all AI measures were markedly high suggesting that the SFTSI was measuring the same ability as the three measures of abstract intelligence.

The correlations were factor analyzed with a varimax rotation. Factor 1 had high loadings (above .65) on all measures of AI and on the SFTSI. The HES also loaded above the conventional cutoff of .30 on factor 1. Factor 2 did not obtain the conventional 1.0 eigenvalue cutoff and therefore did not seem interpretable other than as pointing out some sex differences.
The following conclusions seemed warranted: the measures of SI did not show strong evidence of measuring the same trait; intercorrelations and factor analysis showed that the SI measures loaded low to high on a verbal factor and did not lend support for the existence of an independent SI factor; finally, slight evidence supporting the contention that females are more socially intelligent than males was found on one measure of SI (CHSIT). This last finding partially disconfirmed hypothesis 4.
REFERENCES


Allport, F. H., & Allport, G. W. Personality traits: Their classification and measurement. *Journal of Abnormal and Social Psychology*, 1921, 16, 6-40.


Bender, I. E., & Hastorf, A. H. On measuring generalized empathic ability (social sensitivity). *Journal of*
Abnormal and Social Psychology, 1953, 48, 503-506.


Borgatta, E. F., & Corsini, R. J. The Quick Word Test (QWT) and the WAIS. Psychological Report, 1960, 6, 201. (a)


Dymond, R. F. A scale for the measurement of empathic ability. *Journal of Consulting Psychology, 1949, 13,* 127-133.


Flavell, J. H. Role-taking and communication skills in children. *Young Children*, 1966, 21, 164-177.


Grotelueschen, A., & McQuarni, D. Cross-validation of the Quick Word Test as an estimator of adult mental ability. *Adult Education*, 1970 (Fall), Vol. 21 (1), 14-19.


Kannes, L. Judging emotions from facial expressions of emotions. Psychological Monographs, 1931, 41, No. 3 (Whole No. 186).


Lemann, T. B., & Solomon, R. L. Group characteristics as revealed in sociometric patterns and personality ratings. Sociometry, 1952, 15, 7-90.


APPROVAL SHEET

The thesis submitted by Patrick McKian, has been read and approved by members of the Department of Psychology.

The final copies have been examined by the director of the thesis and the signature which appears below verifies the fact that any necessary changes have been incorporated and that the thesis is now given final approval with reference to content and form.

The thesis is therefore accepted in partial fulfillment of the requirements for the degree of Master of Arts.

Date

Signature of Advisor

10/30/73