

# Field Testing the Hand Therapy Certification Examination

## **Mary C. Kasch, OTR, CVE**

Director, Hand Rehabilitation  
Center of Sacramento  
President, Hand Therapy  
Certification Commission, Inc.,  
Sacramento, California

## **Russelyn Roby Shindoll, MA**

Senior Test Specialist  
Professional Assessment Services  
American College Testing, Iowa  
City, Iowa

## **Rachel K. Adams, Ph.D.**

Program Associate  
Professional Assessment Services,  
American College Testing, Iowa  
City, Iowa

**ABSTRACT:** In 1985, the American Society of Hand Therapists (ASHT) conducted a role delineation study (occupational analysis) of hand therapy. The data from the study established the basis for a formal process of certification in hand therapy. As a result, a test blueprint was developed, and test items were constructed. The focus of this paper is on the development and results of eight field tests conducted in 1989 and 1990 to select items for the inaugural examination in 1991. The items with the most appropriate level of difficulty and which conform to the test specifications will be selected for the examination.

The paper includes a discussion of the use of data derived from the role delineation study to influence the test blueprint. A description of item development and item review follows. Overall, the results of the field test were very good. Over 50% of the 600 items field-tested have statistical indicators in the appropriate range for use on the certification examination. The mean scores and reliabilities of the eight different field-test forms were quite similar, indicating that the items were similar in content and difficulty level. The items were fairly difficult. This result, however, is desirable, since the field-test examinees had not yet prepared for the certification examination. Subgroup scores based on demographic information collected from field-test examinees generally supported the appropriateness of the items for advanced competency certification. These results are presented in detail and the implications for the certification program are discussed.

## ROLE-DELINEATION STUDY

### Purpose of the Role Delineation Study

In 1985, the American Society of Hand Therapists (ASHT) conducted a role-delineation study of hand therapy. Chai, Dimick, and Kasch outlined the method used and the results of the study in the inaugural issue of *Journal of Hand Therapy*; a detailed account of the study and its results are discussed in that article.<sup>1</sup> Briefly, however, the study included four populations: (1) members of the American Physical Therapy Association (APTA) Special Section on Hands; (2) members of the American Occupational Therapy Association (AOTA) Physical Disability Special Interest Section (those who worked primarily in hand therapy); (3) ASHT members; and (4) other therapists who subscribed to ASHT's supplemental mailing list.

A self-report questionnaire divided into four sections was the survey instrument for the study. The purpose of the survey was to collect demographic information on the role of activities, skills, and modalities (ASMs) in hand therapy, and to obtain information on major categories of theory and knowledge fundamental to hand therapy practice. A fourth section of the study requested therapists' opinions

concerning the necessity for a certification program in hand therapy.

Ultimately, the results of the role-delineation study served to define hand therapy and its scope of practice, provided a focus for educational objectives, supplied a data base for further research, and established the empirical basis for a formal process of certification in hand therapy.

The first step in the design of a standardized testing program is the development of a test blueprint. A test blueprint is a guide for building the examination. That is, the blueprint prescribes exactly how many test items should cover topic A, how many test items should cover topic B, and so forth on a specific form of an examination. The test blueprint for the certification examination in hand therapy is shown in Figure 1. After the test blueprint is developed, the process of item development, item review, and examination construction may proceed.

## ITEM DEVELOPMENT

### Item-Writing Workshop

The purpose of an item-writing workshop is to train content experts in writing and reviewing test items. An item-writing workshop is designed to be a supportive and efficient means for producing large numbers of high-quality test items. Compared to items written by untrained item writers, test items generated at an item-writing workshop are generally clas-

This paper was presented at the American Society of Hand Therapists annual scientific meeting, Toronto, September 24, 1990. Correspondence and reprint requests to Mary C. Kasch, OTR, CVE, Hand Rehabilitation Center of Sacramento, 1020 29th Street, Suite 250, Sacramento, CA 95816

**EXAMINATION BLUEPRINT**

**KNOWLEDGE DOMAINS**

<b>ASM DOMAINS</b>		<b>ANATOMY</b>	<b>HISTOLOGY</b>	<b>PHYSIOLOGY</b>	<b>KINESIOLOGY</b>	<b>PHYSICS</b>	<b>SURGERY</b>
1.0 Wounds and Scars							
No. of items:	27	7	7	7	0	0	6
Percent:	13%						
2.0 Edema							
No. of items:	12	3	3	3	0	0	3
Percent:	6%						
3.0 Pain							
No. of items:	10	3	2	3	0	2	0
Percent:	5%						
4.0 Neurovascular & Neuromuscular							
No. of items:	38	6	6	7	7	6	6
Percent:	19%						
5.0 ROM							
No. of items:	29	6	5	6	6	0	6
Percent:	14%						
6.0 Dexterity							
No. of items:	27	6	0	7	7	0	7
Percent:	13%						
7.0 Pros/Orthotics							
No. of items:	16	3	3	3	4	0	3
Percent:	8%						
8.0 Work							
No. of items:	7	0	0	3	4	0	0
Percent:	3%						
9.0 Tech/Modalities							
No. of items:	34	3	3	8	8	8	4
Percent:	17%						
<b>TOTAL</b>	<b>200</b>	<b>37</b>	<b>29</b>	<b>47</b>	<b>36</b>	<b>16</b>	<b>35</b>
	<b>100%</b>	<b>18%</b>	<b>14%</b>	<b>23%</b>	<b>18%</b>	<b>8%</b>	<b>17%</b>

**FIGURE 1.** Blueprint of the Hand Therapy Certification Examination.

sified correctly according to the test blueprint and are more likely to survive the item review process.

The primary focus of the item-writing workshop is training. The item-writing workshop conducted for the certification examination consisted of an item-writing training session led by a measurement specialist and an experienced medical editor from American College Testing (ACT). The training session consisted of group instruction on how to avoid common item-writing flaws, such as lack of focus and possible item bias. A series of group discussions and one-on-one instruction followed, in which item writers scrutinized sample items, reviewed items written by other item writers, and received feedback from ACT staff on how well items conformed to item-writing principles.

To ensure that all areas of the test blueprint are well-represented, item-writing assignments are necessary. Assigning specific topics to item writers helps to ensure that the item bank includes a variety of items in all topics covered in the test blueprint. As

a result of the first item-writing workshop, item writers produced over 500 high-quality test items.

### Item Edit and Review

After the items were written, they were sent to ACT for thorough examination by a staff of experienced medical editors and measurement specialists. The editorial staff followed the guidelines printed in the *AMA Manual of Style* when reviewing the items for clarity, conciseness, and accuracy of technical terminology. For items that were incomplete or for some reason did not conform to item-writing principles, a query was recorded, and the item was forwarded to a content expert/item reviewer. In addition, items were reviewed for grammatical accuracy, freedom from possible bias, option homogeneity, and logic of construction.

After the items were reviewed by ACT staff, the Examination Committee of the Hand Therapy Cer-

tification Commission (HTCC) conducted two item review meetings—one in October 1988 and a second in March 1989—to review the technical accuracy of the items. The item review meetings provided a forum in which the Commission could review the items to ensure that all items were appropriate in both form and content. After the item review meetings, ACT staff incorporated all suggested revisions to the items in preparation for producing the field tests.

## CHARACTERISTICS OF THE FIELD TESTS

### Composition

A field test is a mechanism by which test items may be "tried out" to obtain information about item difficulty and discrimination before the items appear on a scored examination. The Examination Committee of the HTCC requested that field tests be conducted to determine the quality of the test items before they appeared on the certification examination. Each field test contained only 75 randomly selected items to ensure that only a sampling of the items would be seen by any one group of individuals. Any one of the field tests contained items that may appear on the 1991 certification examination. However, the proportion of items on the field tests did not represent the proportion required by the test blueprint. Those items that are found to have the most appropriate level of difficulty and that conform to the test blueprint will be selected for the 1991 certification examination.

The field tests also provide an opportunity to "try out" a variety of illustrations in a number of test items. Some illustrations have been used as item options and some have been used in conjunction with the stem (question) of the item. These illustrations have been provided to avoid geographic bias; that is, some devices may be routinely identified by different names in different parts of the country.

### Administration

Eight different field tests were administered at four different sites, beginning in November 1989 and ending in May 1990. The first field test was administered on November 10, 1989 in Baltimore, Maryland. The second field test was administered on January 26, 1990, in Atlanta, Georgia. The third field test was administered on March 30, 1990, in Los Angeles, California, and the fourth test was administered on May 4, 1990 in St. Louis, Missouri. The field tests were timed to simulate an actual examination. Examinees were given 90 minutes to complete their field test. The results of the field tests are reported in the next section.

## RESULTS

### Demographics

Examinees were asked to complete a demographic survey as a part of the field tests so that the

Hand Therapy Certification Commission could identify the salient characteristics of field test participants. The HTCC also needed data for planning the 1991 examination, since all the numbers used prior to field testing were "best guesstimates" made by the HTCC.

Table 1 summarizes the demographic data collected. A total of 839 examinees took the field test at the four test sites. Of the examinees completing the demographic survey, a total of 89% indicated they were occupational therapists and 11% indicated they were physical therapists. Ninety-six percent of the respondents indicated certification or licensure prior to 1985. A little more than half of the respondents had maintained a 50% or higher case load of upper extremity patients more than 6 years, whereas 96% had maintained a 50% or higher case load for 2 years or more. Most of the respondents (96%) planned to take the certification examination in 1991 at the same site as their field test. Additional sites have been added, because the data from the field tests indicated that more therapists would take the certification examination than had been anticipated by the HTCC. Given the clinical nature of hand therapy practice, it

TABLE 1. Field Test Survey Results

	FRE- QUENCY	PER- CENT
1. Do you plan to take the Certification Examination in Hand Therapy in 1991?		
(0) No	31	4 %
(1) Yes	790	96 %
2. If you plan to take the examination in 1991, which test site will you attend?		
(0) Baltimore, Maryland	201	25 %
(1) Atlanta, Georgia	171	21 %
(2) Los Angeles, California	266	33 %
(3) St. Louis, Missouri	163	20 %
3. Are you an occupational therapist (OT) or a physical therapist (PT)?		
(0) Occupational therapist	736	89 %
(1) Physical therapist	90	11 %
4. If you are a certified OT or PT, when were you certified?		
(0) 1970 or earlier.	69	9 %
(1) 1971-75	113	15 %
(2) 1976-80	243	32 %
(3) 1981-85	307	40 %
(4) 1986-90	36	5 %
5. Please indicate the number of years you have maintained a case load of 50% or more upper extremity patients.		
(0) More than 15	37	4 %
(1) 15 years	110	13 %
(2) 6-10 years	306	37 %
(3) 2-5 years	350	42 %
(4) 1 year or less	21	3 %
6. Please indicate your primary area of practice (select one).		
(0) Clinical	777	94 %
(1) Administration	39	5 %
(2) Academic	4	0.5%
(3) Research	5	0.6%
7. Are you currently a member of the American Society of Hand Therapists?		
(0) No	614	75 %
(1) Yes	209	25 %

was not surprising that most of the respondents (94%) described their primary area of practice as clinical. Of the remainder, 5% indicated administration and less than 1% each indicated academic or research as their primary area of practice.

The examinees were also asked if they were members of the American Society of Hand Therapists. The Society has been the primary professional association for hand therapists since it was founded in 1977, and it has played an active role in developing the HTCC. There are many therapists engaged in upper extremity practice, however, who are not members of ASHT. The HTCC wanted to know how many therapists would be interested in hand therapy certification outside of ASHT, as well as what percent of ASHT members chose to take the field test. Only 25% of the respondents identified themselves as ASHT members, indicating a high level of participation by nonmembers. About 60% of the 369 active and associate members of ASHT took the field test.

### Test Statistics

The reliability of each field test was estimated by use of the Kuder-Richardson formula 20 (KR20). The KR20 coefficient provides information about the consistency of examinee responses to each item within the test. The magnitude of this coefficient indicates the likelihood that an examinee's score would be similar if a different form of the field test were administered. The KR20 coefficients ranged from .61 to

.77, with an average of .69, an expected and acceptable level of reliability for a 75-item field test. KR20 reliability estimates tend to be higher with longer tests composed of similar items. The certification examination will contain scored 200 items—over twice as many items as the field-test forms. In addition, the consistency of candidate responses is likely to increase after specialized content review and study. Therefore, the reliability of the certification examination is quite likely to be in desirable range of .80 or higher.

Field-test scores were analyzed at the test and item level for the total group of examinees and for examinee subgroups. The summary test statistics for the total group for each field test form are shown in Table 2. The average raw scores for the total group ranged from 35.31 to 41.61, meaning that, on average, 35 to 41 items out of 75 were answered correctly. Item difficulty was determined by analyzing how often an item was answered correctly. A difficult item will be answered correctly by fewer people and will therefore have a lower item difficulty value. Average item difficulties ranged from .47 to .56 with an average of .54, indicating that each item was answered correctly by about half of the examinees. Table 3 shows the number of items in each of the five ranges of item difficulty. Fifty percent of the items had item-difficulty values greater than .40 and less than .79. Only 9% of the items had difficulty values less than .20. Figure 2 shows the distribution of item-difficulty values for all of the field test items. It shows that although most of the field-test items had difficulty val-

TABLE 2. Summary Test Statistics

FORM	TOTAL GROUPS					
	NUMBER TESTED	AVERAGE SCORE	STANDARD DEVIATION	AVERAGE ITEM DIFFICULTY	RELIA-BILITY	STANDARD ERROR OF MEASUREMENT
1891	74	39.70	6.03	.53	.63	3.67
1892	72	41.61	7.54	.56	.77	3.62
1901	95	41.52	7.52	.55	.74	3.70
1902	95	35.31	6.39	.47	.64	3.83
2901	145	40.33	6.93	.54	.70	3.80
2902	144	38.90	6.60	.52	.69	3.67
3901	108	39.92	7.58	.53	.76	3.71
3902	106	41.58	6.02	.55	.61	3.76

TABLE 3. Field Test Item Difficulties

FORM	ITEM DIFFICULTY				
	.00-.19	.20-.39	.40-.59	.60-.79	.80-1.00
2901	7	12	26	17	13
2902	8	17	17	20	13
1901	4	17	20	19	15
1902	7	22	27	11	8
1891	7	20	17	15	16
1892	9	15	11	23	17
3901	8	17	18	17	15
3902	4	14	26	24	17
Total Items:	54	134	162	136	114 (600)
% of Total:	9%	22%	27%	23%	19%

## Field Test Item Difficulty Distribution

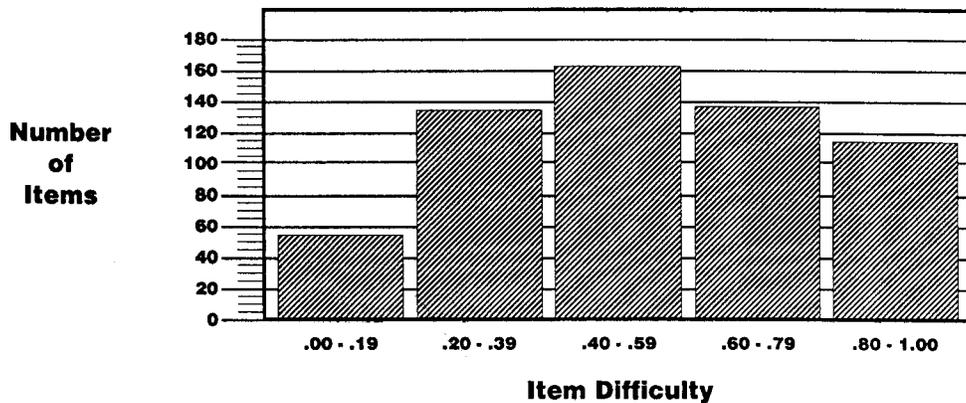


FIGURE 2. Distribution of the item difficulty values for the 600 field-test items.

TABLE 4. Total Field-tested Items Selected for the Certification Examination Pool

DOMAIN	FIELD TESTS 1, 2, 3, AND 4		
	NUMBER SELECTED	NUMBER TESTED	TOTAL NUMBER NEEDED
1.0	0	1	0
1.1	31	50	9
1.2	16	21	6
1.3	1	7	4*
1.4	11	19	8
2.1	3	8	4*
2.2	13	20	8
3.0	0	1	0
3.1	6	11	5
3.2	7	13	5
3.3	0	1	0
4.0	1	1	0
4.1	8	12	3
4.2	28	48	10
4.3	3	6	2
4.4	2	4	2
4.5	9	17	6
4.6	9	18	5
4.7	7	13	4
4.8	14	24	6
5.1	26	43	10
5.2	36	50	19
6.1	18	27	9
6.2	14	25	10
6.3	11	16	8
7.1	7	14	4
7.2	12	24	12
8.0	10	19	7
9.1	16	28	12
9.2	16	29	10
9.3	16	30	12
Totals	351	600	200

\* An insufficient number of satisfactory items are available for the 1991 certification examination for these domains.

ues above .39, there was a set of items with very low difficulty values that reduced the average difficulty value for the whole set of items. When the most difficult items are eliminated, the average item-dif-

ficulty value is .67, which is close to the target item difficulty of many examinations. It also shows a desirable distribution of items, in that most are in the moderate range of difficulty, with fewer at the extreme ends of the distribution.

Item difficulty is a concern for all therapists who plan to take the certification examination. A test that is perceived as being "too easy" would not represent the complexity of hand therapy practice. On the other hand, a test that is perceived as being "too difficult" might prevent qualified hand therapists from obtaining certification. A few points should be kept in mind when considering the item-difficulty values of the field test items. First of all, the item data reported are for the field tests and, as such, the summary statistics describe the characteristics of the whole pool of items, not the characteristics of the certification examination. Also, since the number of examinees taking each field test was quite small (74 to 145), these statistics are only rough estimates of the certification examination results. These data permit us, however, to eliminate items that appear to be too hard or too easy. Items with field-test item-difficulty values between .40 and .79 will be given priority in selection for the examination. The average item-difficulty level for the examination is, therefore, likely to be higher than the average item difficulty level for the field tests.

Second, examinees who took the field tests did so prior to studying review materials and prior to having a handbook available to them advising them about examination content and the references to use in the examination preparation.<sup>2</sup> Scores after content review and examination preparation based on the examination blueprint and recommended references are also likely to be higher than field-test scores.

### Item Selection

After field testing, statistics were calculated and reviewed to eliminate any items from the certification examination pool that appeared to be too easy or too difficult. A total of 351 items were selected and are

TABLE 5. Results of the Subgroup Analyses

SUBGROUP	AVERAGE NUMBER TESTED	AVERAGE RAW SCORE	PERCENT CORRECT
Occupational therapist	735	39.30	52%
Certified before 1981	448	40.75	54%
Certified 1981 or later	343	38.53	51%
Case load of 50% or higher for 6 years or more	452	41.42	55%
Case load of 50% or higher for less than 6 years	371	37.84	50%
ASHT: Nonmember	614	38.65	52%
Member	<u>208</u>	<u>43.23</u>	<u>58%</u>
Totals	839	39.86	53%

available for construction of the 200-item scored portion of the certification examination.

Of the 27 specified content domains, 25 had sufficient items selected to meet the test blueprint. Table 4 provides a summary of the items field-tested and selected by content domain. The Examination Committee will need to obtain only four additional items to meet the test blueprint for domains 1.3 (Scar Management for Hypertrophy) and 2.1 (Evaluation of Edema). Information on response patterns of poorly performing items in these content domains will provide guidelines for revising field-test items to make them appropriate for the certification examination.

The 249 field-test items that did not meet selection criteria for the certification examination pool may be reviewed by the Examination Committee, revised or rewritten as needed, and pretested on the 1991 or 1992 examinations.

### Subgroup Analyses

The results of the subgroup analyses of field-test data are shown in Table 5. Subgroups were based on the demographic information collected on occupation, length of certification, experience with upper extremity practice, and ASHT membership. Subgroup results for physical therapists are not reported, because too few physical therapists took the examination to provide reliable data. Examinees who indicated they were members of the ASHT had the highest average score (58%). Those who indicated they had case loads of 50% or higher for 6 years or more and those certified before 1981 had the next highest average scores (55% and 54%, respectively). The range of average scores among subgroups was relatively narrow. The higher average scores for subgroups with more experience in the field and more specialization in upper extremity practice, however, would tend to support the items as appropriate for advanced competency assessment.

### Respondent's Ratings of Perceived Item Difficulty

An optional survey was conducted at the end of each test administration to give examinees an op-

portunity to rate the perceived level of difficulty of the field-test items and comment on the field tests in general. A total of 809 examinees completed the survey. Of the respondents, less than 1% rated the items as very easy for therapists meeting certification examination eligibility requirements; 2% rated them as moderately easy; 30% rated them as appropriate; 48% rated them as moderately difficult; and 20% rated them as very difficult. Thus only 23% of the field test participants rated the difficulty level of the items as not appropriate for certification examination candidates. Many respondents commented that, although they rated items as moderately or very difficult, they felt that the difficulty level was appropriate due to the nature of the certification examination.

### SUMMARY AND CONCLUSIONS

Overall, the results of the field tests were quite good. The demographic information collected indicated that the target population for the certification examination was well-represented in the field-test sample of examinees. This indicates that the field-test statistics can be used as good indicators of certification examination statistics. The demographic data collected also helped the HTCC to better estimate the number of test sites needed, to help ensure that candidates choosing to take the certification examination have access to convenient sites.

The test and item statistics collected indicate that, as a whole, the items were appropriate for the target candidate population, i.e., those therapists with considerable experience and expertise in hand therapy practice. Field-test reliabilities indicated that the reliability of the certification examination is also likely to be in the appropriate range.

Sufficient items were selected based on the field-test performance to meet the certification examination blueprint, with the exception of two subdomains for which four additional items will be developed. A total of 151 good-performing field-test items will be available for future examinations after 200 are selected for the 1991 examination. In addition, the 249 poorly performing field-test items may be revised, using performance patterns as clues to effective re-

vision, and used as pretest items on the certification examination.

Analyses of field-test scores by subgroups confirmed the appropriateness of the items for advanced competency certification. In addition, examinee ratings of perceived item difficulty levels indicated that most of the examinees perceived the items to be at the appropriate level of difficulty.

Taken together, the field-test results indicate that the preparation undertaken by the HTCC in conjunction with ACT and other consultants for the 1991 certification examination has been well planned and

thorough. A solid foundation has been built for a reliable and valid examination program that is also perceived as valuable by candidates.

## REFERENCES

1. Chai SH, Dimick MP, Kasch MC: A role delineation study of hand therapy. *J Hand Ther* 1:7-16, 1987.
2. Available from the Hand Therapy Certification Commission, 1002 Vandora Springs Road, Suite 101, Garner, North Carolina 27529.

## RESEARCH FOR THE CLINICIAN

### Evaluating Published Research: Method — Population Sample

The introductory section of a research article narrows in focus from the literature review to identification of the problem and statement of purpose, and comes to completion with a hypothesis statement. Following the introduction, all methods and techniques used to conduct the study are described. While many currently published articles fail to do so, the method portion of a research article should provide enough detail to allow complete replication of the study.

A comprehensive description of the sample population studied is usually provided at the beginning of the methods section. Of uppermost importance to the reader is identification of the number of subjects studied. Far too many articles are published with insufficient numbers of subjects, allowing potentially skewed or erroneous assumptions and conclusions. Generally speaking, the greater the numbers of subjects included in the study the better. Unless an article is presented as a case study or a preliminary study, conclusions based on fewer than 20 subjects should be viewed with extreme caution. Granted, there are statistical instruments that are applicable to small subject groups, but the conclusions generated by such statistics often are not considered to be as strong or reliable as those from large population samples. Forty subjects are better than 20, 80 even better, and in most cases

160 would lend considerable credence to the conclusions drawn. Unfortunately, specific numbers for subject population cannot be dogmatically assigned and applied universally to all studies. There are times when, depending upon the entity studied, 1000 subjects would be required, or when less than 15 may be appropriate. Statistical measures exist that identify for researchers the number of subjects required for study, and if these instruments have been used, it should be so stated, allowing the reader to proceed with confidence.

While seemingly large numbers of subjects may look good initially, close inspection may reveal major problems. If the number of subjects is broken down into subgroups according to diagnosis, handedness, sex, age, or other appropriate variable, the number of subjects within each subgroup must be scrutinized carefully. For example, if the overall number of subjects studied is 30, but they are divided into three subgroups according to diagnosis (10 subjects each), and then further grouped according to sex (5 subjects per group), conclusions drawn from the study are, in reality, based on appallingly few subjects, 5 to be exact! Unfortunately, this fallacy is especially apparent in some of our hand-related normative studies that are considered "classics" and are cited repeatedly in the literature. Three hundred normal

subjects divided into 12 subgroups according to sex and age results in far too few subjects per group from which to draw reasonable and responsible conclusions. Three hundred subjects *per category* would be more appropriate for normative statements!

A converse problem, failure to identify and control pertinent population variables, may enhance subject numbers, but the conclusions drawn can be equally as misleading and inappropriate. Reports that fail to consider differences in diagnoses, age, sex, hand dominance, treatment, occupation, pain, etc., although exhibiting greater subject numbers, are identified and rendered valueless by astute and knowledgeable readers.

As noted in previous columns, "the buck stops" with each of us as individual therapists. Ultimately, we alone make the decisions of what is appropriate and not appropriate for our patients. Our most important source of current information comes from published articles. It is therefore critical that we be able to identify quality research. Understanding the importance of an adequately sized and appropriately controlled subject population is but one of many of the criteria by which we may judge the value of published research.

Elaine Ewing Fess, MS, OTR, RA