THE ISSUE

A Response to
Testing and Measurement in Occupational Therapy:
A Review of Current Practice with Special Emphasis on the
Southern California Sensory Integration Tests

Elizabeth J. Yerxa

Criticism of occupational therapy rarely appears in the literature. Critical thought can either be seen as a stimulant to knowledge development, dialogue, and emerging consensus, or an irritant that must be encapsulated by defensive reactions in order to be forgotten as soon as possible. I wish to respond to criticism that appeared recently in "Testing and Measurement in Occupational Therapy: A Review of Current Practices with Special Emphasis on the Southern California Sensory Integration Tests," by Patricia Evans, a doctoral candidate in educational psychology at the University of Minnesota, and Mary Ann Peham, a law student, both of whom are registered occupational therapists (1). Their treatise will be referred to as the Monograph.

I will examine several issues raised in their monograph in the spirit of encouraging reflection that could contribute to the further development of the knowledge base supporting the occupational therapy profession. This discussion will concern 1. the audience to whom the monograph is addressed, 2. the quality of scholarship demonstrated, 3. the semantic content of the language employed, 4. the standard of test reliability adopted, and 5. the generalizations made about occupational therapy education.

Audience
Scholarly works are usually written for the purpose of communicating to a particular audience. A review of the monograph results in some confusion or ambivalence about the constituency of the intended audience. In view of the publisher (the Institute for Research on Learning Disabilities) and the focus of content of the other monographs and research reports published by the Institute, one can assume that the intended audience was special educators and educational psychologists interested in learning disabilities. This observation is supported by the fact that neither author used the designation "OTR" after her name on the title page, although both were identified as "registered occupational therapists" in the footnote on page 42. One is therefore surprised to find, on page 30, the first reference to occupational therapists as "we" when previously, occupational therapists were referred to as "they" (1). Three other references to "we" or "our" profession followed.

If the authors meant this monograph to prod occupational therapists (as it certainly could), why was it published outside the mainstream of professional literature by authors who were not initially identified as occupational therapists? Why did the authors wait until their criticism was almost complete before demonstrating a desire to change "their" profession? A reasonable conclusion seems to be that the monograph was addressed primarily to special educators, members of a discipline that may overlap with occupational therapy, especially under the provisions of PL 94-142. This work could thus be used as the basis for arguing the need for occupational therapists in the schools.

Scholarship
A risk one takes in leveling criticism is to expose one's work to assessment by the same criteria. This "persons who live in glass houses..." principle will now be applied to the quality of the monograph's scholarship just as the authors criticized the breadth and depth of occupational therapists' knowledge.

First, the authors overgeneralized from their criticisms of the Southern California Sensory Integration Tests (SCSIT) conclusions regarding the entire occupational therapy profession and its knowledge base. For example, they stated, "occupational therapists do lack knowledge in the area of testing and measurement." (1, p 32) This generalization is so all-encompassing as to be meaningless and seems a nonsequitur from

Elizabeth J. Yerxa, Ed.D., OTR, FAOTA, is Associate Professor and Chairperson, Department of Occupational Therapy, University of Southern California, Los Angeles, California.
the body of the monograph, which consists of criticism of the SCSIT and their interpretation. No other instruments used by occupational therapists were mentioned in spite of the fact that the SCSIT are employed by a minority of occupational therapists. The monograph should have been titled "... with Total Emphasis on the SCSIT," rather than "... with Special Emphasis. . . ."

Such a general statement about lack of knowledge could be applied to members of almost any discipline since knowledge of the theory and practice of tests and measurements is becoming increasingly sophisticated and specialized. Every instrument developed, no matter how conscientious its author, has weaknesses and limitations. Tests are developed and refined over time, just as theories are.

The reader may well question the validity and appropriateness of the quotations and citations used to support the authors' point of view. For example, I was cited as questioning the future of the profession and identifying the negative impact of "educational and political issues" as evidence of the need for the profession to engage in research (2). Thus the authors implied that occupational therapists "misuse" tests. The authors cited their "clinically honest." (6, p 1292) Kephart, a recognized authority in the field of learning disabilities, stated, "In my remark was the need for self-assessment (1, p 34). I certainly believe in the value of self-assessment, but the actual target of my remark was the need for occupational therapy faculty members to engage in research (2). Thus the statement was taken out of context and used to convey an erroneous and unintentionally pessimistic perspective of the profession.

Objectivity in tone and caution about reaching unfounded conclusions are hallmarks of scientific writing. The monograph would have received a "D" had it been written for a graduate course in occupational therapy. It presented only one side of some very controversial issues and completely ignored different views that might be held in equally good faith. For example, psychometrically oriented criticism of the SCSIT covered 18 pages, or 49 percent of the monograph's content. Much of this criticism centered upon the content of the SCSIT test manuals published between 1962 and 1976; however, the most recent manual, published 9 months before the monograph appeared, was not referenced (5).

The authors cited three reviews of the SCSIT that appeared in Buros' Seventh Mental Measurements Yearbook (4) and two from the Eighth (5), all of which were negative, or unfavorable. The reader is thus led to believe that only negative reviews exist. However, one finds these comments in Buros: "Ayres' new test . . . SCPMT appears to be the well-thought-out product of an acknowledged expert in special education." (6, p 1291) "On the whole, the manual of the tests is high in quality and psychometrically honest." (6, p 1292) Kephart, a recognized authority in the field of learning disabilities, stated, "In spite of its shortcomings, the SCKT can provide important information for perception in general. It provides this information in an area where few measures exist and where major problems of perception probably have their origin. It deserves expansion and intensive study." (7, p 1289) In omitting these positive reviews, the monograph presented a negatively biased view of expert opinion regarding the quality of the SCSIT.

Evans and Peham cited nine references from the occupational therapy literature as evidence that the SCSIT "is (sic) used with populations other than that for whom it was intended." (1, p 14) In reviewing the same articles, I found that Clark and colleagues did not use the tests; they employed one sensory integrative technique along with other methods in the treatment of profoundly retarded adults (8). King also did not employ the SCSIT, but rather, hypothesized about the presence of sensory integrative deficits in adults who had chronic schizophrenia (9). Thus, the conclusion that the tests have been used with schizophrenic and profoundly retarded adults is untrue and leads the reader to question the accuracy of other citations. It is also noteworthy that De Pauw (10), referenced as one who does not interpret the SCSIT in the correct manner, is a developmental and remedial physical education teacher, and not an occupational therapist.

It is generally considered appropriate to provide evidence to support conclusions, especially those crucial for validating an argument. For example, on page 4 the authors implied that occupational therapists "misuse" tests. Page 5 mentioned that data collected by occupational therapists may be used for "mislabeling" children. On page 15, Evans and Peham contended that individual SCSIT scores are "often" erroneously used in clinical interpretation. The authors cited their "clinical experience" as the basis for stating that occupational therapists do not apply the standard error of measurement to individ-
eral scores as recommended by Ayres. All of these generalizations were made without the provision of any empirical evidence even though they were used to raise serious questions about occupational therapists' practices and ethics.

In writing a critical review, an author is expected to reflect the current thinking of experts or authorities in the field being reviewed. The more scientists can demonstrate that their views are supported by a consensus of experts, the more credible are their generalizations about scientific approaches (11). In discussing educational measurement as applied to children who have disabilities, the authors included 21 citations of Ysseldyke's work originating from the University of Minnesota's Institute for Research on Learning Disabilities (the publishers of the monograph). The excessive reliance on this "home-based" author is demonstrated by the fact that only one other author (Cronbach) was cited in discussing standards of measurement as applied to handicapped children. One wonders why other widely cited experts on testing such as Anastasi or Thorndike were not cited and why Cronbach was overshadowed so completely by Ysseldyke. Perhaps Ysseldyke directed this work but was loath to appear as a coauthor since criticism from within a discipline (occupational therapy) might seem less politically motivated than that emanating from without (special education).

In discussing "standardization of tests" on page 13, the Wechsler Intelligence Scale for Children was cited as a "well-standardized" test and the Frostig Test of Visual Perception and the Purdue Perceptual Motor Survey as "poorly standardized." The reader was told nothing more. Naturally, employing the tenets of scholarship, that statement must be classified as opinion, rather than a conclusion based upon verifiable evidence since no evidence whatsoever—not even a single reference—was provided to back it up.

On page 29, the authors stated that there is "inadequate empirical support" for the idea that underlying deficits cause academic difficulties as posited by tenets of sensory integration and other theories. The source for this view was none other than Ysseldyke and Salvia (12). Only one of Ayres' empirical studies, that published in 1972, was referenced of six relating to this topic published up to 1978. In addition, the criticism of "inadequate empirical support" could be leveled at just about every theory developed that attempts to establish causality in relation to human learning.

Finally, an innuendo appeared on page 31 in which the AOTA Principles of Occupational Therapy Ethics were cited to support the idea that occupational therapy researchers ought to use "accepted scientific methodology." However, no evidence was presented to demonstrate that this principle has ever been violated by Ayres or any other occupational therapy researcher. It should also be noted that honest differences of opinion exist in many fields about what constitutes "accepted scientific methodology," including disciplines as diverse as psychology, education, biology, and medicine (13-16).

In summary, the scholarship of the monograph does not meet acceptable standards since it overgeneralized unbalanced criticism of the SCSIT to the practices of the entire occupational therapy profession, employed quotations out of context, leveled unsupported accusations regarding misapplication of the SCSIT, presented negatively biased criticism of the SCSIT from Buros' Mental Measurements Yearbook, made unwarranted generalizations about the misuse of the SCSIT by occupational therapists, demonstrated overreliance on a single author as a psychometric expert, presented opinions rather than facts regarding the quality of test standardization, ignored the majority of recent research published in the areas of sensory integrative theory, and hinted at unethical practices by occupational therapists without providing evidence that those practices do, in fact, exist.

Semantic Quality of Language

A scientific paper is required to maintain an objective and unbiased tone so that evidence, and not emotion, is the basis for drawing conclusions. The words employed in this monograph frequently failed to meet that standard and created confusion about whether the monograph was intended to represent fact or persuade via rhetoric. Italics have been used by this author to identify words that seem emotionally loaded. For example, on page 15 the statement appeared that "... These published reports of alleged misuse of the SCSIT undoubtedly represent only a fraction of the instances of clinical misuse." On what basis was this an "undoubted" conclusion? Such a statement is inconsistent with a primary principle of science,
namely, that conclusions must be supported by verifiable data. This is particularly crucial when accusing professional colleagues of "clinical misuse" of tests. On page 17 a fact regarding the SCSIT's standardization sample was called highly disturbing. Page 20 stated that "poor reliability is further rationalized." "It is a sad note" was seen on page 30 referring to the SCSIT Manual's alleged failure to follow Ayres' procedures. Page 33 stated that "Obviously, thousands more therapists administer this (sic) test" than the 800 certified by CSSID. On what basis was this statement obvious? On the same page occupational therapists were classified as psychometric innocents and illiterates, accused of indiscriminate acceptance of tests, and by extension as operating under license to misuse tests and to use inadequate tests.

On page 94 the statement appeared that "the professional credibility of occupational therapists is not well-established, as suggested by reimbursement difficulties." These "difficulties" were never made explicit nor were they documented. Thus no support was provided for sweeping statements referring to the practice of more than 35,000 occupational therapists and certified occupational therapy assistants. On page 97, the reader was left with a final warning "...a reluctance or refusal to recognize and respond to these concerns from within the profession by the profession may result in societal mandates for changes in our professional practices." In other words, if occupational therapists (we) don't clean up their (our) act, some outside authority will dictate their (our) practices.

Were the authors now counting themselves among the observers or the observed? If these "concerns" were aimed at inspiring action from within the profession, why were they published outside of the profession, particularly in a monograph that is likely to be distributed to groups having territorial anxieties concerning occupational therapists' assuming new roles in the school systems? In summary, the semantic content of this monograph reflected an emotional tone out of place in a scientific document. Facts and data are the stuff of science, not affectively loaded words that have an accusatory and judgmental tone. The frequent substitution of emotion for evidence was particularly damaging to the monograph's credibility.

Proposed Standards of Reliability

On page 9, the monograph quoted Nunnally (1) in establishing the range of reliability coefficients for the use of tests for decision making at .90 (minimum to be tolerated) to .95 as a "desirable standard." This standard was presented as if it should be applied to all categories of reliability. Reliability coefficients of the majority of research instruments used in occupational therapy would probably not achieve the minimum standard of .90 as is true of most of the tests included in the SCSIT. Does this mean that such instruments are too unstable or inconsistent to be employed in serious endeavors like research or diagnosis?

If Nunnally's standard of reliability, accepted in the monograph, were applied to all tests, the use of many venerable and almost universally accepted instruments would be eliminated. For example, the Wechsler Intelligence Scale for Children—Revised (WISC-R) demonstrates subtest scaled scores with median across all ages and subtests at .77 (5). Tests that measure characteristics of human beings, particularly children, are likely to have lower test-retest and split halves reliability coefficients than the .90 to .95 Evans and Peham have adopted as a standard, partly because of the imperfect science of test construction and partly because of variability in children's attention, moods, or actual performances from moment to moment. For example, in The Assessment of Psychopathology and Behavioral Problems in Children, 44 scales were evaluated as "suitable for epidemiological and clinical research with children" (17). Ten instruments selected to measure specific syndromes (such as hyperactivity) reported a wide range of reliability coefficients (.61 to 1.0) with five out of ten in the .61 to .78 range. As is usually the case, higher coefficients were reported for interrater than for either test-retest reliability or internal consistency. All of these instruments were considered sufficiently reliable for research use by authorities from the National Institute on Mental Health regardless of the ranges of reliability reported.

A comparison of the SCSIT with other perceptual motor assessments reviewed in the Eighth Mental Measurements Yearbook shows that several published and widely used instruments do not even report reliability coefficients. Of those that do, test-retest coeffi-
cients were: Primary Visual Motor, .82; Developmental Tests of Visual-Motor Integration, .80 to .90; Spatial Orientation Memory Test, .60 to .74 (median .62); Motor Problems Inventory, .77; and Frostig Movement Skills Test Battery, .44 to .88 (median .60).

Thus it is evident that few instruments available to measure perceptual-motor functions in children reach the standard of reliability demanded by Evans and Peham. Efforts need to be made to develop and validate more reliable instruments, if possible, to measure all perceptual-motor functions. Until this is done the reliability of the SCSIT should be compared with that of other published perceptual-motor tests as well as tests with a longer history of development.

To summarize, standards of reliability employed in the monograph, although well worth striving toward, are unrealistically high when applied to the gamut of tests in the perceptual-motor area. Establishing a range of .90 to .95 for test-retest reliability and internal consistency would eliminate almost all measures currently used to assess characteristics of children, including the WISC-R, one of the most widely used tests of intelligence.

Generalizations about Occupational Therapy Education

After presenting unsupported conclusions regarding the ineffectiveness and misapplication of the SCSIT and generalizing these to the entire occupational therapy profession, the authors stated on page 30 that “occupational therapy curricula (sic) do not provide adequate preparation in statistics and measurement; background in the areas of perception and sensation is similarly lacking.” Evans and Peham seem to be criticizing all occupational therapy curricula. It is obviously not feasible to “adequately” prepare anyone in all of these areas at a baccalaureate level; however, the authors provided no recognition of the existence of 22 graduate programs in occupational therapy, many of which do provide such background. Again, on page 32, only baccalaureate and continuing education programs were mentioned when the authors criticized occupational therapists’ minimal background in test and measurement theory. On page 36, while it was acknowledged that “expanding the knowledge base for students is not a simple task” and that “assessments may be more suitably learned at the graduate level than at the bachelor’s level,” graduate education in occupational therapy was not mentioned.

Were the authors unaware of graduate occupational therapy programs, or did they choose to ignore them because educational preparation for entry into the profession remains primarily at the baccalaureate level? In the monograph, the authors concluded that occupational therapists, taken as a whole, are undereducated for understanding the theory, interpretation, and limitations of tests.

Discussion

Think, for a moment, of the most recent political speech you have heard. Chances are that it was characterized by some or all of the following: sweeping generalizations; conclusions based on little, if any, evidence; emotionally loaded words; selective presentation of only one side of the issue; erroneous quotations or quotations lifted out of context; application of an impossible standard as though it were the norm; excessive reliance on one source for “facts” or opinions; and innuendo regarding the ethics of the “opposition.”

These qualities are also found in the monograph (shown above). It seems evident that this document is unfortunately an imposter, because it is really a political statement posing as a scientific critique.

In spite of the criticism that can so easily discount the validity of the monograph, as a political treatise it will be available as ammunition for use by those who wish to discredit the roles and functions of occupational therapists as they move into new areas of practice. Our profession is vulnerable to such political attacks, particularly to the charges of undereducating occupational therapy entry-level students since we deem them ready to enter the field with only a few years’ professional preparation. As demonstrated, Evans and Peham overgeneralized to the entire occupational therapy profession as if it were represented totally by occupational therapists who employ sensory integrative therapy. They also leveled erroneous and unsupported criticism at the latter group. This author, however, believes that the question of the adequacy of educational preparation remains a valid one that should be looked at seriously by every occupational therapist.

Occupational therapy knowledge is developing at an explosive rate, while at the same time new roles and functions, requiring highly sophisticated knowledge, are being performed by large numbers of occupational therapists. A science of human occupa-
tion is being developed that deals with understanding and developing one of the most complex phenomena known to humankind—the adaptive capacities of persons with chronic and often severe disabilities. This science has remarkable breadth and depth since it encompasses all intentional action, with all disabilities across all age groups.

In order to deal with the complexity of civilization and its institutions, occupational therapy students need a strong liberal education as the foundation for professional knowledge. This process requires at least 2, and possibly 3, years at the undergraduate level, leaving only 1, or at the most, 2 years to develop both the knowledge and practical skills to deal effectively with extremely complex issues. As the need for knowledge and skills grows, curricula are expected to add more and more content to the undergraduate course of study. For example, entry-level programs were asked recently to add content on research and administration to the already overloaded curriculum. Educators wonder how much more can be packed in before the students and faculty, like overstuffed sausages, begin to burst from sheer overload. However, since all of this content is perceived by the profession as basic or entry level, the curriculum is limited in its ability to develop the advanced conceptual abilities necessary for critical thinking, that is, analysis, synthesis, and evaluation. Evans and Petham’s work, with all of its deficiencies, can only be employed politically to threaten us because our profession is vulnerable at the level of developing and transmitting knowledge. We still consider education at the baccalaureate level sufficient for entry into and practice of the myriad of sophisticated and autonomous roles that increasing numbers of occupational therapists occupy.

What alternatives exist for the occupational therapy profession in responding to the unscholarly but politically potent accusations of these authors? We can encapsulate ourselves and try to forget these issues as soon as possible. We can respond by criticizing the scholarship of the authors, discrediting their work, and letting it go at that. Or, we can give serious consideration to the overriding issue that makes such a political document possible. Is education at the baccalaureate level adequate preparation for the roles and functions that society and the profession increasingly expect of occupational therapists? Does it allow for the transmission of what needs to be known about the science of human occupation including how to criticize it, practice it with a healthy skepticism, and go on to improve upon it? If the monograph stimulates discussion of these questions within our discipline, then it is worth the $3.00 price tag and the time necessary to read it.

REFERENCES