

.....

The **BEHAVIORAL** **MEASUREMENT** *Letter*

**Behavioral
Measurement
Database
Services**

.....

Enriching the health and behavioral sciences by broadening instrument access

Vol. 9, No. 1
Summer 2006

Introduction to the Spring 2006 Issue of *The Behavioral Measurement Letter*

There is a growing demand in the market place for expertise in psychometrics, at a time when fewer and fewer people are receiving training in measurement. According to a recent article on the front page of the *New York Times* (Herszenhorn, D.M. "As Test-Taking Grows, Test-Makers Grow Rarer," May 9, 2006, A1, A19), testing company executives are urging federal action to increase opportunities for training in measurement, including government-paid fellowships to increase the numbers of students entering graduate programs to become experts in psychometrics. As demand for measurement specialists continues to outstrip supply, it will become increasingly critical in the years ahead to broaden and strengthen educational training in measurement in the health and psychosocial sciences.

In this spirit, the Spring 2006 issue of *The Behavioral Measurement Letter* is devoted to a topic that is currently missing from the background training of most researchers in the health and psychosocial sciences—namely, how to select appropriate measurement instruments for research. Students receive extensive undergraduate and graduate training in how to formulate hypotheses, design research studies, and collect, analyze, and interpret research data in their various empirical disciplines. They learn how to maximize the validity of research conclusions using fully factorial experimental designs in randomized laboratory experiments, cross-sectional and longitudinal designs in survey research, and quasi-experimental designs in applied settings. They learn how to use computer software to conduct a wide range of sophisticated

statistical analyses, and how to write and publish research reports communicating their findings. But most students are never taught how to go about finding an appropriate measurement tool for use in their research—they are never formally trained in the process of instrument selection, or "instrumentation."

It is particularly ironic that students typically receive no training in how to select suitable measurement tools for research, given that the quality and precision of measurement determines the quality and precision of research findings. Imagine that space scientists were trained to operate the way undergraduate and graduate students in the health and psychosocial sciences are trained to operate with respect to instrument selection. On the one hand, NASA would spend billions of dollars designing and constructing a sophisticated space probe capable of flying millions of miles through space to another planet, landing safely on the planet's surface, and relaying information back to Earth electronically; and scientists would devote years of effort in carefully scrutinizing and interpreting the data the probe relayed back to Earth. On the other hand, scientists would spend little time or money in deciding what measurement tools to include on-board the space probe in the first place. Instead, they would choose instruments haphazardly from among whatever options they could find, never being sure whether they were missing an important tool in the process. Space scientists might also grab the first available measure of a particular variable of interest to put in the space probe, without considering whether it actually reflects the construct of

interest. Surely there's a better way.

Accordingly, the Spring 2006 issue of *The Behavioral Measurement Letter* addresses this gap in health and psychosocial research training by presenting three articles, each aimed at helping researchers optimize the process of instrument selection. This issue begins with a reprint of a 1999 article from *Eye on Psi Chi* by Daniel Moore, Fred Bryant, and Evelyn Perloff, "Measurement Instruments at Your Finger Tips." Moore, Bryant, and Perloff remind readers that measurement surrounds us in everything we do, yet we know so very little about how to select the most appropriate measures for a given purpose. This article also introduces the HaPI database as a resource for optimizing researchers' use of the measurement toolbox.

Also reprinted in this issue is a 1998 article from *Teaching of Psychology* by Jennifer Brockway and Fred Bryant, "You Can't Judge a Measure by Its Label: Teaching the Process of Instrumentation," describing an exercise in instrument selection that increases awareness of the issues involved in measuring conceptual variables. This article is intended as a resource for classroom instructors to use in courses on research methods in the health and psychosocial sciences, including counseling, education, medicine, nursing, psychology, social work, and sociology. Brockway and Bryant provide empirical evidence that the exercise is effective in raising student's consciousness about how to select suitable instruments for use in their research.

Finally, in "Selecting Instruments for Behavioral Research: Advice for the Intermediate User," Thomas Hogan of the University of Scranton's Department of Psychology provides advice for intermediate-level researchers (which includes most instrument users) about the informational resources available and how to use them to identify suitable measurement instruments, retrieve copies and scoring protocols for instruments, and find the right measure for the job at hand. Dr. Hogan highlights available electronic databases and printed compendia that provide researchers with the tools necessary to locate, evaluate, and select appropriate measurement tools in the health and psychosocial sciences.

All three articles in this issue emphasize the need

for a better understanding of the process through which researchers select instruments in their research, so as to enhance the quality of research conclusions. This issue of *The Behavioral Measurement Letter* is intended to serve as a resource for instructors in research methods courses within the various disciplines that constitute the health and psychosocial sciences. It is hoped that instructors will use this issue of the newsletter as a tool for raising students' awareness of the issue involved in selecting appropriate measurement instruments for research.

Behavioral Measurement Database Services (BMDS) is grateful to Lawrence Erlbaum Associates, Inc. (Mahwah, NJ), for permission to reprint Brockway, J. H., & Bryant, F. B. (1998). You can't judge a measure by its label: Teaching students how to locate, evaluate, and select appropriate instruments. *Teaching of Psychology*, 25, 121-123.

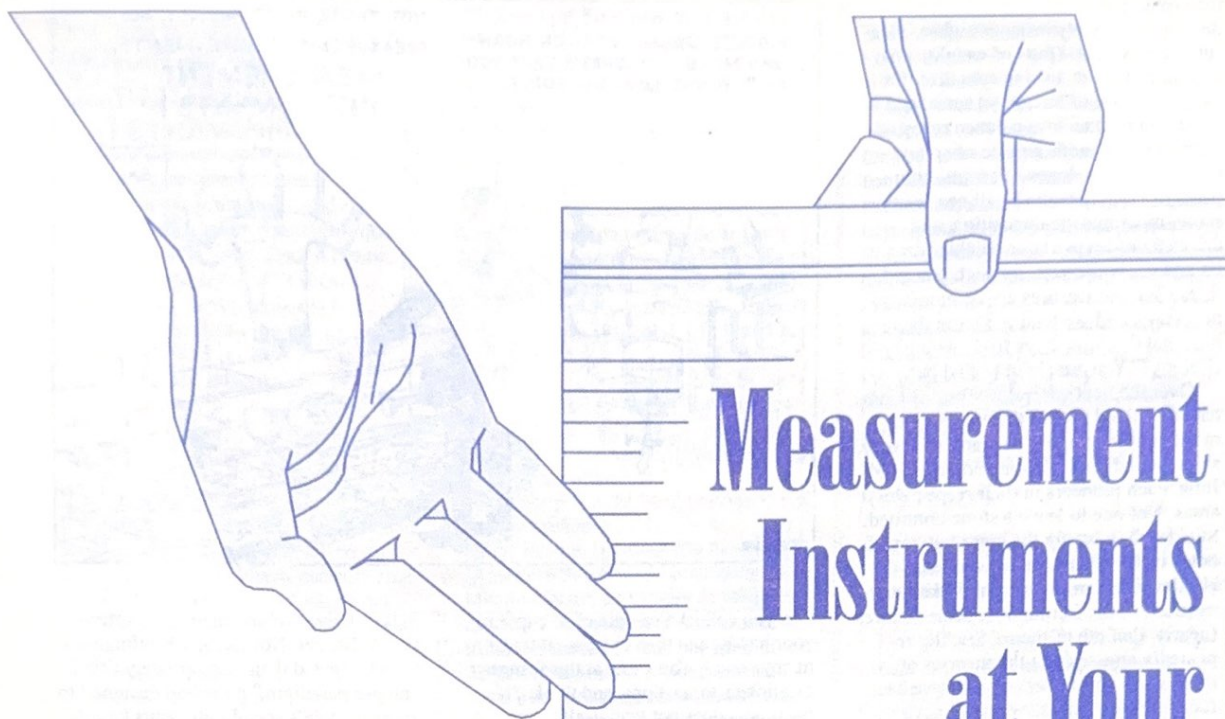
BMDS also thanks the Psi Chi National Office (Chattanooga, TN) for permission to reprint Moore, D., Bryant, F. B., & Perloff, E. (1999). Measurement instruments at your fingertips. *Eye on Psi Chi*, 3, 17-19.

As always, we invite written responses from our readership. Please address comments, suggestions, letters, or ideas for topics to be covered in future issues of the journal to: The Editor, *The Behavioral Measurement Letter*, Behavioral Measurement Database Services, P.O. Box 110287, Pittsburgh, PA, 15232-0787. Email: bmdshapi@aol.com.

We also accept short manuscripts to be considered for publication in *The Behavioral Measurement Letter*. Submit, at any time, a brief article, opinion piece, or book review on a topic related to behavioral measurement, to The Editor at the above address. Each submission will be given careful consideration for possible publication in a forthcoming issue of *The Behavioral Measurement Letter*.

HaPI reading...

The Editor



Measurement Instruments at Your Fingertips

Daniel Moore, MA
Fred B. Bryant, PhD
Evelyn Perloff, PhD

IMAGINE YOU ARE A COLLEGE senior. For the past three years, you've seen and met a variety of academic challenges. You've endured grueling daily lectures, tutored struggling students, participated in numerous class projects, and written endless essays and papers. In every case, you completed your assignments promptly and well.

However, this time is different. Sitting in class early in the term listening, you look dumbstruck as Professor Stickler—a strict disciplinarian with a demanding teaching style—describes an assignment that leaves you wondering why you majored in psychology.

IT ALL STARTS ROUTINELY ENOUGH. Professor Stickler announces that the semester would be dedicated, as is this paper, to exploring a concept that most people take for granted: measurement. Professor Stickler says: "We usually think of measurement as a person's or an object's dimensions, quantity, or capacity, which, of course, it is. But it is also much more." The professor pauses while reaching for the *Oxford English Dictionary*, then continues, "This dictionary, the unassailable authority on our language, allots over four pages of text to measurement and its many manifestations (Perloff, 1994). Why?" Professor

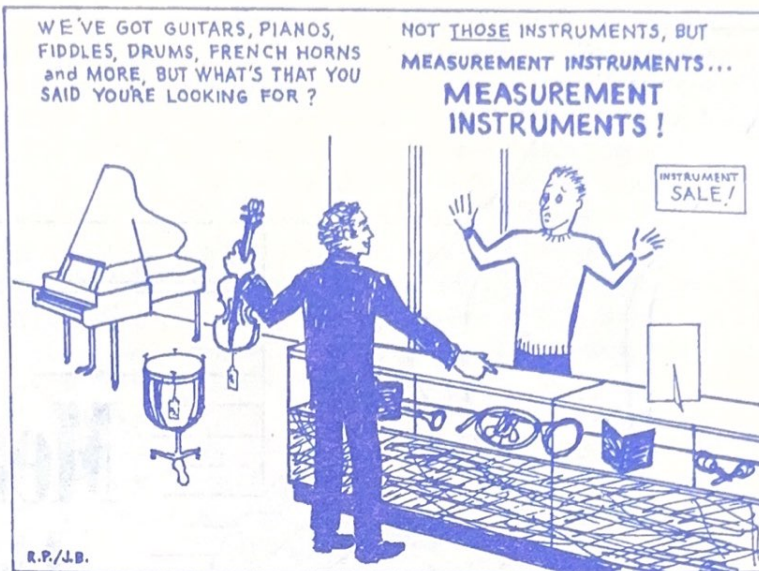
Stickler pauses again. "The answer should be obvious. Just look around and you will see some sort of measurement in nearly every area of life."

During the rest of the hour, Stickler cites dozens of ways that measures lend order and understanding to the world. But, in keeping with Stickler's style, the first examples are from space. You learn that a light-year is equivalent to 5.88 trillion miles, a nearly unimaginable distance, but it pales in significance as Stickler recounts how astronomers have used light-years to measure the seemingly infinite vastness of the entire universe—which is still expanding. Stickler

then returns to Earth and reminds you of the many everyday instances where measures are evident. Units of weight, area, and time allow us to conceptualize, for example, a ton of bricks, 40 acres, and a millennium. One buys a dozen roses, wears a size 8, adds a pinch of salt, checks the Dow Jones Averages, compares prices, scans the headlines, keeps score, measures the opposition, and drives 55 miles per hour. Stickler drones on and on, while you, somewhat bored, think, "True, measures are common and necessary, and we tend to ignore them, but what does that have to do with psychology?" You are about to find out.

Over the next several weeks, the class takes a guided tour of the evolution of modern psychology. You learn, of course, about Wundt and Watson and Freud and Jung, each pioneers in their respective areas. Not one to leave a stone unturned, Stickler then covers the contributions of noteworthy others: Pavlov, James, Harlow, Erikson, Rogers, and Skinner, to mention a few. Throughout these lectures (nearly a month of them), Stickler repeatedly stresses that the theories of these early pathfinders would have been meaningless—that they would have amounted to little more than unsubstantiated speculation—unless their underlying concepts could be measured. To underscore this point, Stickler declares, emphatically, "Measurement is the cornerstone of science. In fact, just as in the physical sciences, advances in psychology are proportional to advances in measurement. As Robert Pool (1988) stated in the case of the physical sciences, "These advances are vital, because science's understanding of the physical world is limited by the accuracy with which science can measure that world."

At the beginning of the next class, Professor Stickler asks how many students hope to live the good life. Every hand in the room shoots up. Stickler then asks for a definition of a "good life"; not a hand stirs. Stickler continues, "Is it happiness, and if so, how is it measured?"



R.P./J.B.

Can you collect a sample of it, cup it in your hands, and then examine it under a microscope?" You smile at the absurdity of Stickler's questions, and think, "If happiness does not lend itself to precise quantification, then how does one measure it, and the kaleidoscope of other forces, that shape attitudes and behavior?" Stickler echoes your thoughts, and explains that generations of researchers have responded to this problem by developing a vast array of measures or instruments that indirectly assess people's psychological states.

Most numerous are the traditional "paper-and-pencil" tests, questionnaires, checklists, inventories, rating scales, and interview schedules, many of which now have computerized versions or scoring systems. Stickler then reviews instruments that employ alternative methods. For example, Csikszentmihalyi and Larson (1987) devised the Experience Sampling Measurement, a measurement technique that uses an electronic pager to signal participants to record their feel-

ings as they go about their daily activities. Vallacher, Nowak, and Kaufman (1994) created their appropriately titled "mouse paradigm," a method designed to assess people's social judgments based solely on where they position a cursor on a computer screen.

Stickler continues this emphasis up to and through the midterm examination. Immersed in measures and measurement issues, you struggle to grasp concepts such as construct validity, internal consistency, forced-choice, ratio scale, multitrait-multimethod matrix, measurement modeling, and Cronbach's alpha. You take copious notes, do the assigned readings, and try to appear knowledgeable in class discussions. This strategy must have worked, for your midterm grade is a B+. At this point, you learn that in place of a final exam, Stickler has devised an assignment in which you must explore a particular construct that he will speak to shortly.

Although you do find this information of interest, you uncomfortably await Dr.

Daniel Moore earned his MA from the University of Pittsburgh. As an undergraduate, Moore participated in the National Honors Society Honor Lecture Series and the 1987 and 1988 Northeast Region of the National Collegiate Honors Councils at Providence College and Syracuse University. As a graduate student, he completed a research project for the Center for Social History exploring novel ways historians measure the past. His awards and distinctions include an undergraduate Intercession Travel Award to China and three graduate fellowship awards. Moore is chief analyst at Behavioral Measurement Database Services.

Fred Bryant is professor of psychology at Loyola University, Chicago. He has 80 publications in social psychology, personality psychology, measurement, and behavioral medicine. In addition, he has coedited five books. Dr. Bryant has extensive consulting experience in many applied settings, including marketing firms, medical schools, and public school systems, and has been a methodological expert for the U.S. Government Accounting Office and an expert witness in several federal court cases involving social science research evidence. His current research interests include happiness, psychological well-being, Type A behavior, the measurement of cognition and emotion, and structural equation modeling.

Evelyn Perloff is founder and director of Behavioral Measurement Database Services, the producer of Health and Psychosocial Instruments (HaPI). In addition to working in the public sector and for the federal government, Dr. Perloff was a professor at the University of Pittsburgh, Northwestern University, and Purdue University. At Purdue, she founded and served as the first faculty advisor to an active and productive Psi Chi chapter. Her interests and publications have been directed primarily to measurement issues in health psychology.

Stickler's assignment which then follows all too quickly, but is, you agree, appropriate to your immediate feelings. The assignment concerns the construct of anxiety. Each student is required to select two instruments designed to assess anxiety and to compare and contrast the measures for a class presentation.

Topics to be addressed could include: strengths, weaknesses, theoretical orientation, type of measure (e.g., questionnaire, interview schedule, checklist, vignettes) intended population (e.g., children, adolescents, the elderly, college presidents), number of questions. Professor Stickler then posts a sheet for students to list their selected titles as soon as they have identified them, because no instrument may appear twice.


Now, you do become anxious. With 20 students in your class, you realize that 40 different measures need to be located, and you wonder whether there are as many as 40 instruments to measure anxiety. Surely, you think, Professor Stickler would not make such a demand, if this were not the case.

The thought urges you to get started as soon as the period ends, but where to start? As you leave class, a friend recalls seeing an instrument store around the corner. Sounds good, you think, and decide to make this your first stop, although you wonder can it be so easy.

Of course, you should have known better, and you head for a more likely source: the library. At the Reference Desk, you request information on how to find materials about measurement instru-

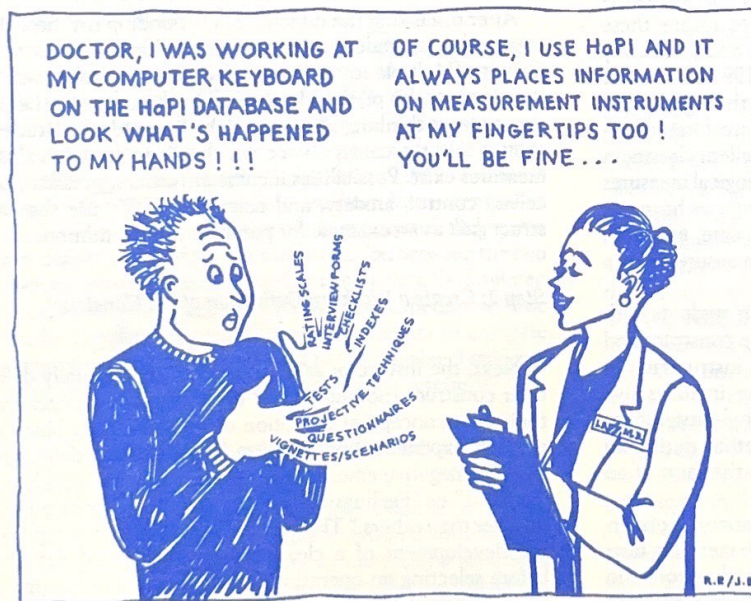
ments for the behavioral and social sciences. The librarian quickly directs you to a section containing several shelves of relevant books. Feeling better, you begin to review their contents. After several hours of browsing and finding interesting but not specifically useful information, you return to the Reference Desk and ask about other sources you could consult. The librarian then refers you to The Mental Measurements Yearbooks (Buros, 1938-1978; Conoley & Impara, 1989, 1995; Cramer & Conoley, 1992; Impara & Blake, 1998; Mitchell, 1989) and Test Critiques (Keyser & Sweetland, 1984-1994), both containing reviews of commercially produced tests. But the prospect of wading through 13- and 11-volume collections does little to lessen your growing anxiety.

At this point, the librarian asks if you have consulted the Ovid Technologies online databases, especially PsychINFO (American Psychological Association), Medline (National Library of Medicine), and HaPI (Behavioral Measurement Database Services). You are familiar with the first two, which are abstract databases of information on the behavioral and medical literature, respectively. The last database, however, is less familiar. You inquire further and learn that HaPI is the acronym for Health and Psychosocial Instruments, a computerized database that consists of more than 60,000 records describing tests, questionnaires, checklists, interview schedules, and other types of measures used in the health and psychosocial sciences.

At first, you refuse to believe that a database dedicated solely to instruments exists, but after typing the word "anxiety" and seeing the results, you have a change of heart. HaPI returns more than enough anxiety measures! That's not all, for you notice that many of the records contain additional information about the instrument, such as the number of questions, a statement of purpose, response formats, and relevant references. You quickly realize that HaPI provides the necessary information to complete your assignment. Moreover, the anxiety that has accompanied you since you started this project now gives way to a welcome sense of well-being. But what's happened to your fingers! 

References

- Buros, O. K. (Ed.). (1938-1978). *The first-eighth mental measurements yearbooks*. Lincoln: University of Nebraska, Buros Institute of Mental Measurements.
- Conoley, J. C., & Impara, J. C. (Eds.). (1989). *The tenth mental measurements yearbook*. Lincoln: University of Nebraska, Buros Institute of Mental Measurements.
- Conoley, J. C., & Impara, J. C. (Eds.). (1995). *The twelfth mental measurements yearbook*. Lincoln: University of Nebraska, Buros Institute of Mental Measurements.
- Cramer, J. J., & Conoley, J. C. (Eds.). (1992). *The eleventh mental measurements yearbook*. Lincoln: University of Nebraska, Buros Institute of Mental Measurements.
- Csikszentmihalyi, M., & Larson, R. (1987). The Experience Sampling Method. *Journal of Nervous and Mental Disease*, 175, 526-536.
- Impara, J. C., & Blake, B. S. (Eds.). (1998). *The thirteenth mental measurements yearbook*. Lincoln: University of Nebraska, Buros Institute of Mental Measurements.
- Keyser, D. J., & Sweetland, R. C. (Eds.). (1984-1994). *Test critiques* (Vols. I-X). Austin, TX: pro-ed.
- Mitchell, J. V., Jr. (Ed.). (1989). *The ninth mental measurements yearbook*. Lincoln: University of Nebraska, Buros Institute of Mental Measurements.
- Perloff, R. (1994). For good measure. *The Behavioral Measurement Letter*, 1(1), 4-5.
- Pool, R. (1988). Advances in measurement science. *Science*, 240, 604-605.
- Vallacher, R. R., Nowak, A., & Kaufman, J. (1994). Intrinsic dynamics of social judgment. *Journal of Personality and Social Psychology*, 67, 20-34.



Eye on Psi Chi Winter 1999

In physics the truth is rarely perfectly clear, and that is certainly universally the case in human affairs. Hence, what is not surrounded by uncertainty cannot be the truth.

Richard P. Feynman

19

You Can't Judge a Measure by Its Label: Teaching the Process of Instrumentation

Jennifer Howard Brockway
United States Air Force Academy

Fred B. Bryant
Loyola University Chicago

In this article, we describe an exercise in instrument selection applicable to both undergraduate and graduate courses. The exercise consists of 5 progressive steps that involve choosing and defining a theoretical construct (Steps 1 and 2), using computer-based technology to obtain 2 distinct instruments that measure this construct (Steps 3 and 4), and comparing and contrasting these 2 instruments along multiple dimensions (Step 5). This activity generates awareness of the issues involved in measuring latent constructs and teaches at least 3 lessons: (a) The 1st step to accurate instrumentation is precisely conceptualizing the construct of interest; (b) there is more than 1 way to measure latent constructs, and these multiple approaches should be compared; and (c) it is crucial to measure a construct in a way that best matches its underlying definition.

A difficult concept for many students to grasp is the notion that there is no one universally appropriate measure for any given psychological construct. Typically, several good (reliable and valid) instruments already exist for most constructs with which psychology students are familiar. Thus, an important skill to teach students is the ability to locate these measures and to choose the most appropriate one (Brockway & Bryant, 1997). According to Brewer et al. (1993), it is better to teach such abstract methodological issues through hands-on exercises as opposed to traditional lecture format. Although instructors have suggested several excellent classroom exercises with respect to creating new psychological measures (e.g., Benjamin, 1983; Davidson, 1987), instructors have not offered exercises to teach students how to locate, evaluate, compare, and select the most appropriate measure from a database of existing measures.

The following exercise helps to familiarize students with the variety of measures that exist for any one construct and enables students to compare these different instruments to make better informed choices. The exercise includes five progressive steps: Steps 1 and 2 are part of a single assignment outside class, Steps 3 and 4 take place together during an out-of-class library session, and Step 5 takes the form of an in-class discussion or presentation.

This exercise is appropriate for use in a variety of undergraduate psychology classes, including research methods, tests and measurements, and various psychology laboratories in which students collect and analyze data. Graduate students may also benefit from this exercise by including a more rigor-

ous analysis, critique, and comparison of existing measures. At Loyola University Chicago we have incorporated this activity into an undergraduate social psychology laboratory course and a graduate research methods course. The following example comes from the undergraduate course.

Procedure

This activity is best preceded by a review of psychological constructs and the inherent difficulty of measuring phenomena that are intangible, dynamic, and subject to multiple definitions. For this exercise to be effective, students must understand the difference between conceptual and operational definitions.

Step 1: Choose a Psychological Construct

After discussing the difficulties surrounding psychological measurement, students choose a psychological construct of interest. (Multiple students can select the same construct.) A list of constructs previously covered in the class may help to get students thinking along the right lines. Ideally, students should select a construct for which a number of available measures exist. Possibilities include altruism, aggression, perceived control, anxiety, and compliance. We use the construct *guilt* as an example for purposes of presentation.

Step 2: Create a Working Definition of the Construct

Next, the instructor asks students to conceptually define their construct thoroughly and carefully. Students generate their own conceptual definition of the construct, based on personal experience and intuition. Thus, one may define guilt as, "The negative emotion experienced after knowingly misbehaving" or "Feelings of distress when one's social position is better than others." The purpose of this step is to emphasize the development of a clear, precise conceptual definition before selecting an operational definition (i.e., instrument). Thus, if guilt is one's construct, it is important to distinguish it from shame, embarrassment, and regret, for example.

Step 3: Use an Instrument File to Generate List of Existing Measures

After ensuring that students have precisely defined their construct, instructors ask students to utilize their library's measurement database to generate a list of existing measures of their construct. A powerful new measurement database called the Health and Psychosocial Instrument (HaPI) File (1995) is particularly useful in this regard. HaPI provides information about thousands of behavioral and social measures through abstracted descriptions summarizing instrument characteristics (e.g., intended audience, validity and reliability information, means of obtaining copies). HaPI is available in hundreds of college libraries both in the United States and internationally.

The HaPI File has distinct advantages over traditional catalogs of instruments. First, HaPI provides a larger number of measures (more than 40,000) than other sources. HaPI also provides a more efficient means of managing the volume of measurement information that exists. We believe the HaPI File provides a more thorough and cost-effective measurement tool than other measurement volumes.

To use HaPI, students simply type in the name of their construct and generate a list of corresponding references. For instance, for the guilt construct, HaPI generated a list of 73 instruments that the originators of these measures described as assessing some form of guilt.

Step 4: Choose Two Measures With Distinct Conceptual Definitions

From the list of measures, students choose two references for distinct instruments and obtain these articles from their library. One article should define the construct in a way that resembles the students' conceptual definition. The other article should define the construct in a way that is different from the students' definition.

From the list of guilt measures, for example, we chose two articles with different theoretical orientations (Kugler & Jones, 1992; Montada & Schneider, 1989). Whereas the first article approached guilt from a macrolevel (sociological, cultural) perspective, the second article assessed guilt from a more microlevel (psychological) perspective. We chose these measures not because they are the best instruments, but because of the contrast between their conceptual and operational definitions. The greater the distinction between the two chosen measures, the easier it will be to complete the final step of the exercise. Although our students have experienced little difficulty in selecting two different instruments to compare and contrast, instructors may want to be available if students need assistance with this critical step of the exercise.

Step 5: Compare and Contrast Alternative Instruments

After ensuring that students have found distinct measures, ask students to make an in-class presentation comparing and contrasting the measures on various dimensions generated either by the instructor or by the students and the instructor. Besides each measure's overall strengths and weaknesses, other dimensions could include:

1. Theoretical orientation (e.g., social justice vs. psychological conceptualizations of guilt);
2. Duration and frequency of construct manifestation (e.g., state vs. trait guilt);
3. General format of the instrument (e.g., reactive vs. unobtrusive measures of guilt, vignettes vs. self-report questions, closed- vs. open-ended items);
4. Intended audience (e.g., children, English speaking adults, etc.);
5. Number of items and scaling issues (e.g., single item vs. composite index, Likert vs. semantic differential response format).

For example, Montada and Schneider (1989) defined *existential guilt* as a prosocial emotion felt when one perceives oneself as better off than others suffering hardships. To measure guilt, Montada and Schneider embedded three guilt items within a larger questionnaire designed to tap other "prosocial emotions" such as sympathy and moral outrage. The general form of the measure is a written scenario describing the misery of a group of disadvantaged people (e.g., the unemployed). Respondents use a 6-point scale to rate the degree to which three statements reflecting guilt express their thoughts and feelings. The instrument's intended audience seems to be at least the age of young adults because it assumes respondents possess some moral awareness concerning the status of disadvantaged populations. Although the instrument appears in English, Montada and Schneider developed it using a German sample, so it is unclear whether it is applicable to other populations.

Taking a very different approach, Kugler and Jones's (1992) measure is a guilt inventory. Ninety-eight items, presented on 5-point scales, tap three specific content domains (trait guilt, state guilt, and moral standards). Both college students and adult nonstudents have completed this inventory.

Evidence of Pedagogical Effectiveness

To assess the effectiveness of this exercise, 10 undergraduates enrolled in a social psychology laboratory course answered open-ended questions addressing several goals of the exercise both before and after completing the exercise. Results revealed that before the exercise, only 1 student (10%) knew that the first step in measuring a construct is to carefully create a conceptual definition, whereas all 10 students (100%) gave this correct answer after the exercise (Fisher's exact $p = .00006$). Also, before the exercise, only 1 student (10%) stated that when confronted with three equally valid and reliable measures, one should choose the measure that most closely matches the conceptual definition, unlike 7 of 10 students (70%) at the posttest (Fisher's exact $p = .0099$).

As a control condition, a comparable group of 6 laboratory students completed pretest and posttest measures, but did not participate in the measurement exercise. Results revealed no significant changes in knowledge from pretest to posttest regarding the crucial first step in selecting measures (Fisher's exact $p = .50$) and how to choose from among three psychometrically equivalent measures (Fisher's exact $p = .23$). Although the size and representativeness of these samples are

far from ideal, these data nevertheless support the effectiveness of the exercise.

Additional Suggestions

Instructors can simplify or expand this exercise to fit a particular time slot, lesson plan, or student population. For example, instructors can eliminate the library portion of the exercise and simply supply students with the results of the HaPI search for existing measures for a particular construct and with a copy of two preselected articles and instruments. Instructors may then ask students to compare and contrast the two measures with respect to the dimensions (or a subset of those dimensions) discussed previously.

Conversely, instructors can expand this activity to include issues more appropriate for advanced psychology students. For example, students can compare and contrast instruments with respect to the various validities (i.e., face, construct, criterion, content) and reliabilities (i.e., test-retest, parallel forms, interrater, internal consistency). Also, advanced students can locate instruments tapping similar (yet conceptually separate) constructs and highlight the subtle distinctions between the constructs (e.g., guilt vs. shame vs. embarrassment). This approach offers instructors a concrete means of teaching students about the multitrait-multimethod matrix (Campbell & Fiske, 1959) and how to implement it. Finally, and perhaps the most challenging task of all, instructors can ask students to uncover the "missing measure" after reviewing all existing instruments. By highlighting both the gaps and overlaps in current measurement options, future researchers can begin to distinguish between areas that need new instruments and areas in which new measures are unnecessary.

Conclusions

This measurement exercise teaches students several important lessons. First, it stresses the importance of having a clear conceptual definition before choosing an instrument to measure a psychological construct. Second, it teaches students that there almost always is more than one way to measure a construct and that they should compare these multiple approaches. Third, it teaches students the importance of measuring a construct in a way that best matches one's underlying

definition. Finally, it introduces students to new information technologies that they can use to locate, evaluate, compare, and select the best measures. Indeed, maximizing the match between conceptual and operational definitions is the essence of construct validity (Cook & Campbell, 1979). Thus, it is important to teach students that you can't judge a book (i.e., a measure) by its cover (i.e., its label).

References

- Benjamin, L. T., Jr. (1983). A class exercise in personality and psychological assessment. *Teaching of Psychology*, 10, 94-95.
- Brewer, C. L., Hopkins, J. R., Kimble, G. A., Matlin, M. W., McCann, L. I., McNeil, O. V., Nodine, B. F., Quinn, V. N., & Sandra. (1993). Curriculum. In T. V. McGovern (Ed.), *Handbook for enhancing undergraduate education in psychology* (pp. 161-182). Washington, DC: American Psychological Association.
- Brockway, J. H., & Bryant, F. B. (1997). *Spotlighting the second fiddle: The importance of teaching measurement*. Manuscript submitted for publication.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81-105.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design & analysis issues for field settings*. Chicago: Rand McNally.
- Davidson, W. B. (1987). Undergraduate lab project in personality assessment: Measurement of anal character. *Teaching of Psychology*, 14, 101-103.
- Health and Psychosocial Instrument (HaPI) File (Version 4.1) [CD-ROM]. (1995). Pittsburgh, PA: Behavioral Measurement Database Services [Producer and Distributor].
- Kugler, K., & Jones, W. H. (1992). On conceptualizing and assessing guilt. *Journal of Personality and Social Psychology*, 62, 318-327.
- Montada, L., & Schneider, A. (1989). Justice and emotional reactions to the disadvantaged. *Social Justice Research*, 3, 313-344.

Notes

1. An earlier version of this article was presented at the American Psychological Society Institute on the Teaching of Psychology, San Francisco, June 1996.
2. Correspondence concerning this article should be sent to Jennifer Howard Brockway, HQ USAFA DFBL, 2354 Fairchild Drive, Suite 6L47, USAFA, CO 80840-6228; e-mail: brockwayjh.dfl@usafa.af.mil.

Staff

Director Evelyn Perloff, PhD
Associate Director..... Linda Perloff, PhD
Measurement Analyst..... Judith Myers, PhD
Measurement Analyst..... Adam DeHock, MA
Measurement Analyst..... Jennifer Diamond, BS
Indexer Brian Hightower, MA
Informationalist..... Barbara Brooks, MLS
Newsletter Editor Fred Bryant, PhD

Computer Consultant Alfred Cecchetti, MLS
Measurement Consultant Fred Bryant, PhD
Information Specialist Consultant..... Ellen Detlefson, DLS
Assistant Analyst..... Olga Goldenberg
Database Services Manager Diane Cadwell
Journal Searcher..... Bryan Fuller, BA
Clerical Assistant Stephanie Jones
Record Checker..... Betty Hubbard

Selecting Instruments for Behavioral Research: Advice for the Intermediate User

Thomas P. Hogan
University of Scranton

General Overview

Selecting appropriate measurement instruments is a crucial part of any research study. For most psychological constructs, instruments already exist. Fortunately, there are a variety of sources of information to help locate suitable instruments. Although novices are best served by simply adopting whatever is most frequently used (or whatever they are told to use), persons at an intermediate level of training and experience should become proficient in using standard sources of information to find and evaluate instruments. I outline these sources, along with their strengths and weaknesses, in four major categories: electronic listings, hard copy listings, test reviews, and test publishers' catalogs and websites.

Users of Instruments

For didactic purposes, it is convenient to divide people who need to select an instrument for behavioral research into three groups. I will call the three groups of users novices, intermediates, and experts, much like a classification of tennis players or do-it-yourself home repairers. Of course, there is an underlying continuum of knowledge and skill for all three groups, but the discrete groupings will be helpful for formulating advice.

Novices in the world of instrument selection include undergraduate students and others with little experience or training in behavioral research. The entry-level users might have had one course in statistics and possibly a course in research methods. In fact, novice users might need to select a behavioral instrument for a first project as part of a course in research methods or psychological testing.

The intermediate group of instrument users includes individuals who have developed a reasonable base of training and experience in behavioral research. These mid-level users have completed several courses in statistics, research methods, and measurement. Intermediate users have likely designed and

carried out one or two simple research projects, but always under the supervision of a more experienced researcher. Many graduate students would fall into this middle category of instrument users.

The expert group, of course, includes experienced researchers who regularly publish the results of their research. These high-level users often supervise and teach novice and intermediate users. Expert users include professors and full-time researchers at college, universities, research institutes, and major test publishers.

Getting the Target Construct Right

In fundamental ways, all three groups have the same needs when it comes to selecting measurement instruments for research projects. As nicely described by Brockway and Bryant (1998) [reprinted in this issue of the newsletter], research studies typically revolve around theoretical *constructs*. These abstract constructs receive their concrete *operational definitions* in the form of some *measurement instruments*. A crucial question is how well a particular measurement instrument fits the construct(s) of interest.

Textbooks and journal articles tend to use cryptic labels (often one or two words) for constructs, for example, depression, self-esteem, nonverbal intelligence, or home environment. In many ways, these labels facilitate communication. Indeed, later sections of this article will use these short-hand labels for purposes of keyword searching. However, an overreliance on such labels belies the underlying richness and complexity of the construct the user wants to measure.

The *Standards for Educational and Psychological Testing* (Standards; American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999) provides a useful conceptual framework and terminology for this issue of the fit between construct and measure. The *Standards* refers to *construct underrepresentation* and *construct irrelevant variance*.

Selecting Instruments (continued)

Construct underrepresentation occurs when the measurement instrument does not fully capture the construct of interest. Construct irrelevant variance refers to measuring, in part, constructs in addition to what the user has in mind. The *Standards* note "Nearly all tests leave out elements that some potential users believe should be measured [construct underrepresentation] and include some elements that some potential users consider inappropriate [construct irrelevant variance]" (p. 10).

Before searching for appropriate measurement instruments, it is essential to spend time clearly defining the construct of interest. In other words, it is crucial that users get the target right. The following simple strategy may help this process. Start with a short-hand label for the construct. Then develop an elaborated definition or description of the construct. This might be one or two paragraphs. The expanded version might include synonyms, definitions from standard sources such as field-specific dictionaries, and textbook elaborations. Finally, try to identify what is *not* included in the construct. For example, you may want to ensure that the construct of quantitative ability is not contaminated with a heavy computational load or reading ability; or that anxiety is distinguished from fear. The important point here is that this process of clarifying the target construct should occur *before* you start searching for appropriate measurement instruments.

Although this basic issue of defining the target construct accurately pervades the work of people in all three groups of instrument users, the practical approaches to selecting appropriate instruments differ across the three groups. In this brief article, I concentrate on advice for the intermediate group of instrument users. But, first, I briefly consider the other two groups of users. As for the expert group, no advice is needed. People in this group already "know the ropes."

As for the novice group of users, advice is easy. It consists of three simple points. First, resist the temptation to build your own instrument. While it sounds expedient to develop your own instrument, it is almost always a bad idea to try to do so. Why? Because instrument development and validation is a complex and time-consuming task, one that requires

considerable technical expertise and experience. Second, find out what measurement instruments are most frequently used for the construct(s) you are studying and use these instruments. For example, if most of the studies covered in your literature review use the *Piers-Harris Children's Self-Concept Scale*, then use that instrument. Adopting commonly used instruments helps to integrate your findings with previous research. Third, use simple search strategies, such as those outlined by Moore, Bryant, and Perloff (1999) [reprinted in this issue of the newsletter], to get more detailed information about the instruments you plan to use, that is, the instruments most frequently used in studying your topic.

Advice for the Intermediate Group

Now for the middle group of instrument users: What more is expected of them than is expected of the novice group? The research community, including the research mentor, has at least the following expectations for persons in the intermediate group of instrument users.

1. You should have increased sophistication in searching out appropriate instruments for specific applications. Later sections of this article help to ensure that this expectation is met.
2. You should have the ability to evaluate the reliability and validity of instruments for specific applications. Such evaluation may be beyond novices, who probably do not have sufficient training to make independent evaluations of these matters. But people in the middle group do have such training – that is partly why they get classified as intermediate users. You need to apply this training to make your own judgments about reliability and validity.
3. You need to be able to judiciously combine two or three instruments in a single study. This requires not only knowledge of technical matters like reliability and validity, but also the practical common sense that comes from having conducted at least a few research studies in the past.

Practical Tips About the Search Process

As you begin your search for suitable instruments, here are a few practical tips.

Selecting Instruments (continued)

1. The first bit of advice is the same as for the novice group: Resist the temptation to build your own instrument. Leave that for the experts. Building your own instrument should be a last resort. Almost certainly, several instruments already exist for the constructs you intend to study. The trick is finding them – which is the main point of this article. (The exception to this first bit of advice is a project where the entire focus of attention is on building a new instrument.)

2. Use the invaluable sources of information described here to find an existing instrument or instruments for your research project.

3. Devote ample time to your search for instruments. You cannot do a respectable job in a mere 20 minutes of Internet searching. It will take several hours for your initial search, followed by hours of more detailed study of the most promising possibilities, followed by more time securing copies and perhaps permission to use the instrument.

4. Do not rely entirely on brief descriptions of instruments. You have to get your hands on the actual instrument, examine its directions, and review the items and administration procedures before deciding whether the instrument is right for your purposes.

5. Be sensitive to the varied roles that measurement instruments play in the research enterprise. A measurement instrument provides the operational definition of the dependent variable in a research study. However, measurement instruments can also be important in describing characteristics of the participants and as covariates or blocking variables in more sophisticated research designs.

6. It is tempting to use multiple instruments in order to get a rich description of the construct of interest. In fact, use of multiple instruments is the mark of a researcher with some experience. However, be careful not to overdo it. You will carry out your research with real people, who probably don't care that much about your project. Be careful you don't bore respondents to the point that you jeopardize the validity and reliability of their responses.

7. If you are reproducing or adapting an existing

instrument, make sure you obtain proper permission from the author or publisher. Be aware that obtaining such permission can be time-consuming.

Sources for Your Search

Now for the search itself. There are a variety of excellent sources available to help you find suitable instruments for a research project. Here I identify these sources and describe their strengths and weaknesses. The sources for finding instruments fall into four major categories.

Electronic Databases. Currently, there are two available electronic databases specifically devoted to cataloging existing instruments:

- * Health and Psychosocial Instruments (HaPI), a product of Behavioral Measurement Database Services (BMDS), available Online and on CD-ROM from Ovid Technologies.

- * Educational Testing Service Test Collection, available at <http://sydneyplus.ets.org>

Probably the best place to start your search is with one or both of these two electronic databases. Each database contains brief descriptions of thousands of tests and has excellent searching capabilities. Simply enter the name of the construct of interest (e.g., anxiety or self-esteem) and get a list of measurement instruments relevant for that construct.

These databases have three main advantages. First is their easy searching capability. Your list of *possibly* useful instruments is just a few clicks away. Second is their comprehensiveness. Each database contains information about thousands of instruments. For the more commonly studied psychological constructs (such as depression, self-concept, anxiety, introversion, intelligence, spatial ability, etc.) it is not unusual to find at least a dozen existing instruments listed in these databases. Third, the databases contain not only the names of instruments but also brief descriptive information, including the instrument's purpose, intended target groups, scores, and publisher or other source.

The electronic databases have two main drawbacks. First, although they contain brief

Selecting Instruments (continued)

descriptive information about each entry, they do not provide evaluations of the quality of the entries. The bad get equal billing with the good. Thus, you, as the ultimate user, still have work to do in deciding if the instrument is appropriate for your project. Second, although the databases guarantee that a particular instrument existed somewhere at sometime, they do not guarantee the ready availability of that instrument. Users must investigate availability separately. Some of the instruments in the databases can be purchased from one of the major test publishers. BMDS also provides copies and scoring instructions for many of the instruments in the HaPI database. On the other hand, some of the instruments may be out-of-print or may be available only in the appendix of a journal article, thus requiring permission from the journal's publisher for reproduction.

Hard Copy Listings. There are currently at least three hard copy listings of tests. Although these hard copy listings are similar in many ways to the electronic listings, they are also different in several important ways.

Tests in Print (Murphy, Plake, Impara, & Spies, 2002), usually referred to simply as *TIP*, is now in its sixth edition. New editions appear about every four years. *Tests: A Comprehensive Reference for Assessments in Psychology, Education and Business* (Tests; Maddox, 2003) is quite similar to *TIP*; new editions of *Tests* appear about every five years. Both *TIP* and *Tests* limit their entries to tests that are in English and are available from regular publishers. (A "regular" publisher means a commercial organization that is in the business of developing and selling tests.) The *Directory of Unpublished Experimental Mental Measures* (Directory; Goldman & Mitchell, 2003), as suggested by its title, lists tests that are *not* available from a publisher. Entries in the *Directory* are taken, either by reference or in full, from a journal article. Thus entries in the *Directory* are published in the sense that they appear in print somewhere, but they are not available from a regular publisher.

Whereas you use typical "keyword and click" methods to search the electronic databases, you use a printed index to search these hard copy listings.

Although it is certainly easier to complete a quick search electronically, browsing a hard copy listing is potentially useful; you are more likely to encounter a serendipitous result by leisurely browsing a hard copy. As with the electronic listings, all three of these hard copy listings provide brief descriptions of the tests they include (e.g., purpose, target group, and scores). Also like the electronic listings, these sources do *not* provide evaluative commentary about the tests. Most academic libraries have copies of all three of these hard copy listings.

Test Reviews. Two major sources provide professional, qualitative reviews of tests. The most well known is *Buros Mental Measurements Yearbook* (Spies & Plake, 2005), now in its 16th edition. Known simply as *Buros* or *MMY*, new editions appear about every two or three years. Each edition contains reviews of approximately 400 tests, concentrating on new or recently revised tests. Two independent reviews are given for most entries.

Buros reviews are available in three forms. First, there are the traditional hard copy volumes, available in most academic libraries. Second, some libraries subscribe to the *Buros* on-line reviews (an Ovid Technologies product, like HaPI), which include all reviews since the 10th *MMY*. For libraries subscribing to this service, there is no user fee for accessing a review. Third, for a fee, a *Buros* review can be accessed via the Internet at <http://www.unl.edu/buros/>.

The second major source of test reviews is *Test Critiques* (Keyser, 2004), now available in 11 volumes, with new volumes appearing at varying intervals. *Test Critiques* covers fewer instruments than *MMY* and tends to concentrate on the more widely used tests.

Reviews in *MMY* and *Test Critiques* provide a very important professional service. They are the only sources that give evaluations of the quality of a wide variety of tests. You should definitely consult these sources to see if they contain evaluative reviews of any tests you are considering for your project and, of course, read the reviews if they are available.

These collections of test reviews have two principal drawbacks. First, they cover only regularly published tests and not even all of those. Thus, many of

Selecting Instruments (continued)

the tests you identify in searching electronic or hard copy listings of tests will not have been reviewed. Second, because the reviewing process takes time, reviews will not be available for tests that have become available only recently. For both of these reasons, you need to rely on your own judgment about the suitability of many of the instruments you are considering for your research. Members of the intermediate group of test users should be able to do that.

Publishers' Catalogs and Websites. Like L.L. Bean and Sears, the major test publishers describe their products in catalogs, available in hard copy as well as on the Internet. Of course, many of these published tests are also briefly described in the electronic and hard copy listings presented earlier. However, a publisher's catalog will contain more complete descriptions of a test. Thus, if you are considering use of an instrument from one of the major test publishers, you should definitely consult the publisher's catalog. Especially important are the detailed facts about the latest editions, types of answer media, scoring services, and costs. Regarding costs, be aware that most publishers give discounts, usually from 25 – 40%, for research use of their tests. A publisher's website usually contains the form needed to secure this research discount.

Although the publisher's catalog is the best source of information about such practical matters as cost, latest editions, etc., it is not an unbiased source of information about the quality of a test. After all, the publisher is in the business of selling the test and will, therefore, present it in the best possible light. Look elsewhere, including your own judgment, to evaluate test quality.

Other Sources

I have outlined here the main sources of information about tests appropriate for use at the intermediate level. Several more advanced sources, especially appropriate for technical work in test construction and specialized applications, are also available. For descriptions of these other sources see Hogan (2003).

Concluding Thoughts

Selecting the best instruments for your research project is a crucial part of the entire research process. An instrument is an operational definition of the construct under investigation. Novices should simply use whatever has been used most frequently in the past. More is expected of those who, by training and experience, have moved beyond the novice category. Showing sophistication in the use of sources of information about existing tests is part of this "more." It is worth taking the time to become proficient in the use of these sources of information and then actually use them. Doing so should contribute to a richer, more meaningful research project.

References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Brockway, J. H., & Bryant, F. B. (1998). You can't judge a measure by its label: Teaching students how to locate, evaluate, and select appropriate instruments. *Teaching of Psychology*, 25, 121-123.
- Goldman, B. A., & Mitchell, D. F. (2003). *Directory of unpublished experimental mental measures* (Vol. 8). Washington, DC: American Psychological Association.
- Hogan, T. P. (2003). *Psychological testing: A practical introduction*. New York: Wiley.
- Keyser, D. J. (Ed.). (2004). *Test critiques* (Vol. XI). Austin, TX: Pro-Ed.
- Maddox, T. (2003). *Tests: A comprehensive reference for assessments in psychology, education, and business* (5th ed.). Austin, TX: Pro-Ed.
- Moore, D., Bryant, F. B., & Perloff, E. (1999). Measurement instruments at your fingertips. *Eye on Psi Chi*, 3(2), 17-19.
- Murphy, L. L., Plake, B. S., Impara, J. C., & Spies, R. A. (Eds.). (2002). *Tests in Print VI*. Lincoln, NE: University of Nebraska Press.
- Spies, R. A., & Plake, B. S. (Eds.). (2005). *The sixteenth mental measurements yearbook*. Lincoln, NE: University of Nebraska Press.

Coming Soon to a Website near you . . .
BEHAVIORAL MEASUREMENT DATABASE SERVICES

About Us
Mission Statement
Type of Instruments
Explanation of Database Fields
Tips on Using the Database
Document Delivery
Newsletter
Meet the HaPI Staff
Frequently Asked Questions
Contact Us



HaPI Advisory Board

Aaron T. Beck, MD
University of Pennsylvania School of Medicine

Timothy C. Brock, PhD
Ohio State University, Psychology

William C. Byham, PhD
Development Dimensions International

Donald Egolf, PhD
University of Pittsburgh, Communication

Sandra J. Frawley, PhD
*Yale University School of Medicine,
Medical Informatics*

David F. Gillespie, PhD
*Washington University,
George Warren Brown School of Social Work*

Robert C. Like, MD, MS
*University of Medicine and Dentistry of New Jersey,
Robert Wood Johnson Medical School*

Joseph D. Matarazzo, PhD
Oregon Health Sciences University

Vickie M. Mays, PhD
*University of California at Los Angeles,
Psychology*

Michael S. Pallak, PhD
Behavioral Health Foundation

Kay Pool, President
Pool, Heller & Milne, Inc.

Ora Strickland, PhD, RN, FAAN
*Emory University Woodruff,
School of Nursing*

Gerald Zaltman, PhD
*Harvard University,
Graduate School of Business Administration*

Stephen J. Zyzanski, PhD
Case Western Reserve University, School of Medicine

.....
*Search for measurement instruments
in the Health and Psychosocial Instruments (HaPI)
database ONLINE or on CD-ROM!*



Produced by BMDS • Behavioral Measurement Database Services

SUBSCRIBE TO HaPI!



*Call Ovid Technologies at 800-343-0064
for pricing and to order today!*

*Obtain copies of measurement instruments
from BMDS Instrument Delivery!*



*Call BMDS at 412-687-6850, fax 412-687-5213, or
e-mail bmdshapi@aol.com (\$20/\$30 processing/handling fee)*

.....

HaPI Thoughts

BMDS

**Behavioral
Measurement
Database
Services**

**Measurement
Thesaurus**

First Edition

measure, v. assess, appraise, apprise, gauge, evaluate, survey, observe, judge, monitor, calibrate, tabulate, figure, quantify, tally, count, rate, rank, estimate, weigh, value, equate, differentiate, assay, plumb, fathom, size up, examine, probe. See **HaPI**.

Special Issue
on PSYCHOLOGY

The BEHAVIORAL MEASUREMENT Letter

Behavioral
Measurement
Database
Services

In This Issue:

- Introduction to This Issue
Fred B. Bryant, Editor 1
- Measurement Instruments at Your Fingertips
Daniel Moore, Fred B. Bryant & Evelyn Perloff 3
- You Can't Judge a Measure by Its Label: Teaching
the Process of Instrumentation
Jennifer Howard Brockway & Fred B. Bryant 6
- Selecting Instruments for Behavioral Research:
Advice for the Intermediate User
Thomas P. Hogan..... 9
- HaPI Thoughts..... 15

Special Issue
on NURSING to follow