DOCUMENT RESUME

ED 051 300 TM 000 621

AUTHOR	Gardner, Eric F.
TITLE	Measurement in Education: Interpreting Achievement Profiles - Uses and Warnings.
INSTITUTION	National Council on Measurement in Education, East Lansing, Mich.
PUB DATE	Jan 70
NOTE	12p.; Special Report, v1 n2 1970
AVAILABLE FROM	National Council on Measurement in Education, Office of Evaluation Services, Michigan State University, East Lansing, Michigan 48823. $2.00 per year (4 issues); Single issues 0.25 each in quantities of 25 or more
EDRS PRICE	EDRS Price MF-$0.65 HC Not Available from EDRS.
DESCRIPTORS	*Achievement, Achievement Tests, Comparative Analysis, Correlation, *Norms, Performance, *Profile Evaluation, Raw Scores, *Scores, Statistical Analysis, Student Records, *Test Interpretation, Test Results

ABSTRACT

The interpretation of test profiles, graphic devices which indicate the overall performance of an individual or group of individuals, is discussed. Some uses of profiles are presented, together with a three step profile analysis procedure for interpreting test results. Because profiles are relatively simple to construct and appear to be easily interpreted, considerable emphasis is placed upon possible errors in profile use. (AG)

Volume 1, No. 2

January, 1970

measurement in education

A SERIES OF SPECIAL REPORTS OF THE NATIONAL COUNCIL ON MEASUREMENT IN EDUCATION

Interpreting Achievement Profiles--Uses and Warnings

Eric F. Gardner

ABOUT THIS REPORT

In the press of dealing with great quantities of information about large numbers of students, there is a natural inclination to seek ways to simplify and summarize. The test profile is valuable for this purpose because it can capture the essence of important relationships and present them in a manner which is immediately apparent. In that sense the profile has heuristic value which enhances understanding of scaled test scores.

Dr. Gardner emphasizes these virtues of the profile -- particularly as they apply to achievement tests. He takes special care, however, to warn against corollary dangers in using profiles. These stem mainly from oversimplifying and overinterpreting information which is presented in distilled, visual form. As this report makes clear, test scores still have the same measurement characteristics and limitations regardless of the mode of presentation. This article enumerates specific cautions as well as useful advice on the use of profiles.

The author is highly qualified to write on the subject. In his years as Professor and Chairman of the Psychology Department at Syracuse University, Dr. Gardner has been recognized as a national authority on measurement in education. As a co-author of the well-known Stanford Achievement Test, he has had unique experience in fostering good measurement practice. This report continues the author's tradition of putting that experience to good use.

WWW

ERIC F. GARDNER

The old Chinese saying that a picture is worth one thousand words is especially applicable to test profiles. Profiles are convenient ways of showing test scores; they are graphic devices enabling us to see the over-all performance of an individual or group of individuals at a glance. They provide an excellent means for gaining a comprehensive picture of a person's or class' strengths and weaknesses. Profiles can be very helpful provided we use suitable caution in their interpretation. In one sense, a profile is like a good map which reflects features existing in reality, however, the appearance of such features on the profile does not *guarantee* their reality. An important point to remember is that although many of us find it surprisingly easy to believe that a score must be accurate if we have seen it on a test profile, its appearance on the profile does not make the score any more or less accurate or valid.

In general, profiles are used when we wish to show two or more scores for the same person or two or more scores for groups of people. We may be interested in sets of scores obtained at the same time or sets of scores obtained at fixed intervals such as those of a student on a group of tests taken in successive grades.

Profiles show the tests along one axis of the graph and the score values along the other axis. Profile forms may show score values along the vertical axis or along the horizontal; there is no particular reason for preferring the one over the other so far as ease of reading is concerned. If, for example, we wish to prepare a profile sheet for our own class we would probably want to list our test variables down the left hand side of the sheet and to plot our scores along the horizontal axis of the profile. We would do this because it is easier to write the complete test identification along a line than to write it along the narrow confines of a column.

Since raw (or obtained scores) test scores may vary considerably in meaning, it is obvious that raw scores cannot be used in plotting a profile. Before a profile can be plotted, then, it is clearly necessary to transform the scores to sets of comparable values. There are two ways of doing this. One is to scale the raw scores on the profile itself so that each scale has an equivalent mean and equivalent units of measurement. The other is to convert the raw scores into some type of derived scores before plotting them. The most common method is to use either standard scores, percentile ranks scaled to proportional standard score distances, or stanines. Note that when this procedure has been followed, the standard scores, stanines, or percentile ranks must be based on the same or strictly comparable populations all of whom have been tested at the same time. A discussion of comparability and other warnings about the construction of profiles and their intepretation will be presented later on in this paper. Let us now consider and illustrate several different kinds of useful profiles.

SOME USES OF PROFILES

1. To Obtain a Picture of the Relative Performance of a Pupil in Several Different Subjects or Areas.

What should you look for in a profile? Is there a systematic way that you can analyze test results? The following three steps of analysis represent a good approach to the interpretation of test results. Diagnose, Evaluate, Plan. The analysis of the test profile in Figure 1 illustrates how these steps can be applied.[1]

Analysis of Susan K.'s Profile

Figure 1 is a sample copy of the Pupil Stanine Profile. The three steps of analysis are applied to this profile as an illustration of how test scores of a pupil can be meaningfully interpreted. At this point I would like to call attention to the statement in the Stanford Achievement Manual which says, "When comparing two subtest stanines for an individual pupil, only differences of 2 or more stanine levels should be considered significant by the teacher."

Diagnose – Examine the profile for the most obvious subject strengths and weaknesses shown by the pupil's performance on this test battery.

Susan, with plotted stanines of either 8 or 9, is achieving best in the areas of word meaning, paragraph meaning, spelling, word study skills, language, and the social studies. When compared with other girls and boys of her grade level, Susan

Stanford Achievement Test
Intermediate Complete Battery

Name	Date of Testing	Grade Placement	Age
K., Susan B.	2-26	4.6	9 yr. 6 mo.

Otis Quick-Scoring Mental Ability Test IQ 124, Stanine 8

	GRADE SCORE	%-ILE RANK	STANINE
Word Meaning	70	94	1 2 3 4 5 6 7 (8) 9
Paragraph Meaning	82	98	1 2 3 4 5 6 7 8 (9)
Spelling	76	96	1 2 3 4 5 6 7 8 (9)
Word Study Skills	71	90	1 2 3 4 5 6 7 (8) 9
Language	72	90	1 2 3 4 5 6 7 (8) 9
Arithmetic Computation	54	80	1 2 3 4 5 6 (7) 8 9
Arithmetic Concepts	40	34	1 2 3 (4) 5 6 7 8 9
Arithmetic Applications	49	62	1 2 3 4 5 (6) 7 8 9
Social Studies	77	98	1 2 3 4 5 6 7 8 (9)
Science	45	46	1 2 3 4 (5) 6 7 8 9

Figure 1. Stanine Profile for Susan K.

shows average achievement in arithmetic concepts and in science, where she has stanines of either 4, 5, or 6. She shows evidence of understanding the application of arithmetic with a stanine of 6 and shows considerable competence in arithmetic computation with a stanine of 7. With an I.Q. of 124 and a corresponding stanine of 8, Susan would normally be expected to achieve stanines of 7 or above in the various subjects.

Evaluate – Relate the pupil's scores on the achievement test to such variables as your estimate of the pupil, his grades, his performance on a test of mental ability, and the like.

Susan's test results indicate that she is a superior student in the language arts and in social studies. Her school marks and judgments of previous teachers should reflect this superiority. If the test was taken in the spring, have school marks through the school year reflected this superiority? If not, why not? What are Susan's personal attitudes? Is she a non-conformist? Does she excel in aspects of a subject not measured by the tests? Is Susan a highly verbal memorizer? Is she a poor reasoner in mathematics and science? What are her interests? Doesn't Susan need special encouragement and help in mathematics and science? These and other questions arise when test scores and other evaluations do not correspond.

[1] The following illustration has been taken from *Stanford Achievement Test, Teachers' Guide for Interpretation and Use of Test Results*, Harcourt, Brace & World, Inc., 1965.

Plan – Plan a program of classroom activities that will remedy some of the obvious shortcomings and will build upon the greater strengths of each pupil.

Diversity of interest in subject areas and of levels of achievement in them is inevitable and even desirable among pupils and within each individual pupil. But no pupil in the elementary grades, especially one of Susan's general level of ability, should fail to learn the fundamental subjects such as arithmetic.

Because of Susan's outstanding work in the language arts, her teacher can be reasonably assured that her inability to score much above average in the area of science or in arithmetic concepts does not stem from any reading difficulty. This then leaves the teacher at least two factors to consider: (1) lack of interest and (2) lack of fundamental knowledge about the underlying concepts in science and in arithmetic. These possible deficiences could have resulted from inadequate experience with these subjects, inadequate interest evidenced in the home in these content areas, lack of a stimulating teacher of these subjects, and the like.

The instructional problem here is a relatively simple one. A careful, thorough discussion with Susan should elicit from her the level of interest in these areas and also some reasons for a lack of understanding in the basic concepts. Her shortages in knowledge of mathematics concepts need to be diagnosed. As a result of such understanding, Susan's teacher will be able to build an instructional program that will improve Susan's performance in these areas.

2. To Compare the Performance of a Single Grade on Several Subjects and with the National Norm.

One common use of achievement test batteries is in connection with some phase of administration or supervision. The supervisor is interested in knowing strengths and weaknesses in specific subjects so that they can be given greater attention. Frequently the national performance is accepted as a standard. Although national norms are useful as one frame of reference, it is important to recognize that achievement at the national average may well be an unreasonable goal for a particular school, class, or system. Norms are not designed to be standards nor should they be so designated unless a consideration of all relevant variables indicates they represent an appropriate level of average achievement for a particular group of students. Even then, by the very definition of a norm, it is expected that half the pupils will exceed it and half will fall below.

The authors of most achievement test batteries provide several scales for comparing local achievement with national norms. We can usually expect to be furnished with grade equivalents, percentile ranks, standard scores, and stanines. In spite of their deficiencies and decrease in popularity, the grade equivalent is still the most commonly used frame of reference for evaluating local achievement.

In Figure 2 a single-grade profile is shown in which the deviations of the local school system medians from the corresponding national normative values are plotted in months of grade equivalent above or below the norm at the time of testing. This profile, which represents the performance of all fourth grade pupils tested the first of November (Grade 4.2) from a community of slightly better-than-average socio-economic level and moderate size, indicates that achievement is above the national norm in all areas.

We have to note, however, that the average I.Q. of this system was 110 on the *Pintner General Ability Test*, and the average age in this grade was three months younger than the national normative group. Hence, it is pertinent to ask, "Is this group exceeding the national norm as much as would be expected?" Many factors previously mentioned and others to be discussed later are at issue e.g.

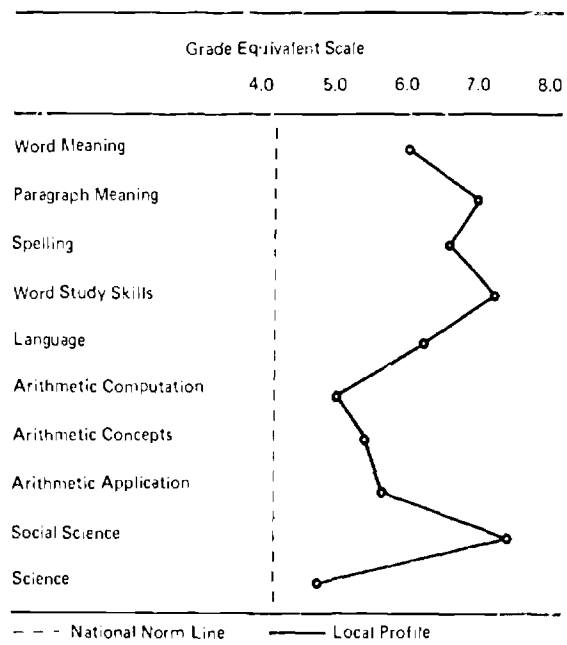


Figure 2. Profile of Fourth Grade Students (Tested November 1st) Plotted in Terms of Median Standard Scores Expressed as Grade Equivalents.

(1) comparability of grade equivalent units across subjects, (2) mental tests generally correlate differently with subject-matter achievement tests from area to area, (3) reliability of test scores.

3. School-by-School Comparisons in Terms of Achievement Tests.

The superintendant of schools, usually the person who eventually has to approve the purchase of test materials and the allocation of time for testing, is interested in knowing how his schools compare with each other and the national norm. As one important datum it is often helpful for him and the supervisor to have a school-by-school comparison based on standardized test results.

Figure 3 shows such a distribution of standard scores by school for one subject (Spelling) in a small school system.[2] Medians have been computed for each school and these median standard scores have been circled and joined in order to make a profile. Although the standard score scales used here lack some of the deficiencies of grade equivalent scales and although we are dealing with medians rather than individual scores, we still have the typical problems associated with determining how large an observed difference must be to be meaningful. Some of these differences are so small that they can be considered chance differences. Others are so substantial that they would undoubtedly maintain upon retesting. A similar profile could be made for each class within a specific grade.

It is desirable for a school system to carry out such a testing program for several years using alternate forms of the same batteries. By relating this kind of achievement test information to other factors such as socio-economic status, aptitude measures, ethnic composition and differences in the characteristics of the instructional staff, the administration will gain an increasingly dependable idea of such school by school variations. Some of these differences, which may be rooted in the background of abilities that the children bring to school, require the focusing of special efforts and resources in particular schools to achieve satisfactory remediation.

Standard Score	SCHOOLS A	B	C	D	E	F
80	1		1			1
76		1	3			2
72	1	1			1	1
70		3	8			
69	1	4	1		2	3
68	1	4		1	2	2
66	3	3		1	1	
65	2	3	2		2	
63	3	4		1	2	1
62	3	3	1		1	3
61	3	(5)	4	1	1	4
60	1	1	2		(2)	4
59	3	4	4	1	2	2
58	(1)	2	4	1		
57	1	1	(2)	1	1	(3)
56	2	4	3	(1)	1	3
55	1	2	1			1
54	2	1	2			1
53	3	1	2			1
52		2	2	1	1	1
51	1	1	2		1	
50	1	3	1	3		1
49	2	1	4		2	1
48		2	2			3
47	2	1	1	1	1	2
46	2		1	1	1	
45	2	2				
44						
43			1			2
42	1	2	1			2
41	1		5			1
40					1	
39			1			
38		1	1			
37						
36	1			1		
35						
34			1			
33	1					
32						2
31						
30						
28					1	
N	46	62	63	15	26	47
Median Stan. Sc.	58	61	57	56	60	57
Corres. Gr. Eq.	7.6	8.4	7.3	7.1	8.1	7.3

Figure 3. School-by-School Comparison of Test Results for Grade 6.7 in a Single Community, Showing Distributions of Standard Scores for the Spelling Test in the Metropolitan Intermediate Battery.

4. Comparing a Pupil's Performance in Successive Years.

Not only is the profile a useful device for portraying the differential performance of a pupil (class or school) on the subtests of an achievement test battery, it can be used to show profiles of the same pupil for several years. The Teacher's Manual for the *Iowa Tests of Basic Skills* presents a standard permanent profile chart on which are

[2] Adapted from the manual of the *Metropolitan Achievement Test*, Harcourt, Brace & World, 1962.

plotted the test profiles of a pupil for two consecutive years. The chart utilizes the principle of plotting scores for all tests in the battery along a "standard scale" – in this case, grade equivalents. This principle is illustrated in Figure 4. The dotted line represents the performance of Frank Smith tested as a fourth grade pupil; the solid line represents his performance when tested at the same time in the fifth grade.[3]

By comparing the dotted and solid lines one may discover what relative progress he has made during the intervening school year in the various areas tested. It appears that Frank's gains on the language tests are larger than the typical gain of 10 points per grade but that they are less on other tests particularly Test W. The usual questions about the confidence one can place in this observation alone are relevant. An examination of the fifth-grade profile confirms the impression from the previous profile that Frank is relatively poor in arithmetic skills. The use of such profiles for consecutive years, not only gives a more complete picture of a pupil's performance, but also gives information about consistency of performance.

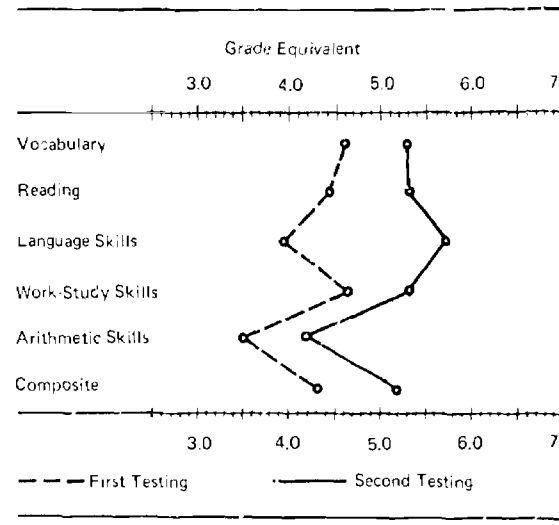


Figure 4. Performance of Frank Smith Tested as a Fourth and Later as a Fifth Grade Student.

WARNINGS ON THE USE OF PROFILES

Since profiles are relatively simple to construct and are at least superficially easy to interpret, they constitute one of the most popular methods of summarizing the results of multiple measurement. They have the obvious advantage that the graphic presentation enables one to view the total set of test scores and their possible interpretations at a glance. In the use of profiles, more than any other aspect of test interpretation, we need to beware of seeming simplicity and to understand the numerous pitfalls into which a naive interpretation would lead us.

For example, the failure to question whether or not the plotted scores are based upon common scales, which permit comparisons, may result in a distorted picture which has absolutely no meaning. Furthermore, most of us find it especially easy to interpret *apparent* differences in scores as real differences. By not being sensitive to the effect on our conclusions of the size of the profile scale we may arrive at erroneous decisions merely because our profile occupies a full page rather than a 2" x 3" corner. By failing to question the reliability of differences between scores and by relying solely on observed differences in arriving at conclusions, we ignore the possible unreliability and invalidity of the overall summation which would be instantly revealed if less "simple" methods were used.

These issues along with others must be understood by those who use or make profiles if the conclusions drawn are to be valid. This comment should not lead to the abandonment of the profile, but to an awareness of the precautions which should be followed in its use. The following portion of this paper will comment in some detail on the background and supplemental information one must consider for adequate interpretation of a profile. The material will be presented in the form of a number of warnings for proper construction and use of a profile.

1. Be sure the scores plotted are in comparable units.

Since raw test scores may vary considerably in meaning, it is obvious that raw or obtained scores do not possess the kind of comparability needed in plotting profiles. For example, the pupils in one class may have scores which range from 40 to 60 on a vocabulary test and from 5 to 25 on an arithmetic test. Hence, it would be impossible for anyone in the class to appear better in arithmetic than in vocabulary. Before any profile can be plotted which will permit desired comparisons it is clearly necessary to convert the scores to scales having comparable values.

Comparable scores are obtained usually by some scaling procedure so that all scores may be expressed in terms of a common reference point and a common unit of measurement. The authors of most tests fill this need by providing tables for

[3] Adapted from the profile chart shown on p. 18 of the *Manual for the Iowa Tests of Basic Skills*, Houghton Mifflin Co., 1956.

interpreting a raw score in relation to "normal" or typical performance. These tables of "norms" are based on statistical operations on scores of a normative or standardization group. A variety of definitions of comparability have given rise to a number of numerical indices used to express such comparability. Such measures as percentiles, standard scores, stanines, and grade equivalents have been derived to be consistent with specific definitions of comparability, and although different from each other, have proved to be useful in the interpretation of performance on tests. Note that comparisons using profiles must use the same unit across variables, even though it may be possible to present the same profile using several different measurement units. (See Figure 5)

2. Be sure the score scales on the profile are based on the same or strictly comparable populations which have been tested at the same time.

The importance of the particular reference population used to determine any such scale cannot be overemphasized. The performance of a student who scores at the 84th percentile or obtains a stanine of 6 in a reading comprehension test where these scores are based on a set of typical 7th grade scores is obviously not the same as the performance of one whose standing at the 84th percentile on the same test is calculated from the distribution of a below-average 7th grade. Likewise, a pupil with a vocabulary grade score of 5.2 obtained for a representative sample of 5th graders in one locality may not be at all comparable to a pupil who makes a score of 5.2 based on a representative national sample.

Note that comparable scores, either standard scores, percentile ranks, or stanines must be based on the same or strictly comparable populations. Also, note that a percentile score of 84 or a stanine of 6 based on a normative population tested in October does not represent as high a level of performance as a percentile score of 84 or a stanine of 6 based on a normative population tested the following April. Figure 6 presents data from which no meaningful statements can be made as to the relationships among the scores of John Jones whether we consider raw scores, percentile ranks, or standard score equivalents.

Name _____

Name of Test and Date of Testing	Norm Group	Raw Score	Stanine	Percentile Rank

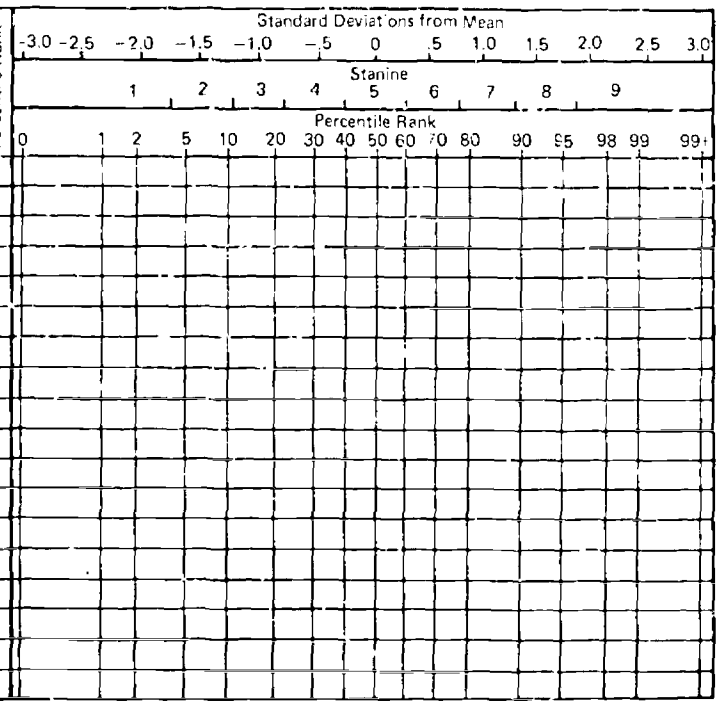


Figure 5. Test Record and Profile Chart.

Test and Normative Group	Raw Score	Percentile Rank
Learning Aptitude (Freshmen at Ohio St.)	76	45
Reading Comprehension (8th Grade Norms)	42	62
Mathematics (H. S. Algebra Class)	38	40
Mechanical Aptitude (H. S. Shop Class)	53	65
Clerical Aptitude (Employed Clerks)	175	80

Figure 6. Sam Jones: Test Results in Non-Comparable Form.

3. Do not depend upon observed differences alone.

One of the most serious abuses in profile interpretation occurs when teachers, counselors, or school psychologists depend upon *visual scanning only* for their interpretation. Too often, only a quick inspection of the profile is used to determine whether Jim is better in reading than in spelling or arithmetic. Two points that appear to be well separated on the profile are assumed to represent real and significant differences in ability. But such may not be the case. Any difference may be made to appear large by increasing the size of the scale, the way a photographer would make an enlargement of a print. In this way a minute difference may be made to look gigantic.

4. Check on the reliability and standard error of measurement of each test and each difference score.

To overcome being unduly influenced by size of scale you should be concerned about the reliability of the measurement and in particular the standard error of measurement of the scores used. Only by considering the size of the errors of measurement associated with the points on the profile can you understand with any certainty the meaning of an observed difference. Suppose that Johnny had a standard score of 60 in a reading test and a standard score of 55 in an arithmetic test. Suppose further that the reading test had a standard error of 7 and the arithmetic test had a standard error of 8. If you asserted, on the basis of these test scores, that Johnny was better in reading than in arithmetic you would be making a very hazardous statement. The errors of measurement are sufficiently large that on a subsequent retesting the scores could easily be reversed.

Some test publishers have provided a variety of ways to aid the profile user in assessing the reliability of the scores and differences presented on profiles using their particular instrument. For example, the manual for the DAT indicates that the authors and publisher have scaled these tests in such a way that a teacher or counselor can use a ruler to determine reliability of differences on the profile form provided. Plotted points must be separated by a vertical distance of one inch for the difference to be considered significant. The Stanford Achievement Test specifies that in comparing pupil performance a real difference can be assumed only when the two scores compared are more than one stanine apart. The STEP Test sets up a band around each score to indicate the extent of its unreliability.

In each of these instances the authors have computed the standard error of measurement and have incorporated it in their instructions as to how to use the profile for detecting meaningful differences. If such information is not given in the manual then it is necessary for the teacher, counselor, superintendent or other profile user to obtain information about reliability and to personally compute the standard error of measurement for each individual score and the standard error of measurement of differences between rubrics being compared. Otherwise he has no information about how much confidence he can place in his observed differences.

5. Be concerned about the independence or lack of independence of the variables shown on a profile.

When interpreting a profile we are concerned not only with the magnitude of the scores but also with those differences among them which constitute the essence of score pattern. We go beyond the interpretive statement: "Sam is very high in reading comprehension, moderately high in arithmetic computation and only average in science." In addition we often make interpretive statements such as: "Sam is higher in paragraph meaning than in arithmetic computation and science; he is higher in arithmetic than in science; and lower in science than either of the others. Therefore, he does not have the score pattern of a person likely to succeed in a field involving arithmetic and science."

Yet even though the individual scores may be reliable for answering certain questions, the unreliability of the differences on which the foregoing interpretation hinges may be such that Sam could actually be equal in all three or higher in paragraph meaning than in arithmetic. Again, the very con-

creteness of the graphic pattern gives it an appearance of accuracy that is wholly spurious. After all don't we have numerical scores, not only in black and white, but as points on a graph? Can't we rely on the pattern we observe for our interpretation?

Obviously the interpretation of a profile depends not only upon the scores, but upon the interrelationships and differences among them. But the reliability (and the interpretation) of the difference between scores for a single individual on two functions involve not only the reliability of the two tests, but also the correlation between them. The reliability of the difference between two measures which are correlated can easily be shown to be less than the reliability of the difference between the same two measures if they are independent of each other. Hence, when one is dealing with a profile involving a series of measures which are highly correlated, it is even more important to be concerned about the reliability of observed differences than when the measures are relatively independent.

6. Be sure that all necessary supporting information is included as part of the labeling on a profile.

The greatest advantage of general profile forms is that several different tests may be shown on the same sheet. The great limitation of such forms is the ease with which we may put tests with drastically dissimilar norm groups on the same sheet. Figure 5 is an example of a good general profile form for it has percentile, standard score, and stanine scales and calls for: title, normative group, the raw score, two different types of derived scores, and the date of testing. When preparing such a profile we should be careful to give complete information on all tests. What seems self evident at the moment of recording may not be so obvious months later. We need to be especially careful to record a complete designation of normative group.

7. Remember that profiles using lines joining points are *not* graphs in the usual sense.

In its usual form a profile is a graphic representation of a set of test scores for a single individual in which the tests are represented by ordinates spaced along the horizontal base line and the magnitude of each score is represented by plotting the point at the appropriate height on that ordinate. In order to aid the eye in locating the points thus plotted, it is customary to join the points by lines, leading to the more or less "jagged" picture that gives the technique its name. It should be remembered that the line thus drawn is not a graph in the usual sense.

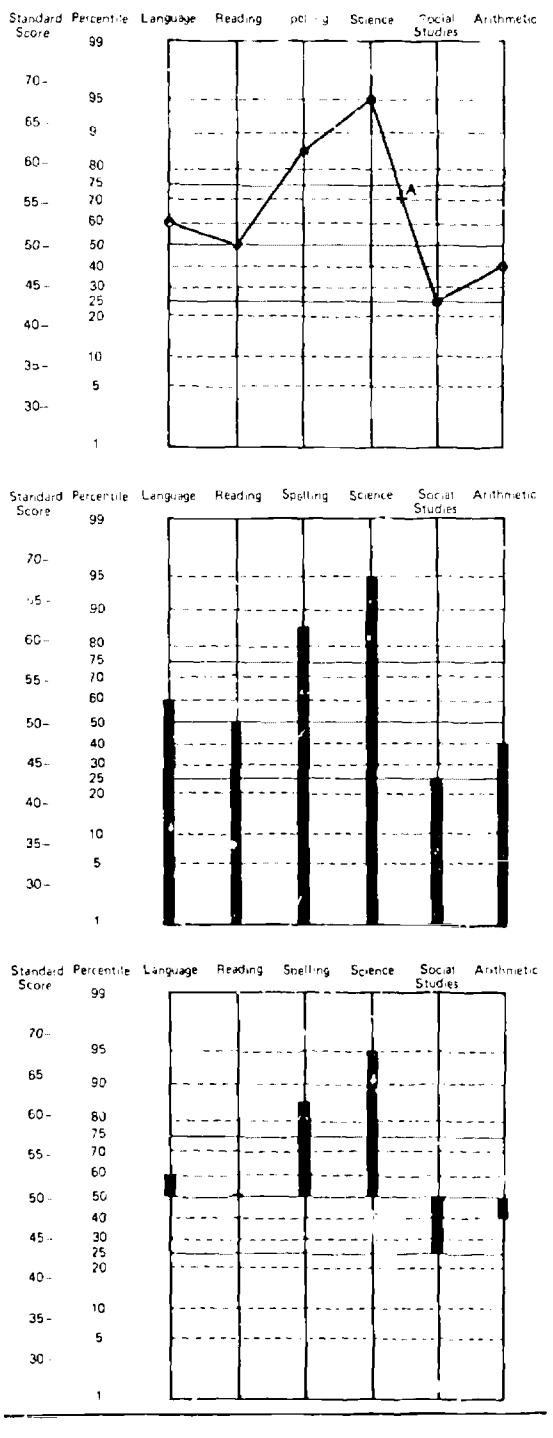
We are accustomed to dealing with graphs of continuous functions even though only a few points may be experimentally determined. In such graphs, the lines have meaning as representing values associated with values intermediate between those plotted. This is not true of ordinary test profiles. What would be the meaning of point "a" in the top profile illustrated in Figure 7? Some of these objections may be overcome by plotting profiles without connecting the profile points. Two methods of doing so along with the more common method are illustrated in Figure 7.[1] These two profiles avoid the false assumption inherent in the connecting of score points on the tests as well as being less subject to configural misinterpretation.

8. Base your interpretation on test scores aided by the profile.

Since the user of a profile frequently tends to think of a profile as representing a pattern of test results the arrangements of the horizontal ordinates is of importance. In which order should a set of tests be arranged for plotting? Since all of the scores are presented and since the lines between plotted points are meaningless, it may be that order on the base line is wholly immaterial. On the other hand, the interpretation of the "pattern" of the profile is often made as a psychological judgment, based not only on the numerical values of the scores, but on their total perceptual configuration.

To the extent that this latter factor enters, order is important. Consider the impression made by the two profiles shown in Figure 8 each based on the same set of test scores. Which is the easier to interpret; which the most open to misinterpretation? To the best of my knowledge no investigation of these problems has been made directly. Without having adequate answers for these questions the practical user of profiles can avoid

[1] Adapted from the Individual Report Form of the *Differential Aptitude Tests*, The Psychological Corporation, 1947.



Three Methods of Profiling the Same Set of

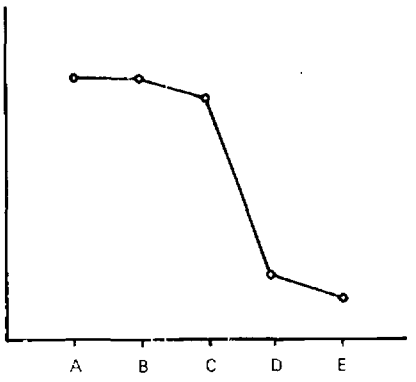
possible errors by regarding the profile as a method of conveniently presenting the actual test results. Interpretation should be based on the scores *aided* by the profile, *not on the profile* aided by the scores.

9. Do not attempt a simplified version of profile analysis.

One of the most extensively researched and complex fields is that of comparing the profile of an individual with the profile of a normative group of individuals having a certain characteristic or who are successful in some particular occupation. On the *Weschler Adult Intelligence Scale* for example, some clinical psychologists have used relationships between scores on certain of the eleven scales as a basis for personality diagnosis. The authors of the California Test Bureau's *Multiple Aptitude Tests* have prepared a number of typical profiles of occupational and other groups, suggesting that important questions can be answered by comparing individual profiles with these examples. In considering the effectiveness of interpretation, many of our questions apply to any normative score, whether on a profile or not. It is worth noting, however, that the use of a profile does not solve the questions, and certainly should not lead to ignoring them.

However, assuming adequate reliability certain meaningful statements can be made about the *resemblence* of an individual's profile to the profiles of specific occupational groups. For example, after administering a battery of tests to Sam, a twelfth grader, we might be able to say that his profile was similar to the profile (pattern of test scores) of the average engineer but was unlike that of the average lawyer. But assuming similarity of profiles, can we say whether he would be successful in either profession? Is the individual engineer more or less successful as his profile lies wholly above or wholly below the average of the group? What shall we say of the individual whose profile, when compared with that for the group in which we are interested, shows several points above the group mean, but one point conspicuously below it? To what extent, then, does superiority in one or more tests compensate for marked deficiency in another?

9



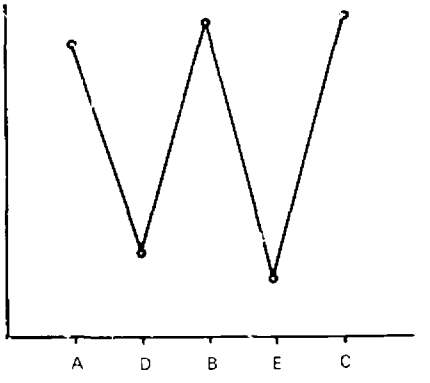


Figure 8. Two Hypothetical Profiles Showing Different Arrangements of the Same Scores.

It is obvious that we are interested in two quite different questions. Question 1 – Do Sam's test scores as profiled permit us to say that he belongs to a particular group, e.g., engineer, when his profile resembles that of the average engineer? Question 2 – If we are able to say that his profile is similar to those of engineers, can we say furthermore that he would be successful as an engineer? To answer these two questions quantitative procedures which are far superior to profile interpretation are available. The first question, namely, group membership, can be answered effectively by the use of the multiple discriminant function, and the second question – how successful a person is likely to be – can be answered best by multiple regression procedures.

Returning now to the actual situation where we are not able to assume reliability for the profile of either the normative group or that of the individual under consideration, what quantitative answer can we give as to the limits of tolerance within which the individual's profile must agree with the criterion profile with which he is being compared? As yet, no one procedure has been found to be completely satisfactory. However, there has been extensive research in this area and a number of procedures have been proposed. Among the number of techniques proposed for measuring profile similarity are the following:

1. Coefficients of correlation (Burt, 1937)
2. Coefficient of profile similarity (duMas, 1946)
3. Coefficient of intra class similarity (Webster, 1952)
4. Coefficient of pattern similarity (Cattell, 1949)
5. Distance measure D (Osgood and Suci, 1952)
6. Dissimilarity index D (Cronbach and Gleser, 1953)

The appropriate statistical technique to be used for any particular set of data will depend upon the assumptions underlying the techniques and how well the investigator understands the nature of the scores he is using. No attempt is made in this paper to describe these particular procedures. They are merely presented as preferable alternatives to profiles for interpreting certain types of sets of test scores. Recent papers by Marx (1968), Nunnally (1962), McHugh and Sivanich (1963), and Heermann (1965) have focused on the general problem of profile comparison.

CONCLUSIONS

This paper has attempted to comment on four issues, although equal weight has not been given to each. The main effort has focused on the first two, namely, the use of and pitfalls in interpreting profiles. I have tried to describe and illustrate in some detail the usefulness of profiles as frames of reference for the interpretation of test scores by school personnel; and I have stressed the need to base interpretations on the test scores aided by the profile *not* on the profile configuration aided by the test scores.

Since profiles are relatively simple to construct and appear, at least superficially, easy to interpret, considerable emphasis was placed upon understanding the numerous pitfalls into which a naive interpretation would lead. Comments in the form of nine warnings were presented to assist the school man in using profiles appropriately.

The other two issues were a consideration of the profile configuration itself and a comment on quantitative procedures that for certain purposes are superior to profile analysis. The complexities involved in answering questions by comparing one profile with another or with a vocational normative group were discussed. Even though such comparisons are useful to researchers, I advise the teacher, supervisor, and administrator to avoid simplified versions of profile analysis. Although I have mentioned some of the methods proposed, I have not described them. References have been given which will permit the interested reader to investigate this complex problem for himself.

Finally, only a casual reference was made to two important quantitative procedures – multiple discriminent analysis and multiple regression analysis. These methods are superior to profiles for answering the questions – Which group does a person most resemble? and, How successful is a person likely to be if he is a member of a certain vocational group? Here also references have been given so that an interested reader with a strong statistical background may pursue these problems further.

BIBLIOGRAPHY

Bennett, G.K., Seashore, H.G., and Wesman, A.G. *Counseling from Profiles.* New York: Psychological Corporation, 1951.

Burt, C.L. Correlation between persons. *British Journal of Psychology* 1937, 28, 59-96.

Cattell, R.B. r_p and other coefficients of pattern similarity. *Psychometrika*, 1949, 14, 279-298.

Cronbach, L. and Gleser, G. Assessing similarity between profiles. *Psychological Bulletin*, 1953, 50, 456-473.

duMas, F.M. The coefficient of profile similarity. *Journal of Clinical Psychology*, 1949, 5, 123-131.

Heermann, E.F. Comments on Overall's multivariate methods for profile analysis. *Psychological Bulletin*, 1965, 63, 128.

Marks, E. Some profile methods useful with singular covariance matrices. *Psychological Bulletin*, 1965, 70, 179-184.

McHugh, R.B. and Sivanich, G. Assessing the temporal stability of profile groups in a comparative experiment or survey. *Psychological Reports*, 1963, 13, 145-146.

Nunnally, J. The analysis of profile data. *Psychological Bulletin*, 1962, 59, 311-319.

Osgood, C.E. and Suci, G. A measure of relation determined by both mean difference and profile information. *Psychological Bulletin*, 1952, 49, 251-262.

Overall, J.E. Note on multivariate methods for profile analysis. *Psychological Bulletin*, 1964, 61, 195-198.

Tatsuoka, M.M. and Tiedeman, D.V. Discriminant analysis. *Review of Educational Research*, 1954, 24, 402-420.

Tiedeman, D.V. A Model for the profile problem. *Proceedings, 1953 Invitational Conference on Testing Problems.* Princeton, N.J.: Educational Testing Service, 1954.

Webster, H. A note on profile similarity. *Psychological Bulletin*, 1952, 49, 538-539.

NCME
measurement in education

Volume 1, No. 2 January, 1970

Warren W. Willingham, Editor

Advisory Committee:
J. Thomas Hastings, University of Illinois
Alice J. Irby, Educational Testing Service
Robert S. Lankton, Detroit Public Schools

Measurement in Education is a series of special reports published four times a year in October, January, March, and May by the National Council on Measurement in Education. These reports are concerned with the practical implications of measurement and related research and their application to educational problems of individuals, institutions, and systems. The emphasis is upon uses of measurement rather than technical or theoretical issues. Subscription rate: $2.00 per year; single copies 30¢ each in quantities of 25 or more.

Application to mail at second-class postage rates is pending at East Lansing, Michigan 48823.

Address Editorial Correspondence to:

Warren W. Willingham
College Board
800 Welch Road
Palo Alto, Ca. 94304

Address Business Correspondence to:

Irvin J. Lehmann
Office of Evaluation Service
Michigan State University
East Lansing, Mich. 48823

Application to mail at second class postage rates is pending at East Lansing, Michigan 48823.

NATIONAL COUNCIL ON MEASUREMENT IN EDUCATION

Office of Evaluation Services
Michigan State University
East Lansing, Michigan 48823