DOCUMENT RESUME

ED 393 884 TM 024 786

AUTHOR Callahan, Joseph P.
TITLE Use of Derived Comparison Scores in Clinical Settings.
PUB DATE [95]
NOTE 9p.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Age Differences; *Clinical Diagnosis; *Comparative Analysis; *Instructional Program Divisions; *Scores; Statistical Distributions; *Test Interpretation
IDENTIFIERS *Percentile Ranks

ABSTRACT

Clinicians use a common practice of reporting age or grade equivalent derived scores, but problems of interpretation exist for such scores. This article examines derived comparison score issues and recommends use of scores of relative standing such as percentile ranks or standard scores in contrast to use of developmental scores like age and grade equivalents. Derivation of both age and grade equivalents is usually based on a hypothetical distribution of means with scores often interpolated and extrapolated along the distribution of means curve. Scores of relative standing such as percentile ranks offer a way for the clinician to interpret test scores when interval or ordinal measurement is used to obtain raw scores. The percentile family is still a derived score but is based on the percent of individuals taking the particular test. The interpretation for a specific percentile rank is that the individual scored as well or better than that percentage of individuals taking the test. Use of percentiles allows for fewer interpretative problems and calculation for any shape of distribution. (Contains six references.) (Author/SLD)

Running head: DERIVED COMPARISON SCORES IN CLINICAL SETTINGS

Use of Derived Comparison Scores in Clinical Settings

Joseph P. Callahan

Dickinson State University

Abstract

Clinicians use common practice of reporting age or grade equivalent derived scores. Problems of interpretation exist for use of age and grade equivalent scores. This article examines derived comparison score issues and recommends use of scores of relative standing such as percentile ranks or standard scores in contrast to use of developmental scores like age and grade equivalents.

Use of Derived Comparison Scores in Clinical Settings

Professionals in clinical settings often use standardized tests to assess characteristics of individuals. Many of the standardized tests or scales used are norm-referenced. The characteristic or domain assessed often represents the performance of a known group called the normative sample or norm group. When comparison scores are derived they fall into one of two categories: (a) developmental scores or (b) scores of relative performance (Salvia & Ysseldyke, 1991).

While many commercial test publishers provide a variety of data tables, raw scores are most often transformed into developmental scores such as grade equivalents (e.g., 5.2, fifth grade-second month) or age equivalents (e.g., 10-6, ten years-six months). Scores of relative performance such as percentiles, standard scores (Z-scores, T-scores) and normal-curve equivalents go beyond the typical average (median or mean) for derivation (Sax, 1989).

Normative data tables for many speech-language pathology tests provide for conversion of raw score data to age equivalent or grade equivalent transformed scores.

This paper argues against using grade equivalent or age equivalent developmental scores in favor of scores of relative performance such as percentiles for clinical test interpretation.

Derivation of Age and Grade Equivalent Scores

Expressing test scores in age equivalents or grade equivalents is a common interpretative practice. Usually, the

test manual provides for age equivalent or grade equivalent specification from a table. For example, on the Peabody Picture Vocabulary Test (1981), a raw score of 114 is expressed as an age equivalent of 11-1 or eleven years and one month. The age equivalents of many norm-referenced tests are derived by standard practices of finding the median or mean performance for an age group equivalent to a particular raw score. Grade equivalents are derived in similar fashion to age equivalents and correspond to a median or mean grade and month in school in tenths of months. Test scores for grade equivalents are usually reported in decimal notation (e.g., 6.2 for sixth grade second month); however, some tests report grade equivalents with a hyphen such as 6-2 for sixth grade second month. Ten months is the maximum for grade equivalents while eleven months is the maximum for age equivalents.

Derivation of both age and grade equivalents is usually based on a hypothetical distribution of means with scores often interpolated and extrapolated along the distribution of means curve (Salvia & Ysseldyke, 1991).

Interpretation of Age and Grade Equivalents

The typical interpretation of age and grade equivalent scores is to state that a particular individual's performance is equal to the average 10 year old, for example, or average third grader's performance on a certain dimension of a scale. Salvia and Ysseldyke (1991) have documented difficulties for the interpretation of age or grade equivalent scores. These problems

relate to use of terminology, such as, a child "performed" at the average of a 10-6 year old. In actuality, the "performance" may be different and the individual is actually obtaining a derived score. Because tables use extrapolation and interpolation, the scores are, in many instances, only estimates of the age or grade equivalent obtained. In a commissioned report for the National Research Council, the Committee on Ability Testing (Wigdor & Garner, 1982) noted "Age equivalent scores are sometimes used, but they are not considered valuable because of difficulties in interpreting them. The age equivalent was popularized as the so-called mental age with early IQ tests" (p. 46). Lawrence (1992) commented on normative scores such as age equivalents and inherent problems: "This developmental norm [age equivalent] makes no attempt to describe a range of normal performance" (p. 7). Lawrence further implied that age equivalent scores often can be falsely interpreted as scores of relative standing by teachers and parents (Lawrence, 1992).

The impression of average performance for a particular age or grade level can be misleading. Caution must be employed in interpreting scores for the highest grade or age levels represented because of lack of equal intervals and resultant ordinality and subsequent curve flattening.

Scores of Relative Standing Offer More for Interpretation

When the median or mean is used as the primary method to determine a derived score, information is extrapolated from the normative group performance and interpolated to age or grade

equivalent scores. Clinicians have an alternative to normative age equivalent or grade equivalent reporting systems. Scores of relative standing such as percentile ranks offer a way for the clinician to interpret test scores when interval or ordinal measurement is used to obtain raw scores. The percentile family is still a derived score but is based upon the percent of individuals taking the particular test. The interpretation for a specific percentile rank (e.g., 84) is that the individual scored as well as or better than 84 percent of those taking the test. Brown (1991) noted that percentile ranks "make it possible to compare the performance of groups that are unequal in size, age, ability, or other ways" (p. 346). A main consideration for the clinician is to reinforce the fact that percentile rank is not the same as percent correct. Other scores of relative standing which would help avoid the interpretative problems of age equivalent scores or percentile ranks are standard scores such as Z-scores, and T-scores; however, they do not appear to be popular in clinical settings.

Age equivalent scores are commonly reported by clinicians. Percentile ranks represent an improved manner in which to show relative standing. Use of percentiles allows for fewer interpretative problems and calculation for any shape of distribution. If a clinician is going to use a derived score, the percentile rank is preferred to the age equivalent score.

References

Brown, J.R. (1991). The retrograde motion of planets and children: Interpreting percentile rank. Psychology in the Schools, 28, 345-353.

Dunn, L.M., & Dunn, L.M. (1981). Peabody picture vocabulary test-revised, Circle, Pinnes, MN: American Guidance Service.

Garner, W.R., & Wigdor, A.K. (Eds.). (1982). Ability testing: Uses, consequences, and controversies. Washington, DC: National Academy Press.

Lawrence, C.W. (1992). Assessing the use of age-equivalent scores in clinical management. Language, Speech, and Hearing Services in Schools, 23, 6-8.

Salvia, J., & Ysseldyke, J.E. (1991) Assessment (5th ed.). Boston: Houghton Mifflin Co.

Sax, G. (1989). Principles of educational and psychological measurement and evaluation (3rd ed.). Belmont, CA: Wadsworth Publishing Company.

Author Note

Joseph P. Callahan, Vice President for Academic Affairs & Professor of Education, Dickinson State University.

Correspondence concerning this article should be addressed to Joseph P. Callahan, 291 Campus Drive, Dickinson State University, Dickinson, North Dakota 58601. Electronic mail may be sent via internet to Joe_Callahan@dsu1.dsu.nodak.edu.