THE RELATIONSHIP BETWEEN EMPATHY AND SUPERVISORS' RATINGS OF STUDENT NURSES

APPROVED:

Earl W. Hooper
Major Professor

Harold A. Holloway
Minor Professor

Harold A. Holloway
Chairman of the Department of Psychology

Robert B. Trouboue
Dean of the Graduate School
THE RELATIONSHIP BETWEEN EMPATHY AND SUPERVISORS' RATINGS OF STUDENT NURSES

THESIS

Presented to the Graduate Council of the North Texas State University in Partial Fulfillment of the Requirements

For the Degree of

MASTER OF SCIENCE

By

Joseph John Libbon, B.S.
Denton, Texas
January, 1970
TABLE OF CONTENTS

LIST OF TABLES ........................................ iv

Chapter

I. INTRODUCTION ........................................ 1
   Statement of the Problem
   Hypotheses

II. RELATED LITERATURE ................................. 8

III. METHOD ............................................. 20
   Subjects
   Instruments
   Procedure
   Statistical Treatment

IV. ANALYSIS AND DISCUSSION OF RESULTS ........... 24

V. SUMMARY AND RECOMMENDATIONS ...................... 31

APPENDIX .............................................. 33

BIBLIOGRAPHY .......................................... 40
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I.</td>
<td>Age Means and Standard Deviations of Student Nurses</td>
<td>20</td>
</tr>
<tr>
<td>II.</td>
<td>The Relationship of Odd and Even Items on the Diplomacy Test of Empathy</td>
<td>25</td>
</tr>
<tr>
<td>III.</td>
<td>The Relationship of Strictly Odd and Even Items on the Diplomacy Test of Empathy</td>
<td>25</td>
</tr>
<tr>
<td>IV.</td>
<td>The Relationship Between Rater 1 and Rater 2 on the Performance Rating Scale</td>
<td>26</td>
</tr>
<tr>
<td>V.</td>
<td>Mean Scores and Standard Deviations on the Diplomacy Test of Empathy</td>
<td>27</td>
</tr>
<tr>
<td>VI.</td>
<td>Summary of Analysis of Variance Computations for Three Levels of Performance</td>
<td>27</td>
</tr>
</tbody>
</table>
CHAPTER I

INTRODUCTION

As used by the layman, the term empathy means having a feeling of compassion for another person. In a study done in 1948, Dymond defined empathy as "... the ability to feel and describe the thoughts and feelings of others, ..." (3, p. 232). She felt that there was a relationship between empathy and childhood experiences.

Empathic ability appears to be related and at least in part due to the character of the relations which surround the child in the home when young; close relations leading to higher empathic ability than distant ones (2, p. 343).

Lois Murphy found the development of sympathy related to empathy.

In the case of sympathy, probably general thresholds for being affectionate, for seeing similarities between others' situations and our own, and for empathic responsiveness underlie sympathetic habits ... If we could find measures for susceptibility to empathic responses, if such they be, we might well find the basis for the most important individual differences in sympathy (8, p. 18).

As used in the above study, empathy forms the basis from which the development of sympathy takes place. Empathy provides an underlying foundation for the growth of various dimensions of sympathy.
Mead also viewed empathy and sympathy as being very closely related.

The attitude that we characterize as that of sympathy springs from this same capacity to take the role of the other person with whom one is socially implicated. Sympathy always implies that one stimulates himself to the assistance and consideration of others by taking to some degree the attitude of the other person whom one is assisting. The common term for this is "putting yourself in his place" (7, p. 366).

Empathy has also been described in psychoanalytic terms.

Empathy can be described as a process of "projection" or "introjection"; both are metaphors referring to the experience of partial identity between the subject's mental processes and those of another with the resulting insight into the other's mental state and participation in his emotions (6, p. 360).

Hoskins hypothesized that distorted empathy may be a prime defect in schizophrenia.

It throws open the possibility that the primary defect in schizophrenia, a defect from which the remainder of the symptomatology stems, is inadequate empathy (4, p. 102).

Perhaps as fundamentally characteristic as anything about the psychosis is the failure of the subject either to achieve or retain adequate breadth or depth of empathy (4, p. 165).

It becomes evident that empathy is very closely associated with interpersonal relations. Any time two or more people interact they must take into account the views and opinions of the people surrounding them. A person who is able to show consideration for another's viewpoint is able to discern what this viewpoint has as its composition.
Brayfield expressed his idea of empathy as "... the ability to put yourself in the other person's position, establish rapport, and anticipate his reactions, feelings, and behaviors" (1, p. 187). So it can be seen that empathy is much more than simple compassion or sympathy. The perfectly empathetic person would be one who could take on the personality facets of another person's way of life. Dymond supported this definition by saying empathy is, "... the imaginative transposing of oneself into the thinking, feeling, and acting of another and so structuring the world as he does" (2, p. 343).

Putting oneself in another's position can be seen in the rationale of client-centered therapy. The therapist strives to understand his client's phenomenological world. He needs to view the client as the client views himself (9, p. 86).

In studies to be cited later it was shown that the field of nursing has been a fruitful area for the study of empathy. This is because the nurse, due to her close interaction with patients, is regarded as one who is in a position to develop a great deal of empathy in her on-the-job situation. Schwartz and Shockley stated that a nurse must be capable of discovering and understanding a patient's needs and that she must be able to fulfill his need for intimacy and companionship (10, p. 225).
By trying to put herself in the patient's place imaginatively and emotionally, she might be able to identify a particular need at a particular time, to determine which need is paramount at the moment, and to gain some appreciation of the meaning of the patient's needs to him (5, p. 230).

The more a nurse is able to transpose herself into the patient's phenomenological world the better able she is to perform the duties expected of her. Kandler and Hyde (5) stated that empathic traits play a large role in the effectiveness of psychiatric student nurses. They observed that there were marked individual differences in nurses' ability to relate to and to understand their patients. "This suggests that the ability to empathize with patients is a major factor in their performance" (5, p. 33).

So it becomes meaningful to study and investigate the role empathy plays in relation to performance. It seems possible that the two are closely related. A nurse's ability to empathize with her patients plays a large role in her overall capability to perform her duties adequately. This involves being able to discover her patients' needs and wants. This is best done by placing herself in the mental set of the patient, being able to empathize with him.

Statement of the Problem

The purpose of this study was to investigate the relationship between a measurement of empathy and performance of psychiatric student nurses. The study was also concerned
Hypotheses

It was hypothesized that there would be a demonstrable relationship between empathy and the performance of psychiatric student nurses. It was believed that there would be a significant difference between those nurses rated in their performance by their supervisors as high, middle, and low performers as evidenced by their score on Kerr's Diplomacy Test of Empathy (see Appendix A). It was further hypothesized that this relationship would be linear, in that as performance scores increased so would the empathy scores.

It was also hypothesized that the odd-even reliability of the scores of the student nurses on the DTE would be high enough to enable it to be considered a reliable measure of empathy of student nurses. In addition it was hypothesized that the reliability of the performance instrument would be high enough to indicate that it was an adequate device for measuring the performance of the student nurses.

In summary, the hypotheses of this study were as follows:

I. The Diplomacy Test of Empathy would be a reliable measure of empathy of psychiatric student nurses.

II. The performance rating scale developed for this study would be a reliable measure of performance of student nurses.
III. There would be a linear relationship between the scores of the nurses on the Diplomacy Test of Empathy and their performance as rated by their nursing supervisors.
CHAPTER BIBLIOGRAPHY


CHAPTER II

RELATED LITERATURE

Among the first formal attempts to measure empathy was a study done by Dymond in 1948 (5). She administered the Thematic Apperception Test to a group of twenty student volunteers. How the characters were used by the students in their stories was rated into three categories. The character was rated good if the thoughts and feelings imputed to him were richly described. A poor rating was given if the character was named but not described in any way. A rating of fair was given if the character was described only on the basis of physical attributes. If 40 per cent of all characters described were given a rating of good then the person was said to have high empathic ability. Anyone scoring below 40 per cent was said to be low in empathy.

In a more statistically oriented study, Dymond (6) again attempted to objectively measure empathy. The test employed four parts, each part a rating of the same six characteristics: self-confidence, superior-inferior, selfish-unselfish, friendly-unfriendly, leader-follower, and sense of humor. In the first part the subject rated himself; in the second he rated each group member (Ns = 6 or 7); in the third part he rated the other group members as they would
rate themselves; and in the fourth part he was to rate himself as he thought each group member would rate him. In an attempt at validation, she first determined if the rating difference scores were significantly different from chance. They were at the .01 level. The ten subjects rated highest and lowest were then given the TAT as a further attempt at validation, using her previous hypothesis. There was a significant agreement between the two methods.

In 1950 Dymond (4) revised her previous test in that she changed the six characteristics used to make the ratings. Once more the difference scores were significant at the .01 level as compared to chance. In an attempt to describe the personality of the empathic person, she studied her high and low scorers on the basis of the Wechsler, Rorschach, TAT, and the California Ethnocentrism Test. She did find differences between her two groups, but she did not submit these differences to any statistical analysis.

Hastdorf and Bender (7) raised the question of how much projection and similarity enter into empathy scores such as those derived by Dymond. It was their hypothesis that if the predictor had values very similar to the person for whom he was predicting, their deviation score would be greatly reduced. In such a case a spuriously high degree of empathy would be measured. Hastdorf and Bender advised a compiling of two deviation scores. One would be an empathy score according to Dymond's techniques and the other deviation
score would result in the removal of the effects of projection and similarity by comparing it with the first score. This would be done by measuring the deviation of the prediction score from the predictor's own test score. If these two deviation scores closely resembled each other it would be assumed that the subjects were highly similar and that it was projection rather than empathy which was bringing about the so-called empathy measure.

In a study done in 1953, Bender and Hastdorf (2) again investigated the interaction between empathy, projection, and similarity. In this research they derived three deviation scores instead of only two. They computed a score for raw empathy, projection, and also similarity. They then used these deviation scores to develop a refined empathy score. To test the validity of this concept, they computed a product-moment correlation between each of the deviation scores. They found a significant relationship between raw empathy and projection and between raw empathy and similarity. After computing refined empathy scores they found no relationship between projection, similarity, and empathy; thus indicating that these were variables extraneous to the empathy measure as previously conjectured.

Bell and Hall (1), using leadership rankings and Kerr's Empathy Test (9) as variables, found a significant correlation of .25. This study used very small groups (N = 5) which were initially leaderless and who met for only one
session to discuss a human relations problem. At the end of the thirty-minute session the group members were asked to rank their leadership preferences. These preferences were then correlated with the test scores. Using empathy scores derived from Dymond's previously mentioned study (6) and again correlating them with the leadership rankings, a non-significant coefficient of .18 resulted.

Remmers and Remmers (12), in 1949, used the test How Supervise? to gain a measure of empathy. They had 100 labor leaders fill out the test according to the standard instructions. They were then asked to fill it out a second time. This time they were to try and answer it as a typical company man would answer it. They found that the labor leaders answering as company men scored significantly lower than they did when answering it as themselves. Evidently, the labor group was unable to conceive of how the questions should be answered from the frame of reference of the management ranks.

In 1950 Miller and Remmers (11) effectively duplicated the above study with one exception. This time management instead of labor filled out How Supervise?, first for themselves and as they felt labor leaders would respond. Again they found a significant difference between the two sets of scores. On this occasion, management consistently rated the labor leaders as having lower scores than the management group. This again indicates that these two groups do not show agreement within these certain areas.
Chambers (3) utilized Dymond's empathy techniques in a study done in 1957. He also incorporated the Bender-Hastdorff correction (2). Chambers correlated empathy with scholastic success and with scholastic aptitude. He used fifty-five college freshmen and divided them into small groups of three and four in order to make their predictions of each other. He found a significant correlation of .37 between their empathy scores and grade-point average. He did not find a significant correlation between empathy and scholastic aptitude. This indicated that empathy is an important factor in the learning process. It seems that it is almost as important to know the empathic ability of a person as it is to know his IQ when predicting scholastic success.

A major forerunner of the test used in the present study was published in 1951. Its title was The Empathy Test (9). This was the only standardized test for the measurement of empathy previous to the Diplomacy Test of Empathy. The Empathy Test was made up of three sections. In the first section the testee is asked to rank in order of popularity, fourteen musical types. In the second section he is asked to rank the popularity of fourteen magazines. The third section calls for a ranking of annoying experiences. The test is scored by comparing the answers given with the actual results of a nation-wide survey conducted in connection with the construction of the test. The test has been significantly correlated with sociometric values of college
men, supervisory attitude, and sales records of automobile salesmen (9). Test-retest reliability has been found to range from .67 to .81. Van Zelst (15) found The Empathy Test to be significantly correlated with six different criteria of leadership. In another validity study Van Zelst (16) found this test to have a significant correlation with job satisfaction and self-judgment.

The test selected for use in the present study was the Diplomacy Test of Empathy (11), hereafter referred to as the DTE. It was devised by Kerr in 1960. It consists of thirty-seven items of the multiple-choice variety (see Appendix A). The first twenty-two questions require two answers, most and least. For example, item 1 asks the examinee to determine from three choices which are most and least annoying to a man between the ages of twenty-four to thirty-six. The instructions on the test require that the person taking the test answer the questions, not as he feels, but as specific reference groups would answer them.

Kerr reported a mean validity coefficient of .46 in the studies cited in the manual of the test. His validity studies were done with salesmen and upper management workers. In a test such as this, intended for use in an employment battery, he concluded that the validity is high enough to be considered worthwhile.

The test is currently being used industrially with good results being obtained. At the present time it composes
part of a national hiring battery being used with good results by a large electronics firm. Industrial competition has prevented publication of their findings (10). There are several other industries currently making use of the DTE, but again competition hampers publication of results.

Korr stated in the manual that he found an odd-even reliability of .56. It could not be determined if this reliability was corrected by the Spearman-Brown formula. It is a modified odd-even test due to the fact that he paired the items according to empathic content and randomly separated them into two groups. He stated that this low reliability is to be expected due to the relatively small number of items included in the test. It was shortened to enhance the length of time of administration.

In 1963 the DTE was used by Smouse, Aderman, and Van Buskirk (14) as part of a three-test battery used to study correlations between empathy measures and tests and rating criteria. The subjects used in their study were junior (N = 110) and senior (N = 119) female nursing students at a large midwestern hospital. The three tests used in the battery were an incomplete sentences test, a fault-value combination, and the DTE. The junior nurses' scores on these tests were correlated with their scores on the Graduate Nurse Examination (GNE) and the Graduate Nurse Qualifying Examination (GNQE). The scores of the senior nurses were correlated with scores on the GNE and rating criteria of
performance as determined by independent observers. The scores on the GNE and GNQE are given in four areas: medical-surgical, maternal and child, psychiatry, and general science. The DTE showed positive significant correlations with all of these areas but the general science. It also showed a significant correlation with the total score of the GNQE. The highest overall correlation was .31 between the DTE and the psychiatry subscores of the GNQE.

Use of the DTE was next reported by Sulzer and Burglass in 1968 (13). They studied responsibility attribution as a function of the psychological variables empathy and punitiveness. It was their hypothesis that the degree to which a person attributed the cause of adverse events as occurring due to someone else's fault would vary inversely with empathy and directly with punitiveness. They used two separate groups of subjects. The first group were female students (N = 112) in an introductory psychology course. The second group consisted of sixty-eight Air Force men tested several months later. The data from each group were analyzed separately. Their results are somewhat contradictory. The correlation coefficients for the psychology students were all close to zero, while those for the airmen were significant and in the predicted direction. The r between punitiveness and responsibility attribution was .42, and the r between empathy and responsibility attribution -.24. One of the factors they cited as possibly contributing to the difference
between the students and airmen was that there were apparent sex differences, but they did not offer any statistical proof of this statement. Another factor they discussed was that the students were probably test-conscious and less spontaneous than the airmen. This seems unlikely as these were students in introductory psychology classes and would not as yet have been exposed to a great variety of psychological tests.

The DTE was again used in 1968 by Kelleher, Kerr, and Melville (8). They attempted to develop a test battery to be used in the prediction of performance of nursing aide applicants. Training program survival, final examination score, salary increment time, and length of job tenure were the four areas of performance they attempted to predict. The mean score of the applicants on the DTE was 31.55, with a standard deviation of 5.96. The subjects (N = 200) who participated in the study were all high school graduates with no previous experience in the nursing field. Their age range was from eighteen to forty, with a mean of twenty-eight and a standard deviation of six years. The coefficient correlations between the DTE and the four criteria were as follows: -.20, .02, -.03, -.06. The only significant correlation was -.20, being correlated with training program survival. The DTE was then combined with two other tests which had correlated with training program survival to a significant level. This small battery was then subjected to
multiple correlation and found to be correlated with the training criterion at the .05 level, $r = .39$. And upon being submitted to cross-validation, these three tests showed a correlation of .45, significant at the .01 level.

Research done with the DTE in the industrial field has shown it to be a useful tool when it is combined in various test batteries (9, 10). Contradictory results were found when the DTE was used as a validational study of punitiveness and responsibility attribution (13). It has also been useful in the prediction of nursing achievement (8, 14).
CHAPTER BIBLIOGRAPHY


17. , "Validation Evidence on The Empathy Test," Educational Psychology, 13 (June, 1953), 474-477.
CHAPTER III

METHOD

Subjects

The subjects used in this study were student psychiatric nurses. They were enrolled at three different colleges in a southeastern state. There were thirty-two subjects from one school, fourteen from the second, and twenty-one from the third. The subjects were undergoing their clinical psychiatric training at a state mental hospital during the summer of 1969. This training period was the culmination of their degree programs at their respective schools. The subjects' mean ages are shown in Table I.

TABLE I

AGE MEANS AND STANDARD DEVIATIONS OF STUDENT NURSES

<table>
<thead>
<tr>
<th>Variable</th>
<th>Means</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>N = 30*</td>
<td>25.80</td>
<td>11.097</td>
</tr>
<tr>
<td>N = 14</td>
<td>24.36</td>
<td>6.34</td>
</tr>
<tr>
<td>N = 21</td>
<td>23.19</td>
<td>6.01</td>
</tr>
</tbody>
</table>

*Two subjects dropped from study.

For purposes of statistical analyses, the groups from the second and third schools were combined to form a composite group of thirty-five. This was the group used to compute
between rater reliability and the odd-even reliability. The remaining group of thirty-two was used in the main study concerned with determining the relationship between empathy and performance.

Instruments

A rating scale, consisting of six items, was developed for the study and was used by the supervisors in their rating of the performance of the student nurses. An original pool of five items was presented to the six supervisors. With their advice, two changes in wording were made and one item was added to the pool. The completed scale can be found in Appendix B. The supervisors expressed the opinion that these items seemed to them to be important aspects of the nurses' performance. It was not intended that these items encompass all areas of nursing performance, only those germane to the study.

Procedure

The first step in the study was the acquisition of the performance ratings from the nurses' supervisors. They were asked to make their ratings independently of each other, without consultation. If they were in doubt as to what rating to give on a certain item, they were told to try and recall a specific incident involving the nurse in mind and to make their decision on that basis.
The nursing supervisors used to determine the performance ratings were all graduate, registered nurses. Each of the three groups of student nurses had two supervisors of equal rank conducting their training. They were in close contact with the student nurses, both during the summer training program and throughout the previous school year. It was felt that these supervisors were sufficiently familiar with the students to be able to give reliable ratings of their performance.

The second half of the procedure involved the administration of the DTE. It was given to the three groups on successive days during their regularly scheduled lecture periods. Because the nurses were from widely separated schools there seemed to be no interaction between them and it was thought that no communication took place between those tested and those untested about the contents of the test.

The nurses were not told the purpose of the study nor that ratings of their performance were being compiled. They were told only that the test being given them had been designed for salesmen and that an attempt was being made to see if it was applicable to nurses also. They were told that their test scores would not be seen by their supervisors and that it would in no way affect their grades in their training course. It was made clear to them that the test was not a regular part of their training and was being done
only for use in a master's thesis. The nursing supervisors, who did the ratings of the nurses, were fully acquainted with the purpose of the study, but they were not shown a copy of the DTE.

The intact group from each school was tested. To allay any fears of lack of confidentiality, they were told to use only their initials as identification. This same procedure was followed by the supervisors when they completed the rating sheets.

Statistical Treatment

1. To test Hypothesis I, that the DTE would be a reliable measure of empathy, a Pearson coefficient of correlation was calculated on the odd-even scores of the composite group \((N = 35)\). The odd-even items used as variables were the same as those outlined by Kerr in his manual for the DTE. The resulting coefficient was corrected by the Spearman-Brown formula.

2. It was also hypothesized that the rating scale would be a reliable measure of performance. This was tested by computing a Pearsonian correlation between rater 1 and rater 2 of the composite group \((N = 35)\).

3. A one-way analysis of variance was used to test the hypothesis concerning the relationship between empathy and rated performance. This was done for the group with an \(N\) of 32.
CHAPTER IV

ANALYSIS AND DISCUSSION OF RESULTS

The first hypothesis of the study was concerned with the reliability of the DTE as an empathy measure. The only previously reported reliability was an odd-even one computed by Kerr (see p. 14). He found an r of .56. Since this was calculated on salesmen only, it was in order to compute an odd-even reliability on student nurses to see if the DTE could be used as a reliable measure of empathy of student nurses.

As mentioned previously, the two smaller groups were combined to form a group with an N of 35. A Pearson coefficient of correlation was computed on this group using their odd-even scores on the DTE as variables. This correlation on the combined group of 35 was .215. With a correction by the Spearman-Brown formula, this became .354, p < .05. The results of calculating this correlation can be seen in Table II. Though this correlation was significant, it was also quite low and only partially supports Hypothesis I.

Kerr's definition of odd-even items was used in the computation of the above correlation. In the manual for the DTE, Kerr paired the items for empathic content and then randomly separated them into two groups which he designated
TABLE II
THE RELATIONSHIP OF ODD AND EVEN ITEMS ON THE DIPLOMACY TEST OF EMPATHY

<table>
<thead>
<tr>
<th>Variable</th>
<th>Means</th>
<th>S.D.</th>
<th>r</th>
<th>$r_{tt}$</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Odd</td>
<td>20.43</td>
<td>3.67</td>
<td>.215</td>
<td>.354</td>
<td>.05</td>
</tr>
<tr>
<td>Even</td>
<td>21.68</td>
<td>3.52</td>
<td>.055</td>
<td>.347</td>
<td>.05</td>
</tr>
</tbody>
</table>

odd and even. A strict odd-even coefficient was also determined as a check on Kerr's rationale of pairing of items. This resulted in a correlation of -.00685. The results of this computation are shown in Table III. This correlation supported Kerr's pairing of items according to empathic content.

TABLE III
THE RELATIONSHIP OF STRICTLY ODD AND EVEN ITEM ON THE DIPLOMACY TEST OF EMPATHY

<table>
<thead>
<tr>
<th>Variable</th>
<th>Means</th>
<th>S.D.</th>
<th>r</th>
<th>$r_{tt}$</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Odd</td>
<td>20.06</td>
<td>4.16</td>
<td>-.00685</td>
<td>-.012</td>
<td>N/S</td>
</tr>
<tr>
<td>Even</td>
<td>21.66</td>
<td>3.92</td>
<td>-.00685</td>
<td>-.012</td>
<td>N/S</td>
</tr>
</tbody>
</table>

The second hypothesis which was tested was that the rating scale would give a reliable measure of the students' performance. This hypothesis was tested by calculating a between-rater reliability through the use of a Pearson correlation coefficient. The variables used were the ratings
given the nurses by their supervisors, designated as rater 1 and rater 2. This coefficient of reliability was computed to be .77, p < .01. The results of this calculation can be seen in Table IV. This highly significant correlation

<table>
<thead>
<tr>
<th>Variable</th>
<th>Means</th>
<th>S.D.</th>
<th>r</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rater 1</td>
<td>13.77</td>
<td>3.27</td>
<td>.77</td>
<td>.01</td>
</tr>
<tr>
<td>Rater 2</td>
<td>16.46</td>
<td>3.29</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

supports Hypothesis II, which states that the performance rating scale is a reliable measure of student nursing performance.

To test the major hypothesis of the study—that there would be a relationship between empathy and performance—a one-way analysis of variance was computed. At this final stage the group had an N of thirty. Two of the student nurses had to be dropped from the study due to their failure to correctly follow instructions on the DTE. A time factor made it impossible to retest them.

The nurses were ranked from low to high according to their total performance score. This score was obtained by adding the numerical ratings given them by raters 1 and 2. On the basis of these rankings the nurses were divided into
equal thirds of high, middle, and low. The other variable used in the analysis of variance was their total score on the Diplomacy Test of Empathy. Their mean scores and standard deviations on the empathy test can be seen in Table V.

TABLE V

MEAN SCORES AND STANDARD DEVIATIONS ON THE DIPLOMACY TEST OF EMPATHY

<table>
<thead>
<tr>
<th>Variable</th>
<th>Means</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>41.2</td>
<td>3.84</td>
</tr>
<tr>
<td>Middle</td>
<td>42.2</td>
<td>6.39</td>
</tr>
<tr>
<td>Low</td>
<td>41.0</td>
<td>5.00</td>
</tr>
</tbody>
</table>

The analysis of variance computation resulted in a non-significant F of .138. Thus, Hypothesis III is rejected.

A summary of analysis of variance can be seen in Table VI.

TABLE VI

THE ANALYSIS OF VARIANCE FOR THE DTE SCORES FOR THE THREE PERFORMANCE GROUPS

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Squares</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between</td>
<td>8.27</td>
<td>2</td>
<td>4.135</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within</td>
<td>807.20</td>
<td>27</td>
<td>29.896</td>
<td>.138</td>
<td>N.S.</td>
</tr>
<tr>
<td>Total</td>
<td>815.47</td>
<td>29</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Discussion

The hypothesis in this research study was not substantiated by the statistical analysis. There was no significant relationship between rated performance and empathy as measured by Kerr's Diplomacy Test of Empathy. There were not even any apparent trends in the predicted directions. This can be readily seen by an investigation of the mean scores of the nurses on the empathy test included in Table V. The range is only 1.2 points. There was only the barest of difference between the three groups on their empathy scores and the middle group obtained the highest score.

Due to the close working relationship between the nurses' supervisors, it was felt that the items selected for the performance rating sheet did provide some measure of performance. And the highly significant correlation between rater reliability shows that the supervisors were able to agree on it to a significant level. However, whatever was being measured did not have a significant relationship with the empathy measure.

An odd-even reliability coefficient of .354 is much too low to be of any use in the prediction of individual differences. The test items are extremely heterogeneous and even when paired as to empathic content, their reliability remains quite low. This low reliability indicates that the test does not consistently measure what it is designed to measure.
The extent to which this test measures empathy does not seem suited for use with student psychiatric nurses. If the mean scores on the empathy test were adapted to Kerr's norms for industry, the nurses would have made the seventy-fifth percentile for routine workers and only the thirty-third percentile for upper management.

It is possible that industrial empathy differs from nursing empathy. The test items may be useful in the industrial field, but have no relevance concerning nursing behavior. This would have a bearing on why Hypothesis III was rejected.

There is also the possibility that successful nursing does not require any high degree of empathy. That is, to be successful a nurse does not have to be able to establish close interpersonal relationships. This too would lead to the rejection of Hypothesis III.

It might also be that this group of student nurses was not typical. Their training program is only two years in length, rather than the standard four-year college curriculum. To them the program may be more a learning of a trade rather than the techniques and skills of nursing. In other words, they may not have been as dedicated as the more typical nurse would be expected to be. And accordingly their development of empathy would not be expected to reach as high a level.

Another factor to be considered is that the scale used to measure performance may have been inadequate. Even though
its reliability was sufficiently high, it is possible that the items did not sample the interpersonal relations which exist between nurse and patient. Thus the performance scores were not true indications of this aspect of nursing behavior. Thus it could be that a nurse has to be able to interact with her patients in order to perform her duties, but the scale did not tap this dimension. This would have contributed to a rejection of Hypothesis III.
CHAPTER V

SUMMARY AND RECOMMENDATIONS

This study was concerned with the investigation of three hypotheses. The first dealt with the reliability of the Diplomacy Test of Empathy as a measure of empathy. The second hypothesis was concerned with the reliability of a performance rating scale developed for this study to measure performance of student psychiatric nurses. The third hypothesis tested the theory that empathy would be related to performance.

Hypothesis I was tested by the calculation of odd-even reliability coefficient for the DTE. The subjects were thirty-five psychiatric student nurses. This coefficient was .354, significant at the .05 level. Due to this low correlation, Hypothesis I could be only cautiously accepted.

Hypothesis II was investigated by computing a between-rater reliability coefficient, using as variables the scores given the nurses by raters 1 and 2 on the performance rating scale. This highly significant correlation of .77, p < .01 led to the acceptance of Hypothesis II—that the rating scale would yield a reliable measure of student nursing performance. The same subjects were utilized, as for testing Hypothesis I.
Hypothesis III was tested by using the ratings given the nurses by their supervisors to classify the nurses into three groups. Their scores on the DTE were used as the criterion measures for a one-way analysis of variance. A different group of nurses were used as subjects in this part of the study, N = 30. A non-significant F of .138, df = 2 and 27, led to the rejection of Hypothesis III.

Recommendations

It should be noted that the empathy measure was highly unreliable in its use with student nurses. It is felt that a test definitely designed for the measurement of empathy in nurses should be employed in other studies of this type.

It is also recommended that a different performance scale be used. Even though highly reliable, the one used in the present study may have been insensitive to interpersonal relations which play so large a part in the interactions between a nurse and her patients. It is possible that the use of rather small N's prevented extreme scores from arising. It might have been possible that these extremes would have produced significant results.
Name: ___________________________ Age: __________ Job Title: ____________________________

Write in the number of times you have held each of the following positions in school or other organizations:  
___ treasurer; ___ secretary; ___ president; ___ vice-president; ___ athletic team captain; ___ board chairman;  
___ committee chairman; ___ fund drive chairman; ___ homecoming chairman.

How many years have you held full-time jobs? ______

Subtract the monthly pay of your first full-time job from your present monthly pay, and record the DIFFERENCE below:

$ __________________________

Directions: This is a test of not what you think and feel but of your knowledge of what OTHERS think and feel. In other words, it measures your ability to put yourself in the OTHER PERSON'S POSITION and think and feel as he does. Skip no item but answer as directed. Write only on the Answer Sheet. Please note that each of the first 22 items requires TWO answers, a "Most" answer and a "Least" answer.

Remember the M or L that you "X-out" is NOT eliminated but is SELECTED as your MOST or LEAST answer!

Copyright 1980
Psychometric Affiliates
Ability and Opinion Measurement Specialists Since 1971
Chicago Plaza
Brookport, Illinois 62910
DIPLOMACY TEST
of Empathic Ability

by Willard A. Kerr
Industrial Relations Center
University of Chicago

Copyright 1960

PSYCHOMETRIC AFFILIATES

Directions: this is a test of not what you think and feel but of your knowledge of what OTHERS think and feel. In other words, it measures your ability to put yourself in the OTHER PERSON’S POSITION and think and feel as he does. Skip no item but answer as directed. Write only on the Answer Sheet. Please note that each of the first 22 items requires TWO answers, a "Most" answer and a "Least" answer.

Instructions: Place this test over all except the extreme right column of the Answer Sheet. Have arrows meet. Now, in each group of 3 possible answers, "X-out" one M (most) and one L (least).

1. Most and Least annoying to persons aged 25 - 39:
   a. A person bragging about himself
   b. A salesman trying to force me to buy something
   c. Hearing sarcasm

2. Most and Least copies printed per issue:
   d. Popular Mechanics
   e. McCalls
   f. Forbes

3. Most and Least important in a job to the average man:
   g. Job security
   h. "Boss"
   i. Fellow employees

4. Most and Least employees think is the best way of getting ahead where they work:
   j. Staying long with the company
   k. Being a good politician
   l. Showing energy and willingness to work

5. Most and Least old employees worried sometime in their lives about:
   m. Appearance
   n. Alcoholism
   o. Work associates

6. Most and Least men worried earliest in their careers about:
   A. Job security
   B. Lack of confidence in meeting people
   C. Alcoholism

7. Most and Least important in jobs in the opinions of married men:
   D. Supervisor
   E. Type of work
   F. Advancement
35

3. Most and Least people worry seriously about:

   G. Neighbors
   H. Sex morality
   I. Money

9. Most and Least women are interested in that which is:

   J. Expensive
   K. Social
   L. Governmental

10. Most and least people move to another City to:

   M. Join relatives or friends
   N. Look for or accept work
   O. Get better housing

11. Most and Least used in talk:

   P. Nouns
   Q. Prepositions
   R. Verbs

12. Most and Least people think it cheerful, jovial, joyful:

   S. Brown
   T. Blue
   U. Yellow

13. Most and Least people who have and who have not been fired agree to be a JUST CAUSE:

   V. Being absent from work too much
   W. Refusing to do the work assigned
   X. Purposely damaging materials or equipment

14. Most and Least old people worried sometime in their lives about:

   a. Neighbors
   b. Envying others
   c. Religious convictions

15. Most and Least employees say they are NOW satisfied in their PRESENT JOB with:

   d. Pay
   e. Freedom to complain or suggest
   f. Effect of company on personal happiness
   g. 25 – 29
   h. 30 – 34
   i. 35 – 39

16. Most and Least employees feel pay to be very important in their jobs when they are aged:

   j. Marital difficulties
   k. Economic matters
   l. Health

17. Most and Least men worried earliest in their careers about:

   m. Coronet
   n. Modern Romances
   o. Business Week

18. Most and Least copies printed per issue:

   m. Coronet
   n. Modern Romances
   o. Business Week
19. Most and Least copies printed per issue:
   p. Life-----------------------3 6------
   q. Factory Management & Maintenance------
   r. Ladies Home Journal---------------------

20. Most and Least women regard as most important in jobs:
   s. Pay--------------------------
   t. Fellow employees----------
   u. Comfortable working conditions------
   v. Policemen-------------------
   w. Railroad section hands-------
   x. Musicians--------------------

21. Most and Least highly intelligent people per occupation:
   y. A person with a gushing manner----------
   z. A person not noticing what I say--------
   i. Coaxing me to do something I don't want to do----

Instructions: each of the remaining questions require only ONE answer.

23. Asked what they think most important in jobs, workers say:
   A. Working conditions------------------
   B. Pay-----------------------------
   C. Advancement opportunity-------------
   D. Congenial co-workers---------------

24. Boys and girls are most likely to have front teeth missing when they are aged:
   E. 3 - 5-----------------
   F. 7 - 9----------------
   G. 11 - 13----------------

25. Children and parents have most disagreements about food when the children are aged:
   H. 0 - 3-----------------
   I. 4 - 7----------------
   J. 8 - 11----------------
   K. 12 - 15----------------

26. People who have been "fired" and those who have not been "fired" from their jobs disagree most on which reason being a justification for discharge:
   L. Stealing things-----------------
   M. Being drunk on the job----------
   N. Refusing to do work assigned-----

27. Men aged 35 are most likely to worry about:
   O. Work associates-----------------
   P. Work efficiency-----------------
   Q. Health--------------------------
   R. Morality of self-----------------

28. What is the typical reaction to canned orange juice:
   S. Too thin or watery----------------
   T. Tastes like fresh orange juice------
   U. Just the right sweetness----------
29. Most degrees are granted to women in:
   a. Biology
   b. Architecture
   c. Music
   d. English
   
30. Which do workers think is the poorest reason for promotion?
   e. Being very loyal to the company
   f. Having a good family reputation
   g. Studying for self-improvement
   h. Very good
   i. Average
   j. Below average
   k. Very hard

31. Legislators introduce more bills to curb Communists when times are:
   l. Germans and Poles
   m. French and Poles
   n. Poles and Italians
   o. Germans and Italians
   
32. The most numerous groups of foreign-born Americans:
   p. Biology
   q. Chemistry
   r. Pharmacy
   s. Sociology
   
33. Fewest female college graduates obtain their degrees in:
   t. Being a good talker
   u. Being a hard worker
   v. Wide experience in various kinds of work
   w. 72%
   x. 82%
   y. 92%

34. Which is thought by workers to be the BEST reason for promotion?
   A. Acquisitive
   B. Conventional
   C. Formal
   D. Independent
   
35. What per cent of workers think "not being able to do the work" is a good reason for discharge?
   E. 9
   F. 7
   G. 5
   H. 3
   
36. The most productive scientists see themselves as:
   Now, return BOTH papers to the Test Administrator.
   Thank you!
## APPENDIX B

### PERFORMANCE RATING OF STUDENT NURSES

<table>
<thead>
<tr>
<th>Rater #</th>
<th>Student Nurse's Initials</th>
</tr>
</thead>
</table>

1. When she makes a mistake does she readily acknowledge doing so?

<table>
<thead>
<tr>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Almost always</td>
<td>Sometimes</td>
<td>Rarely</td>
<td>Never</td>
</tr>
</tbody>
</table>

2. When you talk with this nurse, does she seem cool, disinterested, or indifferent?

<table>
<thead>
<tr>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Never seems cool</td>
<td>Almost never seems cool</td>
<td>Sometimes seems cool</td>
<td>Usually seems cool</td>
</tr>
</tbody>
</table>

3. When you question her about something that has gone wrong with her assignment, does she try to excuse why it is not her fault?

<table>
<thead>
<tr>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Often</td>
<td>Sometimes</td>
<td>Rarely</td>
<td>Never</td>
</tr>
<tr>
<td>excuses</td>
<td>excuses</td>
<td>excuses</td>
<td>excuses</td>
</tr>
</tbody>
</table>

4. When conditions of stress arise, would this nurse wait to be told what to do?

<table>
<thead>
<tr>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Would usually wait</td>
<td>Might wait</td>
<td>Probably would not wait</td>
<td>Never wait</td>
</tr>
</tbody>
</table>

5. Does she make suggestions which are useful?

<table>
<thead>
<tr>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very frequently makes useful suggestions</td>
<td>Frequently makes useful suggestions</td>
<td>Occasionally makes useful suggestions</td>
<td>Very rarely makes useful suggestions</td>
</tr>
</tbody>
</table>
6. Does this nurse let personal feelings develop toward the patients to the point where it interferes with her performance of duties?

<table>
<thead>
<tr>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Often</td>
<td>Occasionally</td>
<td>Rarely</td>
<td>Never</td>
</tr>
<tr>
<td>interferes</td>
<td>interferes</td>
<td>interferes</td>
<td>interferes</td>
</tr>
</tbody>
</table>
BIBLIOGRAPHY

Books


Articles


Kandler, H. M., "Changes in Empathy in Student Nurses During Psychiatric Affiliation," Nursing Research, 2 (Spring, 1953), 33-36.


Van Zelst, R. H., "Validation Evidence on The Empathy Test," Educational Psychology, 13 (June, 1953), 474-477.

Tests


Unpublished Material