Teaching Research: An Introduction to Statistical Concepts and Research Terminology

(schizophrenia studies, measurement)

Lillian R. Greenstein

Concerns for establishing priorities in occupational therapy research and positive responses to these concerns are reviewed. Growth in occupational therapy research can be expected, but it is important that the research process not be limited to educators and graduate students alone. Involvement of occupational therapists in scientific inquiry is often deterred by misconceptions of the research process. Realistic acceptance of research is needed to replace these misconceptions.

One major problem is understanding the language of research. It is suggested that the acquisition of research language can begin in regular occupational therapy classes. To introduce students to research concepts, frequently used statistical symbols and research terms are discussed. Limitations in applying "ideal" research methodology to clinical research are noted.

Lillian R. Greenstein, Ph.D., OTR, was Associate Professor in Occupational Therapy, Eastern Michigan University, Ypsilanti, Michigan.

Concomitant to occupational therapy growth is the increasing recognition that professional credibility requires commitment to scientific research methodology. Pleas for this commitment have been expressed in The American Journal of Occupational Therapy (AJOT) through feature articles (1), Nationally Speaking columns (2-4), Letters to the Editor (5, 6), and American Occupational Therapy Foundation (AOTF) reports (7).

King discussed the need to delineate the role of occupational therapy in psychiatric practice and warned that "whether or not occupational therapy abandons the curative role in psychiatry depends largely on the imagination and vigor with which research is pursued" (1, p 18). Johnson (2) and West (3, 4) pointed out that professional validity is dependent on the development of research support. It has been suggested (Letters to the Editor) that professional survival requires communication in "the scientific language of research" (5, p 515) and that our urgent need is for "fewer innovators and more replicators and validators" (6, p 402). The report of the AOTF-sponsored Research Seminar included comprehensive recommendations for the pursuit of occupational therapy research (7).

A wide range of positive responses to establishing research priorities has been apparent in recent years.
The August 1978 issue of *AJOT* carried multifaceted investigative reports pertaining to King’s caveat on the relationship between psychiatric research and the role of occupational therapy in psychiatry. Huddleston (8), using measures of grasp, posture, and vestibular reactivity, found that people with process schizophrenia differed significantly from a normal group in all three variables; that they differed significantly from people with reactive schizophrenia only in vestibular reactivity. Bailey (9) compared language responses of an experimental group of schizophrenic adults given sensory stimulating activities, to a control group given traditional sedentary activities. The experimental group showed significantly greater improvement in language relevance. Rider (10) showed a decrease in psychotic behavior of those diagnosed as chronic nonparanoid schizophrenic following a sensorimotor treatment program. The improved results did not continue when sensory activities were replaced by other procedures. Endler and Eimon (11) found that, whereas schizophrenic persons differed significantly from a normal group in postural and reflex integration, there was no significant difference on these variables between chronic paranoid and chronic nonparanoid schizophrenic groups.

Organizational components of the profession have also demonstrated support for needed research. The Representative Assembly of the American Occupational Therapy Association (AOTA) has allocated funds for research grants, for a research associate, and for support services. The AOTF has allocated additional grants for research support (12).

The participation of occupational therapists in the research process promises a widespread collection of empirical evidence that can offer substance to occupational therapy theory and can establish a scientific foundation for making decisions on program development. It is important that this research process not be limited to a handful of investigators, such as educators and graduate students. The active involvement of practicing therapists in the search for scientific evidence should be encouraged.

For some, the research role may be investigative, that of conducting and reporting a study. Pragmatic assistance (financial and advisory) is now available to potential investigators from AOTA- and AOTF-funded grants and services. An announcement in the *Occupational Therapy Newspaper* (13) encouraged the submission of research designs to the AOTF Research Coordinator at any stage for critique and review.

For other therapists, involvement in the research process may be as consumers who use published research results in program decision making. These therapists can rely increasingly on the scientific quality of the published reports that are used as guides in program development.

Whether as an investigator or as a consumer, it is necessary for the occupational therapist to understand and feel comfortable with the research process. Throughout the years *AJOT* has published articles that offer guidelines for understanding research. Some have discussed discrete research skills, such as those pertaining to the selection of measuring instruments (14, 15). Other articles have guided the reader through a series of steps required in understanding the entire research process (16-22).

Where does the acquisition of these skills begin? In some curricula they are taught in a separate course in the basic professional program; more often they are part of occupational therapy graduate programs. In many cases, learning these skills is an ongoing process in the undergraduate program, although it may not be recognized as such.

When students synthesize readings relevant to an occupational therapy topic, they are learning how to review literature and select writings pertaining to an area of investigation. When students analyze measures and scales used in assessment, they are learning to judge the scientific qualities of the measuring instru-
When numbers are assigned to the assessment process, students are entering the first step in learning to collect data. When students apply calculation procedures to the numbers in order to arrive at grouped information, they are learning the beginnings of data analysis. When objective and parsimonious written reports are required in coursework, students are learning the skills necessary in writing the research report.

These skills are regarded as necessary for professional education, but they are not often viewed as basic skills of research. The stereotype of the researcher as a "different" type of individual may persist, and the nuts and bolts of the research process that help to build the research study are not recognized.

Kerlinger (23) deplores three such stereotypes that create misunderstanding of scientific research: the white-coated, somewhat peculiar scientist working in a laboratory; the brilliant individual spinning complex theories in an ivory tower; and the engineering technologist inventing improvements in bridges, automobiles, and machines. Kerlinger states that such stereotypes impede student understanding of scientific research and should be superseded by more realistic notions.

Drew, too, regrets the traditional misconceptions of research; he believes that a lack of information has generated views of research as mystical and suspicious, of the researcher as solitary and strange. He points out that research is not at all mystical, that it is a "systematic method of enquiry" (24, p 6).

How can research be taught so that students will recognize it as part of professional involvement, as a searching process to be enjoyed, as a process with which they can feel comfortable? One major problem is the ability to understand the language of research (5). As the uninitiated reader of research articles confronts unfamiliar terms, there is a tendency to regard them as foreign words and phrases. A second problem that mitigates against becoming comfortable with research concepts is the identification of research with statistics, and in turn, of statistics with mathematics. There is a pervasive misconception that the whole process belongs to the exclusive domain of the mathematically elite.

To prevent these problems and to promote favorable conditions for inquiry, common research terms and beginning statistics can be introduced into regular occupational therapy classes. The mastery of these concepts then becomes an ongoing process in the same manner as the mastery of observation and writing skills are ubiquitous objectives in a number of occupational therapy classes. Thus students integrate research terms into their professional vocabulary and learn to view statistics as tools of research that they can use without being specialists in mathematics.

The following sections discuss introductory concepts in statistics and measurement terminology. Material from these sections can be extrapolated as needed for ongoing instruction in regular occupational therapy classes.

**Statistical Concepts**

When journal research articles are studied as part of course content, it

---

**Figure 1**

Frequently used statistical symbols

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>The number of cases, observations, or subjects in the entire group.</td>
</tr>
<tr>
<td>n</td>
<td>The number in a subgroup.</td>
</tr>
<tr>
<td>M, X</td>
<td>The mean or average that takes into account the values of every score in the group.</td>
</tr>
<tr>
<td>s, SD</td>
<td>Standard deviation, a measure of variability.</td>
</tr>
<tr>
<td>p &lt; .05</td>
<td>Less than a 5 percent probability that the observed results are caused by chance.</td>
</tr>
<tr>
<td>I test</td>
<td>A statistical procedure that compares two groups on one measure and determines to what degree the difference could be attributed to chance fluctuation.</td>
</tr>
<tr>
<td>t ratio</td>
<td>A statistic resulting from a t test. The probability level arrived at partially depends on the size of the t ratio.</td>
</tr>
<tr>
<td>df</td>
<td>Degrees of freedom; frequently depends on the size of the sample and/or the number of groups.</td>
</tr>
<tr>
<td>ANOVA</td>
<td>Analysis of variance; a statistical technique that compares differences three or more groups on a measure.</td>
</tr>
<tr>
<td>F</td>
<td>A statistic calculated by ANOVA procedures that, together with the degrees of freedom, determines whether or not the differences in group scores are statistically significant.</td>
</tr>
<tr>
<td>r</td>
<td>Pearson product moment correlation; a statistical technique that tests the degree of relationship of two variables.</td>
</tr>
</tbody>
</table>
frequently becomes necessary for the instructor to offer some preliminary explanation of statistical concepts. It is difficult to interpret a research report without a general understanding of common statistical symbols and procedures.

When introducing statistical concepts, it is important to avoid research anxiety on the part of students to an aversion toward statistics. To allay apprehension, it is best not to delve into formal definitions and mathematical foundations. Neither is it necessary for statistics to be dull. Feinstein states they can be “vivid, vital, and exciting” (25, p 4).

Figure 1 lists frequently used statistical symbols. The defined symbols as well as the following discussion are limited to continuous data and parametric statistics. Discrete data and nonparametric statistics can be similarly presented.

The most familiar symbols are probably those referring to number of subjects, measures of central tendency, and measures of variability. The symbol for the number in the entire group is N; for a subgroup the lower case n is used. The average score is a measure of central tendency; it can represent the scores of a group better than any other score. There are three measures of central tendency, but the most familiar and the most representative of the score values is the mean. It is attained by adding the scores of all the group members and dividing the sum by the number in the group.

The symbols s or SD refer to the standard deviation, an indicator of the variability of a group. The larger the standard deviation for a measure, the more dispersed are the scores and the more heterogeneous is the group. Variability data are necessary in understanding group performance. A group of 8 year olds with a mean IQ of 100 and with scores ranging from 90 to 110 would certainly differ from a group with the same age and the same mean, but with scores varying from 70 to 130.

A majority of statistical procedures can be classified into two general categories: procedures to identify group differences and procedures to identify variable relationships. The first category answers questions such as: Do students who take a series of short quizzes show greater mastery of subject matter than students who take two major examinations during the semester? Group A (short quizzes) is compared with Group B (two examinations).

The second category of statistics answers questions such as: Is there a relationship between grades in anatomy and grades in kinesiology? The procedure analyzes the association between grades in anatomy (variable X) and grades in kinesiology (variable Y) for each student in the group.

At one time, these two statistical approaches resulted in two different disciplines in psychology (26). Although in recent years there is a closer integration of the two procedures (27), for the purpose of explanation to beginners it helps to view them separately. Not uncommonly, novice researchers state that their purpose is to assess the degree of relationship and then use statistical techniques that compare group differences. Relationships have to do with scores on two or more variables for the same subject. Group comparisons have to do with differences in the averages of entire groups.

Group Differences. When comparing two groups, a t test is frequently used. Both groups are given a test or measure, the results of which are analyzed to determine whether Group A, which had a different treatment, scored higher (or lower) than Group B, and the nature of the difference.

Some differences in comparisons of groups are caused by chance fluctuation. In order to attribute differences to the effect of the treatment, the results must be "statistically significant." Before the study, a statement is made on what will constitute a significant difference, because surely there will be some difference in group scores. The investigator may state that a p < .05 level will indicate a statistically significant difference between the two groups. The p < .05 symbol is a probability statement. The investigator is betting that the results can be attributed to chance fluctuation less than 5 percent of the time; that 95 times out of 100 such results are not likely to be caused by chance. The implication is, then, that the results can be attributed to the effect of the treatment. A probability level of p < .01 suggests that 99 percent of the time the results are caused by the treatment.

After the investigator hypothesizes a probability or significance level, such as p < .05, he/she then applies statistical procedures to the data to find out what the actual probability level is for the collected data. A t ratio is computed to determine how large a difference there is in the scores of the two groups of measures. The attained probability level depends in part on the magnitude of the t ratio. It also depends on the degrees of freedom(df), a complicated concept that is best accepted with minimum explanation.

In a t test for independent groups, the df are two less than the number of cases. The size of t ratios required
to meet specified significance levels is related to the degrees of freedom; the larger the $df$ the smaller the ratio required to meet a specified probability statement. Each statistical technique has a formula for arriving at the degrees of freedom. In addition, with the known $df$ and the calculated ratio, appropriate tables can be consulted that indicate the attained significance level. (The tables specify the size of the ratio required to reach a specified significance level.)

If the statistical results support the wager and a probability level of $p < .05$ is attained, it can be concluded that the results are statistically significant and the type of treatment makes a real difference between the two groups. Unfortunately, it is a “winner take all” system. If the attained significance level is $p < .10$, the results are viewed as nonsignificant. In case the investigator is tempted to do otherwise, the American Psychological Association cautions:

*Do not infer trends from data that fail by a small margin to meet the usual levels of significance. Such results are best interpreted as caused by chance and are best reported as such. Treat the results section like an income tax return: Take what's coming to you, but no more.* (28, p 19)

The $t$ test can be used only to compare two groups on one measure. Wittmeyer and Stolov (29), for example, used this technique when they asked: Do spinal cord-injured patients with previous instruction identify architectural barriers more effectively than a similar group without instructional intervention?

When measures for more than two groups are compared, other procedures are used. Analysis of variance (ANOVA) is a statistical technique that can compare three or more groups on a measure. It deals with multigroup questions: for example, is there a difference in grasp strength between process schizophrenic patients, reactive schizophrenic patients, nonschizophrenic patients, and normal subjects? In this study, differences in responses of four groups are analyzed with ANOVA procedures (8).

As with the $t$ test, analysis of variance yields a ratio (the $F$ value) which, together with the degrees of freedom, determines whether the differences are large enough to be statistically significant.

**Variable Relationships.** The most common method for establishing the relationship of variables is the Pearson Product Moment Correlation procedure, often referred to as Pearson $r$ or $r$. The procedure answers questions such as: Is there a relationship (correlation, association) between the Southern California Sensory Integration Tests and the Bender Gestalt (30)? If the variables (the two tests) show a strong association, a high correlation exists. If the scores on the measures vary in the same direction (the correlation is positive; for example, subjects who score high in anatomy also score high in kinesiology. When the scores of the measures vary in opposite directions, a negative relationship exists; for example, subjects who spend more hours studying make fewer errors on a test. In this case, when the scores on Measure A (number of hours) increase, the scores on Measure B (number of errors) decrease.

The correlation statistic (coefficient) may range from -1.00, a perfect negative relationship, to +1.00, a perfect positive relationship. The plus sign is usually deleted and the results may be stated as -.80, -.85, .92, -.97. These would be regarded as strong relationships; a negative coefficient (-.78) is just as strong as a positive coefficient (.78). The smaller the coefficient and the closer to 0 (.28, -.17, 10), the weaker the correlation. As with other statistics, the correlation coefficient can be analyzed to arrive at its significance level.

**Research Terminology**

In most cases common research terms, such as those listed in Figure 2, can be simply explained. When ease in understanding is the goal, the definitions need not be complete. In this way they help destroy the mystique of research terminology and promote an unconstrained environment for learning. Frequently used research terms can be categorized into those relevant to the sample and those pertaining to the measure. **Terms Relevant to the Sample.**

**Population** refers to all the members of a specified group. “Members” of the population may be anything: people, animals, objects, scores. Members may be referred to as elements, subjects, cases, observations. The members of the group have a common characteristic(s), such as grip strength, that is to be studied by the investigator. The investigator is required to specify or define the population, since conclusions of the study are directed back to the population defined and cannot apply to groups outside this population. A study of sensory deficits in stroke patients may define the population as all patients with the diagnosis of cerebral thrombosis admitted to the occupational therapy department of Smith Hospital during the 1978 calendar year.

A *sample* is a subgroup of members selected from the population. For pragmatic reasons, such as time and
<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>population</td>
<td>All the members of a specified group.</td>
</tr>
<tr>
<td>sample</td>
<td>A subgroup of members selected from the population.</td>
</tr>
<tr>
<td>representative</td>
<td>The extent to which the sample stands for the population.</td>
</tr>
<tr>
<td>generalizability</td>
<td>The extent to which the results of the study can be applied to groups other than the one studied.</td>
</tr>
<tr>
<td>random sampling</td>
<td>Selecting a sample from a population in a way that gives every member of the population an equal chance of being included in the sample.</td>
</tr>
<tr>
<td>random assignment</td>
<td>Procedure in which chance alone determines the group to which the subject is assigned.</td>
</tr>
<tr>
<td>variable</td>
<td>Something that varies; an attribute that has different values for different members of a group.</td>
</tr>
<tr>
<td>reliability</td>
<td>The degree of consistency, stability, dependability of a measure.</td>
</tr>
<tr>
<td>validity</td>
<td>The degree to which a measure achieves its purpose.</td>
</tr>
</tbody>
</table>

resources, studies are frequently conducted on smaller groups drawn from the larger population. In many cases, and especially if the population is large, attempting to study the entire population is impractical.

Feinstein shows that using samples for decision making (sampling) is older than the science of statistics. He illustrates that "When an orchard is judged from a few of its apples, a barrel of wine from a glass, or a loaf of bread from a slice, the judge performs a sampling exercise in which the character of a whole is decided from examination of a representative part" (25, p 170).

**Representative** is a key term in sampling. It describes the extent to which the sample can stand for the population from which the sample was drawn. If the results are to be applied to the population, then the sample studied must be representative of the characteristics of the population. The results are generalizable (applicable) only to the population that is accurately represented by the sample. **Generalizability** is the extent to which the results of a study can be applied to groups other than the one studied.

In order to assure representativeness of the sample, **random sampling** procedures are used. Through random selection, each member of the population and all the characteristics present in the population have an equal chance of being included in the sample. Thus random sampling reduces the possibility that a biased or unrepresentative subgroup will be drawn from the population (24).

Any selection procedure that is based only on chance may be used: drawing names from a container, drawing well-shuffled cards, throwing dice. However, the most effective procedure is assigning numbers to all members of the population and using a table of random numbers for sample selection.

When groups are formed for comparison, it is important that varying characteristics are distributed as evenly as possible to each group. To help accomplish this, members are assigned randomly to groups; in this procedure **random assignment** is used. Random assignment is a feasible procedure for occupational therapy studies and can be accomplished even with small groups. Wittmeyer and Stolov (29), for example, selected ten spinal cord-injured patients for their study and randomly assigned them to experimental and control groups.

It is important for the occupational therapy student to recognize the above terms in research methodology. It is also important to recognize the limitations of "ideal" scientific sampling in the real world of clinical practice.

In most occupational therapy studies, random sampling from a larger population is rarely feasible. Studies are usually accomplished with available intact groups. Since the patient groups selected often have specified characteristics that are different from the larger patient population, the subjects studied are not a representative subgroup of the population. In each of the psychiatric studies cited above (8-11), subjects were selected because of specified schizophrenic characteristics.

Limitations of scientific sampling are not confined to occupational therapy studies. Feinstein states that:

*One of the most pernicious scientific delusions now prevalent in the world of medical research is the idea that concepts of "random sampling" can be readily applied to clinical populations.* (25, p 23)

Kerlinger recognizes that in behavioral research random samples are "hard to come by" and that "accidental" sampling or samples at hand are often the only subjects...
available. He suggests that this practice be used with “reasonable knowledge and care” and with caution in the interpretation of the results (23, p 129).

Occupational therapists generally abide by Kerlinger’s advice. The subjects of occupational therapy studies are carefully described in most occupational therapy reports. Huddleston (8) specifies the diagnosis, age range, and sex of the members of the four groups compared in her study. Bailey (9) describes the diagnosis, years of illness, average age, and average range of the subjects in her experimental and control groups. In analyzing results, occupational therapy investigators exercise conservative interpretations. Rider, after noting patient improvement in her study, cautions that “the study was too small to be conclusive” (10, p 455). Endler and Eimon (11), with results suggesting sensory integrative therapy to be a useful treatment with chronic schizophrenic patients, point out limitations in generalizing these findings.

Notwithstanding the sampling limitations that are necessarily present in occupational therapy studies, these investigations offer an important contribution to the professional body of knowledge. Albeit they have recognized limitations, repeated studies indicating favorable results after using sensory integrative techniques with a specified group of psychiatric patients offer valuable evidence in support of sensory integrative theory.

Terms Relevant to the Measure. The definition of a population specifies that the members of the population have a common characteristic (anxiety, verbal response, grip strength). Data describing this characteristic or variable are analyzed in the research investigation. The term variable is not limited to the one characteristic; it can cover a multitude of attributes.

Formal definitions refer to a variable as “a symbol to which numerals or values are assigned” (25, p 29), and as “a property whereby the members of a group or set differ one from another” (31, p 11). Stahl and Hennes (32) simply define a variable as something that varies although it was quite consistent in measurement.

The occupational therapy literature (14-22) stresses the importance of evaluating the reliability and validity of measures used in occupational therapy assessment. Reliability refers to the degree of consistency or stability of the test. If the test is given two different times to the same group, and there is no intervening practice or other intervening variables, will the two scores be similar for each member of the group? A simplified way of understanding the consistency requirement for reliability is using the example of a weight-conscious student who finds that four different readings on the same scale during the day show 115, 115¾, 115¼, and 115 pounds. The slight differences in the readings can be attributed to food consumption, and it is concluded that the readings are quite consistent and the scale is reliable. However, if the readings fluctuated from 110 to 120, the scale would be judged as an unreliable measurement of weight.

The validity of a test indicates the degree to which the measure achieves its purpose. A question on test validity might be: Does a student interest inventory measure the real interests of the student or are responses influenced by some other variable, such as social desirability?

If the four readings on the scale described above were 100, 101, 101, and 101 pounds, and the student really weighed 115 pounds, it would be concluded that the scale was not achieving its purpose. It was not a valid assessment instrument, although it was quite consistent in its measurement.

How do occupational therapists deal with variables, reliability, and validity in their research reports? Variables used in studies are generally carefully discussed. Huddleston (8) specifies the variables of grasp strength, posture, and vestibular reactivity and shows how each variable was measured. Endler and Eimon (11) discuss the three groups of variables measured in their study and list them in table form.

Reliability and validity information on measures, however, are rarely reported. In the four studies on schizophrenic patients (8-11), the reliability of only one instrument is discussed (11). The absence of this information may stem from the writer’s unquestioned faith in the instrument, or it may be attributed to problems in instrument development. Many measures used in the health professions have not been subjected to reliability and validity studies. The problems of ascertaining reliability with clinical measures are similar to the problems of representative sampling; they stem from the complexities of human subjects. Feinstein (25) recognizes reliability problems, but cautions
against the substitution of measures with proven reliability for clinically important measures that are more relevant, but less reliable. He suggests that the answer is in the "development of better methods of observation and classification" when clinically relevant soft (less reliable) data are used (p. 49).

Occupational therapy studies acknowledge the problems of soft data and try to improve the quality of the measure by carefully planning the procedures used with the measure. Rider (10) notes the subjective nature of the Draw a Person Test and analyzes the inter-rater agreement as a contribution to reliability. Huddleston (8) meticulously describes the step-by-step procedures used in measuring vestibular reactivity.

The terms described in this section and listed in Figure 2 provide an introduction to research terminology. The discussion is necessarily incomplete and does not include concepts such as scales of measurement and classes of variables.

Conclusions and Summary

Discussions during the 1976 AOTF Research Seminar indicated that barriers to research production in occupational therapy could be grouped as attitudinal, organizational, and material (7). The classroom may well be the appropriate environment for the development of realistic research attitudes.

Since research is a method of inquiry, an inquisitive attitude is a prime ingredient. Yerxa and Gilfoyle (7) point out that curiosity should be encouraged. A basic understanding of introductory statistical concepts and frequently used research terms is a necessary prerequisite for stimulating curiosity, inquiry, and a realistic interest in the research process. Interpretation and discussion of these research tools can become part of regular classroom instruction in the same manner as instruction in writing skills permeate most occupational therapy classes. As research tools are mastered, they should be recognized as components of research. Thus, students become comfortable with realistic views of scientific inquiry and a receptive attitudinal foundation is established for occupational therapy research.

Acknowledgments

Appreciation is extended to Barbara Barker, Ph.D., and Harry Barker, Ph.D., who instilled in their students an enthusiasm for research and statistics.

REFERENCES

13. Call for research papers. Occup Ther Newspaper. July 1979, p. 6