

1962

## Explorations in the Measurement of Empathy As Analytical Predictive Ability

Calvin J. Daane  
*Arizona State University*

Gordon J. Rhum  
*State College of Iowa*

Copyright ©1962 Iowa Academy of Science, Inc.

Follow this and additional works at: <https://scholarworks.uni.edu/pias>

---

### Recommended Citation

Daane, Calvin J. and Rhum, Gordon J. (1962) "Explorations in the Measurement of Empathy As Analytical Predictive Ability," *Proceedings of the Iowa Academy of Science*, 69(1), 502-512.

Available at: <https://scholarworks.uni.edu/pias/vol69/iss1/79>

This Research is brought to you for free and open access by the Iowa Academy of Science at UNI ScholarWorks. It has been accepted for inclusion in Proceedings of the Iowa Academy of Science by an authorized editor of UNI ScholarWorks. For more information, please contact [scholarworks@uni.edu](mailto:scholarworks@uni.edu).

with Kimble's hypothesis, the M group developed SIR (permanent work decrement), but the D groups did not.

#### ACKNOWLEDGMENT

The writer is indebted to Professor Don Lewis under whose direction the research was conducted.

#### Literature Cited

- Hull, C. L. *Principles of Behavior*, New York: D. Appleton-Century, 1943.
- Kimble, G. A. An Experimental Test of a Two-Factor Theory of Inhibition. *J. exp. Psychol.*, 1949, 39, 15-23.
- Kimble, G. A., & Shatel, R. B. The Relationship Between Two Kinds of Inhibition and the Amount of Practice. *J. exp. Psychol.*, 1952, 44, 355-359.
- Lewis D. The Multipoint Two-Hand Coordinator: Apparatus for Studying the Acquisition and Transfer of Skill in Performing Subject-Paced Tasks. *Iowa Acad. of Sci.*, 1957, 64, 496-503.
- Lewis, D. *Quantitative Methods in Psychology*, New York: McGraw-Hill, 1960.
- Reyna, L. J. An Experimental Study of Work Inhibition in Motor Learning. Unpublished M. A. Thesis, State University of Iowa, 1941.
- Schucker, R. E., Stevens, L. B., & Ellis, D. S. A Retest for Conditioned Inhibition in the Alphabet-Printing Task. *J. exp. Psychol.*, 1953, 46, 97-102.
- Seibel, R. The Relative Permanence of the Decrement in Motor Performance After Practice Under Massed Conditions. *Iowa Acad. of Sci.*, 1955, 62, 457-461.

## Explorations in the Measurement of Empathy As Analytical Predictive Ability

CALVIN J. DAANE<sup>1</sup> AND GORDON J. RHUM<sup>2</sup>

*Abstract.* This was a preliminary effort to develop a scale for the measurement of empathy as analytical predictive ability. The approach embodied the use of 20 words in three contexts; the 60 rating elements comprised the scale. Extent of agreement between self ratings of a person to be judged (reader) and the estimates of those ratings by a group of persons (judges) was used as a measure of empathic ability. An attempt to take into account possible spurious influences on the empathy scores was not fruitful. The standard materials, including a filmed version of the reader's behavior, were used with several samples. Results showed generally better than chance empathy scores. There was some intimation that this type of scale is useful in identifying students for a school of nursing. Other between-group differences are suggested. The findings for a group of student teachers were not revealing.

The word empathy was first used in the field of fine arts and defined in terms of the identification and recognition of shared

<sup>1</sup> Arizona State University.

<sup>2</sup> State College of Iowa.

experiences. A listener might have empathized, for example, by associating the feelings represented by an actor or musical composition to a previous experience of the listener.

Empathy is a word translated from the German "einführung" which was first employed by the German psychologist, Theodor Lipps, in 1903. The literal translation is "feeling in." Garritt (1931) states that according to the Lipps definition "the distinction between the self and the object disappears or rather does not exist." Remmers (1954) also gives credit to Lipps and the term "einführung," but notes that men of literature, art, music, and the theater have throughout the centuries given expression to the ideas underlying this concept.

Rogers (1951) used the term empathy in describing a certain kind of relationship valued during counseling. He was labelling a "feeling of understanding" which the client might perceive from the counselor. A counselor with minimal psychological "blocks" might sense a feeling or attitude in the client and convey in return a feeling of understanding. The feeling empathized may or may not be one which the counselor associates with his own previous experience. When it is conceived as a self-feeling previously held, true understanding of the client's feeling might be difficult to convey. The counselor would be tempted to engage in a self-projection rather than a reflection of the way the client feels. The counselor might say in effect: "I recall now how I felt when *I* experienced this" instead of the desired "I understand how *you* feel when you say this."

The Rogers concept involved an assumption concerning intellectual control over feelings attributed to past experiences—the possibility of cue perception and conveyance of an understanding of how the other person had himself organized the cues without distortion through self-identification or projection. Several questions were apparent and important.

Does this kind of human ability exist? If so, do different persons possess it in varying degrees? Is it measurable? What kind of background of experiences would enhance or detract from the facility? Is the ability something distinct from general intellectual ability? To what extent is it a function of the present environment or the facility of other persons involved in the relationship? Does the facility involve extrasensory perception, or merely the expert freeing of sensory potentialities already widely known?

Since 1955 the investigators have been in touch with the available writings on the empathy concepts and have not been able to find reasonably satisfactory answers to any of the above questions, although hundreds of studies have been conducted.

Variations in subjects, methods, and definitions have contributed to the general state of uncertainty concerning the nature of empathic ability and its functioning, if indeed such a measurable ability exists over and above what is expressed through identification and projection.

The idea of the existence of empathy in some form seems a plausible one. For example, employers often insist on personal interviews with applicants for positions. The research by Thompson and Bradway (1950) showed better than chance predictions of the self-feelings of others even when numbers rather than words were used. And although a review of the various studies does not indicate the use of a common definition of empathy, certain areas of agreement emerge.

(1) Most investigators agree that at least one form of empathy can be thought of as predictive ability, a process whereby one predicts the self-feelings and attitudes of another. Matching of the predictions with the testimony of the person being judged will provide a kind of rough estimate of the subject's (judge's) predictive empathic ability.

(2) It is believed that the prediction of another person is involved with the areas of identification, projection, and stereotyping. When words are used as the basis of a rating scale two additional variables must be taken into account: the degree of similarity in use of definitions, and similarity in preference for either the middle or extreme positions on a continuum scale.

(3) Prediction scores without regard to the spurious influences will almost always constitute a distribution, and the goodness of the predictions will not be accounted for by the operation of chance.

(4) It is probable that two separate kinds of predictive ability exist. One has to do with making judgments about groups and their expected behavior, or of the behavior of individuals as members of a group. This is referred to as mass empathy, and the ability can be exhibited without immediate personal communication with the group or individual. This ability may be related to the judge's sensitivity to accurate stereotyping. It may in turn be highly related to the judge's general mental ability (IQ) and the recall of previous accurate perceptions. High level performance would involve two stages: perception of cues and the retention of the cues perceived. The second kind of predictive ability is the gaining of verbal and non-verbal cues through face to face meetings; this is called analytical empathy. Analytical empathic ability refers to the predictive behavior of one individual as he meets with one other individual. The kind of predictive ability which is concerned with the making of

judgments about groups is dependent upon all the variables involved in retention, whereas the predictive ability utilized in making judgments about *individuals* depends upon immediate cue perception and organization. Bell and Hall (1954) found Kerr's stereotype prediction tests (mass empathy) and Dymond's face to face prediction test (analytical empathy) uncorrelated for 87 subjects.

In seeking to construct an instrument for reliably estimating the degree of predictive ability exhibited during face to face meetings, the writers have used four assumptions.

- (1) A part of one's true analytical empathic ability may be noted through a verbal predictive process. Following a face to face meeting, predictive statements and self statements are compared.
- (2) A part of one's true analytical empathic ability may be observed in terms of the facility to perceive audio-visual cues accurately. Other forms of perception may be involved, but cue perception will be discernible.
- (3) A part of one's true analytical empathic ability will survive deletions from the basic predictive score of reasonable estimates of the influence of certain spurious factors—projection, identification, stereotyping, verbal semantics, and response set.
- (4) The refined scores for a specified population will constitute an essentially normal distribution, and the scores will be reliable.

#### A SUMMARY OF PROCEDURES

The primary effort was in devising an instrument and administrative procedure which would yield refined analytical empathy scores for large samples. All other considerations were secondary, and any correlating of the observed scores to other variables was done incidentally. The work was conducted mainly in a college setting, in four phases: (1) Construction of the prediction scale. (2) Development of a standard process for presenting the individual who is to be predicted. (3) Development of a usable formula for transforming the "rough" predictive score into a refined score. (4) Administration of the instrument to various groups in order to obtain estimates of within-group and between-group variability.

#### *The Prediction Scale*

The first task was to construct, as a part of the prediction scale, a usable scale for the recording of self concepts. For this purpose we started with Allport's (1936) list of 3,000 stimulus words. Three hundred words were selected at the outset. A group of individuals were requested to indicate the degree of meaningfulness a given word held for them. For example, the individual would be asked with reference to a certain word, "To what degree does the word describe you?" Ratings were supplied on a four-category scale (1 low, 4 high). One hundred words with relatively high ratings were selected as the basis for a self-concept scale.

The actual development of the self-concept scale may be understood through a consideration of the stimulus word *mature*. The respondent was to think of the word mature as it applied to himself in his behavior among family and friends: often, not sure but probably often, not sure but probably seldom, or seldom. The respondent was also required to think of the word mature as it applied to his behavior among acquaintances, and if he could attain his ideal. This is illustrated below.

		Often	<u>Not sure, but probably</u> Often                      Seldom	Seldom
MATURE	With family and friends			
	Among acquaintances			
	If I could attain the ideal			

The judge reacted on three bases to each word in each of the three contexts: among family and friends, among acquaintances, and if he could attain his ideal. The first basis was as each word applied to the judge's self concept. A second basis was for a hypothetical, average college student. The third basis was in terms of how the judge felt the person appearing before the judge would rate himself. The self rating (first basis) and the stereotype or blind rating (second basis) were used in the refinement of the prediction score (third basis).

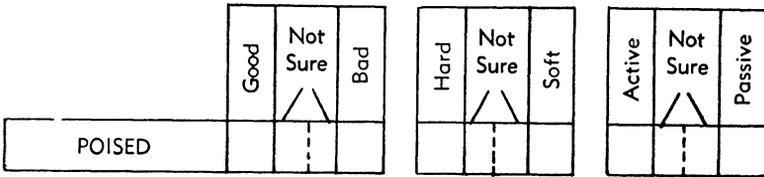
Various combinations of 40 stimulus words from the working supply of 100 words were used in preliminary tryout. In terms of this experience 20 words were selected for final tryout. Two main criteria were employed in identifying the most desirable 20 words from the 100 words which were investigated: (1) Good prediction scores for the judges, i.e., a relatively small difference between what a judge estimated for an individual and what the individual reported about himself. (2) Better prediction scores by judges after actually seeing the "predicted person" than the scores resulting from the prediction of a hypothetical, average college student.

The 20 words selected were:

Bashful	Eager	Humble	Moody
Conceited	Gentle	Jealous	Poised
Constructive	Gracious	Lazy	Reckless
Courteous	Gullible	Loyal	Restless
Cultured	Hesitant	Mature	Worthy

One additional factor, semantics, was considered. Both the judge and the person being judged indicated how they were defining each of the stimulus words. The defining was done through Osgood's (1957) semantic differential approach. One reaction was the choice of one of four possible sections along a

good-bad continuum. Secondly, there was a choice of one of four possible categories in the hard-soft realm. The other four-way classification was for active-passive. Osgood regards these breakdowns as related to evaluative, potency, and activity considerations, respectively. The form employed is illustrated below.



*Presentation of the Person to be Judged*

The person to be judged is referred to as the reader, and it was necessary to arrange for the presentation by a procedure which would be standard, practical, and in keeping with the operational definition of empathy. A number of approaches were tried. A reader was given an unfinished story and asked to tell how he would complete it. Again, a reader was handed some unfinished sentences which he was to complete. In a third approach the reader spoke extemporaneously on questions or topics offered by a second person seated with him at a desk. The other arrangement, and the one which ultimately served as a basis for our work, involved the reading of a short story followed by comments about it by the reader.

In each of the four approaches a mature college student served as a reader before a group of college students who did not know him. After the presentation the reader left the room to fill out the appropriate scales in terms of his self concept. The students (judges) who had viewed and heard the presentation were asked to predict the reader's response to each of the elements in the scales. The four different types of presentation were tried with various readers and with judges (college students) from a variety of academic major fields. Decisions as to which reader and which type of presentation would be best to employ depended upon the mean and standard deviation of obtained distributions of ratings. "Goodness of prediction score" accompanied by appropriate score variance were determining factors.

A college student of junior classification, male, speech major, was selected as the reader. He presented the short story "Quality" by John Galsworthy; about 10 minutes were required for the reading of the story and a similar amount of time for the reader to explain what the story meant to him and what he thought the author had attempted to convey through his characters. The reader's performance was placed on sound film in black and

white, and this filmed representation of the reader was used in the collection of data. It was believed that the film arrangement possessed two necessary characteristics: (1) It permitted a uniform presentation to various groups of individuals from whom the necessary quantity and kinds of data could be secured. (2) It was consistent with the operational definition of empathy as the ability to predict self feelings and attitudes through perception of audio-visual cues.

#### *Development of Scoring Formulas*

A key step in investigating the ability of individuals (judges) to predict the self feelings and attitudes of another person (reader) was to examine the extent of agreement between the manner in which the reader and each of the judges filled out the 20-word scale—selecting for each word one of four positions along the Often-Seldom continuum under each of three situations: with family and friends, with acquaintances, and if the ideal could be attained. Judges' estimates of the reader's self ratings were consistently better than chance estimates. An important question arises as to what could possibly account for such better-than-chance prediction scores other than the predictive ability under investigation. Five possible factors were identified:

1. Similarity between judge and reader in definition and use of the words.
2. The judge's projection of himself under the circumstance of similarity between judge and reader in self concept.
3. Assignment of stereotype ratings by the judge in an instance in which the reader happens to be somewhat like the stereotype.
4. Similarity between judge and reader in response set as related to a forced choice scale with four categories.
5. The given reader's ability to expose or reveal himself in such a way that his self concept might be perceived relatively easily.

The fifth of the above five factors is the possibility that for a given group of judges the predictions of one reader might be quite accurate, whereas the predictions of another reader could be relatively inaccurate. A variation is that some judges may make better predictions of one reader than of another. This contingency was investigated in the early stages of our efforts by using four readers and identical materials (live presentation of a short story, and comments) with several groups of college students enrolled in child psychology courses. The evidence suggests that there may be appreciable, if not substantial, variability among readers. The single reader utilized in the filmed presentation was one of the four who participated in the tryout stage; he was thought to be somewhat typical in predictability. However, this factor merits further investigation.

The first four of the five factors received careful attention. Immediately after the presentation of the reader via film to a group of judges, each judge completed the 20-word empathy

scale by estimating the way in which the reader would complete it. Each word was considered in three contexts, resulting in 60 rating elements. The reader's self ratings constituted a scoring key. For each of the 60 rating elements the divergence in response positions of the judge's estimate from the reader's self rating was determined; the sum of the 60 divergences was the *raw prediction score*. A small score value would be indicative of agreement or of good prediction. For the scoring key actually employed, the maximum possible sum of divergences was 164; a chance sum of divergences was 82.

Attempted corrections or refinements of the raw prediction score were based on four indices, used singly and in combination and with varying weights.

1. *Similarity of definition* of the 20 words by reader and judge.
2. *Similarity of self concept* of reader and judge.
3. Similarity between reader's self rating and judge's estimate of some hypothetical, average reader—*stereotype score*.
4. *Similarity in response set* of reader and judge.

For the first three of these four indices the judge's responses were obtained just prior to the presentation of the reader.

There was an armchair analysis of how the four indices might be used in an attempt to refine the raw prediction score. A total of 43 different scoring formulas were proposed, one of which yielded the raw prediction score. For each judge in the samples employed an empathic ability rating was computed by means of each of the formulas. It was anticipated that this procedure would provide insight as to appropriate scoring techniques for the scale under construction.

## RESULTS AND DISCUSSION

Preliminary tryout of materials and procedures had been conducted with hundreds of college sophomores enrolled in various sections of a child psychology course and with small groups of graduate students who were taking guidance courses. The phases of the project reported in this section were also exploratory in nature, although they represent the latest stages of the work to date.

### *Reliability of the Scale*

After the specific words for inclusion in the scale and the method of presentation had been decided upon, the scale was administered to the 50 members of a college class in order to secure an estimate of reliability. For this sample of sophomore psychology students the coefficient of reliability by the test-retest method was .74. The repeat testing was done in ten days after the first testing. Judges' ratings were expressed in raw prediction score form.

*A Sample of Student Teachers*

The 50 student teachers of one term at a campus laboratory school constituted one sample of judges. There were 21 at the secondary school level, and 29 in the elementary division. The empathy scale was administered to the group; scores were derived by application of each of the 43 formulas. The investigators arranged for an independent rating of each student teacher by the faculty member who was serving as supervisor. The ten statements employed are presumably related to factors in the empathy realm, e.g., "Adapts quickly to pupil motivational changes." For each statement the supervisor selected a position along a line extending from *Often* through *Not Sure* to *Seldom*; the numerical equivalents ranged from 5 to 0. For a random sample of 30 of the 50 student teachers product-moment correlation coefficients were computed for the ratings by the student teaching supervisors vs. each of the 43 empathy test ratings. The coefficients were uniformly low. An  $r$  of .16 was the largest one obtained; it resulted from the raw prediction score in empathy.

It is recognized that these independent ratings may be characterized by substantial fluctuation among raters in ability to rate, in functional basis of rating, and in stringency of rating. The presence of such influences would reduce the meaningfulness of the set of ratings.

*A Sample of Student Nurses*

Data were secured for 30 of the 38 student nurses who comprised the freshman class at a school of nursing. In addition to the empathy scale rating, three variables were involved.

1. American Council on Education Psychological Examination, Total Score. Percentile rank among normative population of nursing education applicants. Median was 72nd percentile.
2. Cumulative grade point index for first year of study at the school of nursing. Median was 2.34.
3. Composite of ranking by each of four members of nursing school faculty in terms of over-all success in nursing preparation. Each of the four faculty members identified the top 5 students, the second 5 students, the middle 10 students, the next 5 students, and the lowest 5 students.

Following are representative correlation coefficients obtained for each of the variables when paired with the empathy test scores:

ACE Psychological Examination	.20
Grade Point Index	.18
Over-all rating by nursing staff	.62

The raw prediction score series produced the highest correlations except in the case of grade point index; in this instance the  $r$  involving raw prediction scores was .16. The relationship between empathy scores and ratings by nursing school faculty members,  $r$  of .62, is sufficiently strong to warrant follow-up investigation.

*Other Student and Occupational Samples*

The empathy test was used with several additional small groups. Certain comparative data are thus available.

Group	N
Student teachers .....	30
Student nurses .....	30
College sophomores .....	50
Ministers .....	10
Business college students:	
Stenographic .....	10
Secretarial .....	10
Accounting .....	10
Business administration .....	10

It is difficult to interpret these small sample means; however, all eight of the samples appeared to achieve relatively "good" raw prediction empathy scores. They were generally in the 30's under an arrangement in which a possible range was 0 to 164 where the score referred to a degree of predictive divergence. Standard between-group differences are noted in means and standard deviations. Further inquiry is warranted in this regard.

*Concluding Statement*

Exploratory work in developing an instrument for measuring empathy as analytical predictive ability has indicated at least five guides for subsequent efforts. One is that there be more investigation of the magnitude of the variability among readers in respects important to empathy measurement. A second observation is that if useful corrections of raw prediction score are to be sought, new avenues should be followed in taking into account similarity between reader and judge in definitions, in self ratings, in response set, and in stereotyping. A third indication is that the empathy concept may be of substantial importance in the selection of students for nursing education. A fourth direction is toward the securing of meaningful independent assessments of student teachers and regular teachers as a basis for a kind of validity check. An extension of the tryout of materials and techniques with sizable groups from various occupational fields is a fifth sphere of activity which might be fruitful.

**Literature Cited**

Allport, Gordon, and Odbert, Henry, "Trait Names, a Psycho-lexical Study," Psychological Monographs, 47: 1-171, 1936.  
 Bell, G. E., and Hall, H. E., Jr., "The Relationship Between Leadership and Empathy," Journal of Abnormal and Social Psychology 49: 156-157, Jan. 1954.  
 Garritt, E. F., *Philosophies of Beauty*, New York: Oxford University Press, 1931, p. 253, Citing Theodor Lipps, *Einfuhlung, Innere Nachahmung Und Organernpfindungen* (Archi fur die gesamte psychologie, i, 1903.)  
 Osgood, C. E., Suci, G. J., and Tannenbaum, P. H., *The Measurement of*

- Meaning*, Urbana: University of Illinois Press, 1957.
- Remmers, H. H., *An Introduction to Opinion and Attitude Measurement*, New York: Harper and Bros, 1954, p. 234.
- Rogers, Carl R., *Client-Centered Therapy*, Boston: Houghton Mifflin Co., 1951, p. 29.
- Thompson, C. W., and Bradway, K., "A Content-Free Method of Training Interviewers," *Journal of Consulting Psychology* 14: 321-323, August, 1950.