The PICA Scoring Scale: Do Its Statistical Shortcomings Cause Clinical Problems?

Joseph R. Duffy
Barbara J. Dale

University of Massachusetts, Amherst

Most aphasiologists are aware that the Porch Index of Communicative Ability (PICA) employs a 16-point multidimensional scoring system which is a validated ordinal scale for assessing disordered communication (Porch, 1967). The ordinal nature of the PICA scale indicates to its users that a score of 10 is better than a score of 5, that a score of 15 is better than a 14, that a score of 7 is better than a score of 6, and so on. Strictly speaking, or statistically speaking, the existence of an ordinal scale does not allow us to assume that the behavior scored as a 10 is twice as good as the behavior scored as a 5, or that the behavioral distance between a 15 and a 14 is the same as the behavioral distance between a 7 and a 6. In other words, with an ordinal scale we do not know if the intervals between all adjacent points on the scale are equal. If these intervals were shown to be equal we would have what is known as an interval scale.

Even though the PICA scale has not been shown to be an interval scale, the routine clinical use of the test employs statistics which require an assumption of intervality. For example, we add item scores and divide to obtain subtest means, and we add and divide subtest means to obtain modality and overall scores. We also use mean scores to predict recovery levels, predictions which are based on regression equations which are also derived from statistics requiring interval levels of measurement.

These apparent statistical violations have not gone unnoticed in the literature. McNeil, Prescott, and Chang (1973) have questioned the statistical validity of using interval level statistics with the PICA. Silverman (1974) critized the use of mean scores with the PICA partly because of the scale's unproven intervality. In response, Porch (1974) excused the possible statistical violation because of the distinct clinical and research advantages provided by more powerful interval level statistics. VanDemark (1974) also defended the use of interval level statistics but cautioned against using mean scores as an index of specific behavior.

Because of the debate over the statistical uses to which PICA results may be put, the current investigation asked two questions which were answered in separate but related experiments. Since no one has really shown that the PICA scale is not an interval scale, the first experiment sought to determine the degree to which the PICA multidimensional scoring scale can be considered as an interval level measure. The second and more important and relevant experiment from a clinical standpoint was designed to determine if currently obtained PICA results would be equivalent to results obtained from the use of an interval level scoring system. A lack of equivalency would suggest that more conservative statistics or a new numerical scale be adopted for the PICA, while evidence for equivalency would strongly support the continued use of interval level statistics with the original scoring scale.

Experiment I

Procedure

In order to determine the degree to which the present PICA scale approximates an interval scale it was necessary to develop an interval scale for the 16 PICA categories. This was accomplished through a pair comparison experiment which followed the experimental and statistical procedures as outlined by Guilford (1954).

Twenty-one graduate students with a basic knowledge of aphasia but no exposure to the PICA were used as judges for the experiment. The written name and description of each of the 16 PICA categories were used as stimuli to be compared by the judges. These were identical to those originally used by Porch (1971, p. 784) to validate the ordering of the PICA scale. The judges' task was to select the more adequate clinical behavior from the 120 pairs representing all possible pairings of the written descriptions of the 16 categories. The 120 pairs were compared in an order which minimized time and space errors (Ross, 1934), and fatigue and/or practice effects were reduced by reversing the order of comparisons for half of the judges. Following the pair comparisons, the data were subjected to statistical procedures (Guilford, 1954) appropriate for the development of an approximation of interval values for the 16 categories.

Results

The newly developed interval values and the original PICA values for each of the 16 categories are shown in Table 1. Inspection of the PICA scale and the interval scale indicates that the PICA is not precisely interval in nature and in several instances the ordering of its categories does not hold up. Without spending an undue amount of time analyzing the differences between the two scales, the following summary observations can be made:

1. The dimension of Accuracy is clearly distinguished in the interval scale in that all inaccurate responses (or scores 8 on the PICA scale) fall below accurate responses. The reversal of the "intelligible" and "error" categories, in our view, may be due to the more positive connotations associated with the word "intelligible" when compared to the word "error". Also, the numerical value of almost all inaccurate categories are below those that are used for the PICA scale and the intervals between each adjacent category are not equal.

2. Several reversals of categories were found for accurate responses (or PICA scores above 7). A number of these, we feel, can be explained by an overall reversal in judges' assessment of the dimensions of Responsiveness and Completeness. That is, the "corrected", "repeated", and "cued" categories, all of which reflect reduced Responsiveness, were all rated above the two Incomplete categories which reflect inadequacies within the dimension of Completeness. This finding is particularly interesting because it is in agreement with those of McNeil and Prescott (1975) who assessed the PICA's ordinality from judgements of actual observed behaviors and found that "the dimension of responsiveness was often interchanged with the dimension of completeness" (p.117).

3. As was true for the Inaccurate categories, the intervals between each adjacent category for the Accurate responses were not equal, with some adjacent

categories being separated very narrowly and others showing separations well exceeding one unit.

4. Finally, in spite of differences in ordering between the PICA and interval scale and the lack of accurate intervality in the PICA scale, the two scales are highly correlated with each other, with Pearson r and Spearman rank order correlations both exceeding .90. Therefore, while the PICA and interval scale are obviously not identical, they do have a great deal in common. With all of these results in mind the most obvious and logical question to be asked next, then, is "does the difference between the two scales make a difference as far as the clinical, statistical uses to which the PICA scale is generally put?"

Experiment II

Procedure

In order to determine if currently obtained PICA results would be equivalent to results obtained with the interval level scoring system, it was necessary to score a number of PICAs using both scoring scales and then determine if results were similar enough to be considered equivalent.

PICA results for 50 aphasic patients were randomly selected from over 200 Mayo Clinic patients who were serving as subjects for another study involving the PICA. The clinical characteristics of these patients are shown in Table 2. All selected patients were diagnosed as aphasic and had neurological evidence of a unilateral left hemisphere lesion. While age, education, time post onset, etiology and sex distribution were not important variables in this study, the data presented in Table 2 indicates that the sample was fairly representative of the range of values usually exhibited by the aphasic population for those variables.

It was important that the patients selected exhibit a range of PICA scores in order to allow for generalizations about the equivalency of the PICA and interval scales across the range of severity of aphasia. The mean overall PICA score of the sample selected was at the 45th percentile. Also, scores ranged from the 3rd to the 90th percentile and each decile contained at least three patients. This distribution was considered quite adequate for purposes of this study.

Following patient selection, the 180 responses from each patient's PICA (18 subtests x 10 items per subtest) were rescored using the newly developed interval values. Results obtained with the PICA scale and the interval scale were then compared.

Results

The results of comparisons between PICA summary scores obtained with the two numerical scales are summarized in Table 3. It can be seen that the mean scores for each scale are quite similar for the Overall, Gestural, Verbal, and Graphic modality summary scores with the interval scale yielding slightly higher scores for all measures. These small mean differences were statistically significant for all but the Graphic modality. These differences, however, should not be construed as indicators of lack of equivalency, as it is the correlation between measures which is the more crucial indicator of the degree to which the results of one test can be used to predict the results of another test (Giolas & Duffy, 1973). In other words, if the results obtained using

PICA's ordinal scale are highly correlated with the results obtained with the interval scale, then they may be considered as equivalent numerical methods of quantifying communicative behavior. As can be seen, the correlations between the two scoring scales are extremely high, exceeding .99 in all instances. Using a formula developed by Gulliksen (1950) we asked if the obtained correlations were high enough to allow results from the two scales to predict each other within five-tenths of a point. This represents a margin of error of less than 3.5 percent which is a very stringent level of precision for a clinical measure of this type. All four of the correlations shown equaled or exceeded the correlation necessary for that degree of precision and for that reason the PICA summary score results obtained with the PICA scoring scale can be considered equivalent to those which would be obtained using an interval level scale.

Table 4 shows the results of comparisons between PICA subtest scores obtained with the two numerical scales. Once again, the mean scores between the two scales are quite similar, with the interval scale usually yielding slightly higher values. While small, the differences between the two scales were statistically significant for 13 of the 18 subtests. Most important, however, all correlations are again exceedingly high with all 18 correlations surpassing the required correlation necessary for equivalency within less than 3.5 percent error or five-tenths of a point. It was, therefore, concluded that equivalency between the PICA and interval scales exists for the 18 PICA subtests as well as for PICA summary scores.

Discussion And Summary

The answer to the first question posed by this study indicated that the PICA is not an interval level scale although it is highly correlated with the interval scale which we developed. We also found that there were some changes in the ordering of categories between the two scales, the most important of which was a reversal of categories representing the dimensions of Responsiveness and Completeness. We do not recommend a change in the ordering of these dimensions on the basis of the findings of this study. However, considering the similar findings of McNeil and Prescott (1975), we feel that the categories representing these dimensions deserve further clinical evaluation regarding the validity of their current ordering on a scale of communicative adequacy.

The second question posed by this study asked if the lack of intervality for the PICA scale results in different scores from those that would be obtained if an interval level scale were used. The extremely high correlations between the PICA and interval scales obtained for PICA subtest and summary scores met stringent criteria levels for equivalency. This evidence for equivalency basically allows us to say that the difference between the PICA scale and an interval scale does not make a difference as far as PICA numerical results are concerned, and that there are no strong advantages to be gained by altering the current numbers used in the PICA scale to correspond to interval scale values. We feel that these results give concrete, clinically relevant evidence is support of the continued use of interval level statistics with the PICA for purposes such as intra and intersubject comparisons, subtest comparisons, making prognostic statements, and evaluation the effectiveness of therapy.

Acknowledgement

This investigation was partially supported by Training Grant NS - 05446 from the National Institutes of Health, Public Health Service.

References

Giolas, T.G. and Duffy, J.R. Equivalency of CID and Revised CID Sentence Lists. JSHR, 16, 4, 549-555 (1973).

Guilford, J.P. Psychometric Methods. New York: McGraw-Hill, 2nd edition (1954).

Guilliksen, H. Theory of Mental Tests. New York: Wiley (1950).

McNeil, M.R., Prescott, T.E. and Chang, E.C. A measure of PICA ordinality. In Clinical Aphasiology Conference Proceedings 1975. R.H. Brookshire (ed.) Minneapolis: BRK Publishers (1975).

Porch, B.E. Porch Index of Communicative Ability: Theory and Development, Vol. I. Palo Alto: Consulting Psychologists Press (1967).

Porch, B.E. Multidimensional scoring in aphasia testing. JSHR, 14, 4, 776-792, (1971).

Porch, B.E. Comments on Silvermann's "psychometric problem". JSHD, 39, 2, 226-227 (1974).

Ross, R.T. Optimal orders for the presentation of pairs in the method of pair comparisons. Journal of Educational Psychology, 25, 375-382 (1934).

Silverman, F.H. The Porch Index of Communicative Ability (PICA): A psychometric problem and its solution. JSHD, 39, 2, 225-226 (1974).

Van Demark, A.A. Comment on PICA interpretation. JSHD, 39. 4, 510-511 (1974).

Discussion

Q. What kind of information did you give your judges?

A. They were merely told to indicate which of the two behaviors in the pair was clinically more adequate. They were given no additional information.
Comment: Dr. Porch presented a lengthy comment on the difficulty judges have in rating the "goodness" of responses. He ended by saying...so I am really surprised that the judging came out so beautifully in this study.

A. I think that you make a very valid point. The only additional comment that I can make is that we have also conducted an experiment where we had judges look at behavior and not merely read descriptions of behavior and the reversal of the dimensions of completeness and responsiveness also occurred, but not quite to the same degree as it did in this experiment.
Comment: It does seem to me that maybe your study says something about the real world and how behavior is perceived by non-successful speakers. I had a patient just recently tell me at a very high level - I asked him what bothered him and he said, "The only thing that bothers me is I can't get it out faster". Apparently he is reflecting that he feels some penalty for that delay in his communicative efforts on the outside or maybe there is some validity for that discrepancy in terms of how non-professionals judge the adequacy of communication.
Comment: I think one problem with the scale is that there are a wide range of behaviors which can occur within each of the categories. You can have a delay of 4 sec. or a delay of 30 sec. I would take a delay of 4 sec. over an incomplete response, but I am not sure that I would take a delay of 30 sec. over an incomplete response.

Table 1. Comparison Of PICA Scoring Categories And Values With Interval Scale Ordering And Values.

PICA Value	PICA Category	Interval Category	Interval Value
16	complete	complete	16
15	complete	complex	15.77
14	distorted	corrected	14.53
13	delayed	delayed	13.65
12	incomplete	distorted	11.93
11	inc.-delayed	repeated	11.71
10	corrected	cued	11.23
9	repeated	incomplete	9.85
8	cued	inc.-delayed	8.49
7	related	related	7.79
6	error	intelligible	4.79
5	intelligible	error	4.41
4	unintelligible	unintelligible	3.61
3	minimal	minimal	2.53
2	attention	attention	1.89
1	no response	no response	0.49

NOTE: Correlations between PICA and Interval scale: Pearson r = .9244
Spearman = .9265

Table 2. Descriptive Data For 50 Randomly Selected Aphasics Whose PICA's Were Rescored Using An Interval Level Scale.

	Age	Education	Time Post Onset (wks)	Etiology	Sex
N	50	35*	46*	CVA=33	male=29
$\bar{X}$	54.10	13.00	12.89	Post OP=15	female=21
s.d.	16.40	4.42	24.75	Trauma=1	
				Tumor =1	
Range	18-87	8-20	1-124		

*Remaining data unknown.

Table 3. Comparisons And Correlations Between PICA And Interval Scale Summary Scores For 50 Aphasic Patients

Summary Score	Pica $\bar{X}$	Interval $\bar{X}$	$\bar{X}$ Difference	Correlation (Pearson r)
Overall	10.13	10.41	-.274*	.9995
Gestural	11.45	11.86	-.417*	.9993
Verbal	10.04	10.33	-.286*	.9991
Graphic	8.44	8.51	-.077	.9985

*p < .01 (t test).

Table 4. Comparisons And Correlations Between PICA And Interval Scale Subtest Scores For 50 Aphasic Patients

Subtest	Pica $\bar{X}$	Interval $\bar{X}$	$\bar{X}$ Difference	Correlation
I	8.09	8.17	-.08*	.9974
II	9.02	9.32	-.31*	.9977
III	10.27	10.65	-.38*	.9982
IV	9.65	9.88	-.23*	.9989
V	9.62	9.74	-.12*	.9970
VI	10.85	11.23	-.38*	.9985
VII	10.70	10.95	-.24*	.9969
VIII	13.84	14.49	-.64*	.9989
IX	10.12	10.41	-.29*	.9986
X	12.84	13.37	-.53*	.9979
XI	14.42	15.15	-.73*	.9995
XII	12.31	12.86	-.54*	.9994
A	5.77	5.51	.26*	.9903
B	7.27	7.14	.13	.9978
C	7.63	7.60	.03	.9977
D	7.98	8.01	-.02*	.9983
E	10.21	10.62	-.41*	.9982
F	11.78	12.22	-.44*	.9986

*p < .01 (t test).