Considerations in Design and Use of Scales in Rehabilitative Audiology

Charles L. Hutton

Abstract

Self-assessment questionnaires in rehabilitative audiology typically possess the measurement characteristics of scales based on forced-choice and rank-order. Data from these ordinal scales neither support arithmetic analysis nor arithmetic interpretation. However, powerful nonparametric statistics are available for application to data from ordinal level questionnaires. Thus, methodologic principles influence the way self-assessment scales are designed, constructed, and analyzed. They also influence the way clinicians interpret responses from individual patients. Within these constraints, well constructed questionnaires can provide information on a wide range of topics in rehabilitative audiology. Suggestions are given for generating interval level questionnaires that are more precise than existing, ordinal level scales.

Key Words: Interpretation, interval scale, nonparametric statistics, ordinal scale, self assessment questionnaire.

In recent decades, the role of behavioral variables in rehabilitative audiology has received increasing attention. This interest has resulted in the development of scales that sample such topics as: (1) perception of amount of hearing handicap; (2) attitude of patient and family toward hearing impairment; (3) patient skills for coping with and management of hearing impairment; (4) communication performance in different settings and under various signal conditions; (5) hearing aid performance; and (6) hearing aid use (Alpiner, 1987; Salomon et al, 1988; Schow et al, 1989). In the words of Schow et al (1989), “Self assessment of hearing is currently enjoying an unprecedented emphasis.”

In view of the increasing use of self-assessment scales in rehabilitative audiology, it seems appropriate to examine the methodology underlying such scales, since the methodology determines which statistical tools are appropriate for analysis and interpretation of the ensuing clinical data. While these methodologic concepts are not unique to audiology, this paper focuses on the use of scales in rehabilitative audiology because both construction and application of these scales are spreading so rapidly in this field. In order to understand the relationship between methodology of measurement and statistics, a discussion of the basic characteristics that underlie Stevens's (1951) four classes of scales is presented.

CHARACTERISTICS OF MEASUREMENTS

Physicists, chemists, and other physical scientists have long used the two most powerful measurement scales, known as ratio and interval scales. As seen in Table 1, ratio scales have four characteristics or features. The most precise characteristic of ratio scales is that they have an absolute reference value. Also, at this level of measurement, the size of the difference between intervals is known, the intervals are ordered in at least one dimension, and members of an interval share a common property (Stevens, 1951). Examples of the use of ratio scales in rehabilitative audiology can be found in the estimates of hours of wear obtained by Brooks (1979) and Hutton (1983). Other examples of ratio scales include the
Table 1  Examples of Measurement Scales in Rehabilitation Audiology and the Characteristics of Each

<table>
<thead>
<tr>
<th>Scale Type</th>
<th>Ratio (Hours of Wear)</th>
<th>Interval (Octave)</th>
<th>Ordinal (Frequency of Occurrence)</th>
<th>Nominal (Site of Lesion)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>125 Hz</td>
<td>99% Always</td>
<td>1. Conductive</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>250 Hz</td>
<td>87% Practically Always</td>
<td>2. Mixed peripheral</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>500 Hz</td>
<td>75% Generally</td>
<td>3. Sensory</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1,000 Hz</td>
<td>50% Half-the-time</td>
<td>4. 8th Nerve</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>2,000 Hz</td>
<td>25% Occasionally</td>
<td>5. Brainstem</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>4,000 Hz</td>
<td>12% Seldom</td>
<td>6. Central</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>8,000 Hz</td>
<td>1% Never</td>
<td>7. Mixed peripheral/central</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>16,000 Hz</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Characteristics

<table>
<thead>
<tr>
<th>Absolute zero</th>
<th>Yes</th>
<th>No</th>
<th>No</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Known difference between intervals</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Intervals are in rank order</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Common property</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

An important characteristic of ratio scales is that scale intervals are of equal size, either initially or through numeric transformation. The fact that the time interval from 2 to 4 hours and the interval from 4 to 6 hours is the same makes it possible to add, subtract, multiply, and divide estimates of hours of wear. Therefore, parametric statistical tools can be applied, if the necessary assumptions are met concerning the population, or the sample(s), or derivatives of the sample(s) (Hays, 1963). These assumptions are mathematically complex and discussion of such topics as the restrictions imposed by truncated variance, abnormalities in the tails of distributions, the role of homoscadasticity, and criteria for application of the Central Limit Theorem is beyond the scope of this paper. Interested readers can consult Hotelling (1961) and Bradley (1968).

An interval scale is less precise because it lacks an absolute zero. Brooks (1981) provides an example of an interval scale. The scale he devised contains five categories of battery usage as defined by battery weight gain. Since the differences between categories were specified in milligrams, and since there was not an absolute zero, this scale has the characteristics of an interval scale. Another interval scale of interest to audiologists is the octave scale. Any frequency can be used as the zero reference for an octave and one octave higher is always twice the reference frequency. Interval scales are sometimes called equal interval scales, because the distance between units of measurement are either linear or can be made linear by multiplying or dividing the data by a constant amount. However, as shown below, there are some pitfalls in data transformation. Equal interval scales are desirable because, like ratio scales, they also support the use of parametric statistics.

Ordinal Scales

A less powerful scale, the ordinal scale, has two characteristics: (1) ordering of intervals and (2) members of each interval have a common property. The ranking or ordering of intervals is a critical feature. The intervals are not necessarily equal in size. For example, many of the items on the Hearing Measurement Scale (Noble and Atherley, 1970) use the structure "Do you ever...?". Presumably, the interval for "No" means never while the interval for "Yes" includes the range from almost never to always. The Hearing Handicap Inventory for the Elderly (Weinstein et al, 1986) contains three response categories, Yes (4), Sometimes (2), and No (0). The numbers 0, 2, and 4 associated with the three intervals on this scale would suggest...
that the intervals are equidistant. However, a study of the meanings associated with qualitative terms suggests that this is an unlikely probability (Simpson, 1963).

Most self-assessment scales in rehabilitative audiology use relative frequency of occurrence for the response scale, including the Hearing Handicap Scale (High et al., 1964), the Hearing Performance Inventory (Giolas et al., 1979), the Hearing Problem Inventory—Atlanta (Hutton, 1987), and the Self-Assessment of Communication (Schow and Nerbonne, 1982). Other scales use amount of agreement-disagreement, including the Modified Denver Scale of Communication (Kaplan et al., 1978), and the Communication Profile for the Hearing Impaired (Demorest and Erdman, 1986). Frequency of occurrence and agreement-disagreement scales appear to have two underlying limitations. First, each user provides her or his unique definition of frequency of occurrence or agreement-disagreement. Second, each user brings to this task a set of attitudes. As a result, a large number of variables have been shown to influence the reliability and validity of responses on ordinal scales (Nunnally, 1978; Hutton, 1980; Johnston and Pennypacker, 1980; Demorest and Erdman, 1986; Stephens et al., 1990).

Given the above limitations, what can well constructed ordinal level self-assessment scales be used for? If the responses are converted to percentiles and the data analyzed using nonparametric statistics, the data can be used in a number of ways to obtain information about the rehabilitative process. A pre-treatment questionnaire can be administered in order to get information for planning rehabilitation. A post-treatment questionnaire can be used to estimate the amount of benefit and to monitor quality assurance outcomes. Interpretation of patient responses on a questionnaire can be facilitated by comparison with responses from other patients. Item responses can be compared with a distribution of responses by a group of patients of similar hearing loss, experience wearing an aid, employment status, and age, i.e., does the patient’s response fall in the upper 25 percent, or between the 25th to 75th percentiles, or in the lower 25 percent for that item (Hutton, 1980, 1983). If questions are grouped by topic, a median score can be computed for each topic and compared with percentile distributions for that topic (Hutton, 1987; Garstecki et al., 1990). In addition, self-assessment questionnaires can be used to study characteristics of contrasting groups of patients. For example, the characteristics of patients who are part-time hearing aid wearers could be compared with those who are full-time wearers. Likewise, the characteristics of patients who show the least amount of pre-post rehabilitative gain (lower 25th percentile) could be compared with those who show the largest amount of gain (upper 25th percentile). Achievement of program goals, e.g., ability to cope with impairment, acquisition of strategies to maximize communication, and use of visual clues, can provide estimates of the success of the treatment program. Similarly, the effects of different fitting strategies, different kinds and lengths of rehabilitation programs can be evaluated. In short, the use of nonparametric statistics does not handicap either designers or users of self-assessment questionnaires.

Within the class of scales that contains the properties of equality and order, it is possible to create a hierarchy of precision. The least precise ordinal scale is one where the response scale is described and the patient supplies her or his interpretations of the intervals. More precise measures would likely result if each interval were defined by physically quantifiable variables. For example, the following variables could be used: signal source, room size, noise level, and amount of light. One end of the response scale could be defined as “Talking to a friend in a small, quiet, well lighted room.” The other end of the scale could be defined as “Talking to a group of strangers in a large, noisy, poorly lighted room.” The intervals between these two ends would be defined by variations of each of the four variables. A sample of the relevant population could be used to determine the difficulty of the different variations. After eliminating sets judged to be similar in difficulty, the remaining sets would constitute an ordinal scale with the precision of physically defined intervals. Many of the existing items on self-assessment questionnaires could be adapted to this kind of response format, an example of which is displayed in Figure 1.

Interval Scales for Rehabilitative Audiology

An ordinal scale with physically defined variables could be used as the basis for generat-
Figure 1 Example of an ordinal scale with physical variables for rehabilitative audiology. (The items shown were adapted from the Hearing Problem Inventory—Atlanta, Hutton, 1987.)

<table>
<thead>
<tr>
<th>Problem Description</th>
<th>Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>I understand when I am</td>
<td>2</td>
</tr>
<tr>
<td>It's too hard to listen when I am</td>
<td>3</td>
</tr>
<tr>
<td>My hearing loss is embarrassing to my family when I am</td>
<td>5</td>
</tr>
<tr>
<td>People have to talk slowly for me to understand them when I am</td>
<td>6</td>
</tr>
<tr>
<td>I feel that people avoid talking to me when I am</td>
<td>7</td>
</tr>
<tr>
<td>When I don't hear, people leave me out and go on talking when I am</td>
<td>10</td>
</tr>
<tr>
<td>When I have problems understanding, I tell people when I am</td>
<td>11</td>
</tr>
<tr>
<td>When I misunderstand people and say things that don't fit in the conversation when I am</td>
<td>12</td>
</tr>
<tr>
<td>When I don't understand, people get annoyed when I am</td>
<td>13</td>
</tr>
<tr>
<td>I can carry on a conversation with people who talk softly when I am</td>
<td>16</td>
</tr>
<tr>
<td>I try to position myself so I can clearly see the speaker's facial expressions when I am</td>
<td>17</td>
</tr>
<tr>
<td>When I can't follow what is said, I feel uncomfortable when I am</td>
<td>19</td>
</tr>
<tr>
<td>When someone talks from behind me, I miss what they say when I am</td>
<td>20</td>
</tr>
</tbody>
</table>

In order to sample perception of hearing aid characteristics, another scale could be defined by variables such as signal-to-noise ratio, fidelity of amplification, amount of loudness, and presence of squeal. One end of this response scale could be defined as "Separates speech from background noise; sounds natural; has appropriate loudness; doesn't squeal." The other end could be defined as "Nothing but noise; very distorted; can't stand the loud noises; squeals uncontrollably." Although the use of these procedures would considerably lengthen the development of scales, the gains would be sub-
Use of Scales/Hutton

stantial. Use of response scales with known, specified intervals would have two principal benefits: (1) since the measures from such scales would be traceable to known quantities, responses should be influenced less by variables not specifically related to communication (Stephens et al., 1990); and (2) the use of interval level scales would facilitate the analysis and interpretation of patient responses. At the present time, few items on assessment scales generate data that support addition, subtraction, multiplication, or division. Statistics appropriate for the analysis and interpretation of data from the existing, ordinal level scales include the use of the percentile to represent individual data, the median to estimate the central location of a number of responses, and the interquartile range to estimate the dispersion of responses (White, 1979). A wide range of procedures to test single and multiple distributions of independent and matched ordinal data for differences, for dispersion, trends, and correlation can be found in Siegel (1956) and Conover (1980). Readers who are not familiar with nonparametric or distribution-free statistics will find Siegel's treatment to be singularly readable.

Nominal Scales

The least powerful member of Stevens's classification is the nominal scale. This designation is unique to Stevens. The traditional view of measurement, as espoused by both physical and behavioral scientists, involves a sense of magnitude (Torgerson, 1958). Nominal scales lack any element of magnitude. Each interval on the scale is a category, each of which has one or more unique characteristics. For example, if each patient were placed in one of seven categories based on site of lesion (see Table 1), such questions could be posed as "Is the proportion of scores above 90 percent the same for all sites of lesion?" While the use of nominal scales has been limited in rehabilitative audiology, a wide range of statistical tools is available for analysis of such data.

Construction and Evaluation of Scales

As noted in the discussion of applications of ordinal scales, use of a self-assessment scale can be limited both by design features and by construction characteristics. While this paper focuses on design aspects of ordinal and interval scales, users of a questionnaire are also concerned with the ability of items to elicit the same response when the same conditions are repeated (reliability), the ability of the items to yield information about the knowledge and skills of the sample (validity), and to what samples the normative data apply. There are many different kinds of reliability, each of which can influence item selection and the overall reliability of the questionnaire. Likewise, there are many ways to obtain estimates of item or scale validity. Each of these areas merits thorough examination by the developer and the user. Discussions of various measures of reliability and validity can be found in Guilford (1954), Siegel (1956), and Hays (1963) as well as in texts devoted to test construction. A demonstration of many of the principles involved in the construction and analysis of self-assessment questionnaires can be found in Demorest and Walden (1984).

STATISTICAL CONSIDERATIONS

When determining which statistics are appropriate for the analysis and interpretation of data, it is necessary to consider (1) the methodology of the measurements; (2) distributional characteristics of the measurements; and (3) the wide variations in sample sizes often encountered in clinical studies (Siegel, 1956; Torgerson, 1958; Hays, 1963; Moore, 1979; Meyers, 1984). This paper is not intended to be an exhaustive or critical analysis of rehabilitative literature. However, informal examination of five journals (British Journal of Audiology, Ear and Hearing, Journal of the Academy of Rehabilitative Audiology, Journal of Scandinavian Audiology, Journal of Speech and Hearing Disorders) disclosed that data from self-assessment scales were either skewed or the dispersions were heterogeneous, or the sample sizes varied from zero to more than 30. Nonetheless, parametric statistics were widely applied to the data. An example of heterogeneity of dispersion in clinical data can be seen in a report of hearing aid wear (Hutton, 1983). The medians and interquartile ranges reflect skewness in one direction at the low end of the scale, skewness in the other direction at the other end of the scale, and systematic decreases in the interquartile range. Changes in direction and amount of dispersion call for the use of nonparametric statistics in both analysis and interpretation of data (Torgerson, 1958).
Before leaving this discussion, it should be reiterated that central to understanding the structure of measurement scales and the statistics they support is knowing the precision with which the response, not the stimulus, is measured (Stevens, 1951; Moore, 1979). For example, when a patient is asked to indicate whether samples of amplified speech are soft, comfortable, or loud, the resulting data are ordinal level measurements, even though the stimuli may be measured in dB SPL. However, this psychophysical methodology results in more precise data than a scale based on psychological intervals, e.g., soft-loud. At the present time, no self-assessment questionnaire has the precision of a scale with physically defined stimuli. However, Guilford (1954) devotes a chapter to the development of scales based on interval and ratio judgments. Finally, the above is an incomplete description of the structure of measurement scales and the statistics they support. More detailed discussions of these relationships can be found in Stevens (1951), Siegel (1956), and Hays (1963).

Given the above information on the roles of scale methodology and data distributional characteristics in the analysis and interpretation of data, there is surprising diversity among statistical sources. One widely used reference for behavioral research (Winer, 1962) contains no discussion of scaling concepts, nor the statistics they support. With few exceptions, the statistical tools presented in this reference assume interval or ratio precision and normal distributional characteristics. Probably the view most often presented in references is that although behavioral data are ordinal, the power of parametric statistics justifies their use, especially if the distributions approach normalcy and the dispersions are relatively homogeneous (Hays, 1963; Nunnally, 1978). However, these same sources restrict the application of parametric statistics to the analysis of the data. Warnings are issued not to give arithmetic interpretation to the data (Hays, 1963). According to this guideline, it would be appropriate to test differences between the means of pre- and post-treatment problem scores using parametric statistics. For illustrative purposes, assume a mean pre-treatment problem score of 2.0, a post-treatment mean score of 1.5, and a sample size of 55 patients (Hutton, 1980; Table 1). Further, assume that application of the parametric “t” test shows the difference between these two means has a chance probability of less than 5 percent. However, it would not be appropriate to give a linear interpretation to this difference. That is, it would not be appropriate to say that this short-term rehabilitation program reduced problems by 25 percent because the distance between 1.5 and 2.0 cannot be assumed to be 25 percent. This interpretation would be appropriate only if the response scale consisted of equally spaced intervals.

Nonparametric Statistics

A more rigorous approach to analyzing and interpreting ordinal data is advocated by Siegel (1956) and Conover (1980). According to these authors, it is not appropriate to use statistics that require parametric assumptions if the data are derived from ordinal methodology. However, use of nonparametric statistics results in a loss of efficiency. In order to give the reader a sense of the amount of the loss of power, five widely used statistics cited in Siegel (1956) will be compared. The Wilcoxon matched pairs signed-ranks test can be used to evaluate pre-post treatment differences of the same patients. Thus it is equivalent to a “t” test for matched pairs. In comparison to the “t” test, the Wilcoxon test is about 95 percent as efficient. That is, in order to test the data at the same level of probability, it would be necessary to increase the sample size by 5 percent. The Mann-Whitney U test can be applied to independent samples. It could be used to compare two different treatments applied to different groups of patients. In comparison to its parametric “t” counterpart, it also has an efficiency of about 95 percent. The Friedman two-way nonparametric analysis of variance has, overall, the same power as its parametric counterpart. Two rank-order correlations, the Spearman rs and the Kendall r, have an efficiency of 91 percent in comparison to the Pearson product-moment correlation. A discussion of the advantages and disadvantages of nonparametric tests can be found in Siegel (1956) and in Conover (1980).

To what degree has information on scale methodology and data distributional properties been applied to self-assessment scales in rehabilitative audiology? Six of the eight reports on self-assessment questionnaires discussed earlier contain descriptions of response data. All six have ordinal response scales. All six contain descriptions of non-normal item distributions, e.g., bimodal, linear, skewed. Although these data do not meet criteria for application of
parametric statistics, means and standard deviations were used to describe the data. In one report, the mean plus one standard deviation is close to one end of the range for each of the five distributions reported. The fact that the mean plus one standard deviation is close to one end of the range clearly shows that these distributions are skewed. For example, pure-tone audiometric data are frequently reported in studies about rehabilitative audiology. Even when standard deviations are as large as the means, they have been used to describe these abnormally shaped distributions (Demorest and Walden, 1984; Wandley, 1984). When one standard deviation is as large or larger than the mean, the distribution is either grossly skewed, bimodal, or very flat. Typically, when sample data indicate that the population from which the sample is drawn is unlikely to be normally distributed, nonparametric statistics should be used. Specifically, the median should be used to estimate central location, the interquartile range should be used to estimate amount of dispersion, and the data should be converted to percentiles to facilitate comparisons.

Sometimes it is possible to normalize distributions, for purposes of analysis, by applying a transformation. For example, by applying a logarithmic transformation to a frequency scale, octave intervals are made equidistant on graph paper. However, the equation needed to normalize the hours of wear data described in the above study would be very complex, since there are different shapes at different points along the scale. Furthermore, the transformed data could not be used for interpretative purposes. In other words, transforming data can facilitate statistical analysis, but also can create data that do not exist in the real world, i.e., wear of less than zero hours and more than 24 hours a day. This illustrates the point that when data are systematically skewed, the investigator must not manipulate the skewness without considering the effects on real data.

Hays (1963) provides a thorough, readable, and balanced evaluation of the effects of violations of assumptions concerning distributional shapes and amount of dispersion. Five points stand out in Hays's discussion of this topic. Firstly, if the necessary assumptions are not grossly inappropriate, violations have only a small effect on statistical probabilities. Secondly, the effects of misapplying parametric statistics vary according to the shape(s) of the distributions, the size of the samples, and the statistic applied. Thirdly, the mathematical procedure required to calculate the exact change in probability associated with the violation of any assumption is very complex. Fourthly, when parametric tests can be applied only by violating assumptions, it is possible to come to false conclusions. Fifthly, appropriate nonparametric tests are available when the assumptions for parametric statistics cannot be met. Unfortunately, there are no exact guidelines as to what constitutes acceptable deviations from assumptions for the use of parametric tests. This is left to the discretion of the investigator. The discretion to apply ratio or interval statistics to behavioral data is one of the sources of a long existing disagreement between physical and behavioral scientists (Luce and Narens, 1987).

**SUMMARY**

Responses on self-assessment questionnaires are influenced by many physical and psychosocial variables, which vary from patient to patient. Data from existing questionnaires have the power of interval order, but not interval magnitude. Also, data from published studies suggest that distributions and dispersions of the populations do not support the use of parametric statistics. However, powerful nonparametric statistics are available for analysis of ordinal data. Within these limits, ordinal data can be used to assess patient needs, to estimate rehabilitation benefit, to monitor quality assurance outcomes, and to identify characteristics of patients who achieve desirable program goals. In order to increase precision of measurement, self-assessment scales developed in the future could use intervals defined by physical variables. Such development could lead to interval level response scales. To the degree that future developments in rehabilitative audiology could be facilitated by the advent of more precise scales for self assessment, such developments have merit.

**Acknowledgments.** The writer wishes to acknowledge, with gratitude, the fruitful discussions with Marilyn Berman during the development of this paper and her constructive comments on the manuscript, and the valuable suggestions of Patricia McCarthy and Sue Ellen Canahl.
REFERENCES


