The BEHAVIORAL MEASUREMENT *Letter*

Behavioral Measurement Database Services

Enriching the health and behavioral sciences by broadening instrument access

Vol. 6, No. 1
Winter 1999

Introduction to This Issue

In this issue of *The* Behavioral Measurement *Letter*, Howard Shaffer examines problems in defining and measuring problem gambling in "Addiction and Gambling Disorders: On Matters of Measurement and Validity." His piece illustrates well a fundamental problem in the health, behavioral, and social sciences, construct definition. In the health sciences, construct definition has obvious implications for treatment as well as measurement, for application as well as theory. Does the construct represent a pathologic condition itself or a symptom of pathology? If it is a symptom, is it part of a syndrome (a configuration of symptoms that characterizes a pathologic condition)? If a condition, does it coexist with other pathologic conditions? In terms of measurement in general, it is obvious that if one cannot define what one is measuring, it is meaningless to speak of validity and reliability. As Shaffer points out in the case of addictive disorders, both theory and the measurements that underlie theory must differentiate between first-order and second-order constructs (i.e., between unitary constructs and those comprised of other, subordinate constructs), and for both theory and its application, one must understand how constructs interrelate.

The use of measurement modeling as a tool for differentiating between first- and second-order constructs and for understanding how constructs interrelate are topics covered by Fred Bryant in his column, "Measurement Modeling: Identifying the Constructs Underlying the Center for Epidemiologic Studies Depression Scale (CES-D)." This article is the third in his series on measurement modeling as an important, increasingly used psychometric tool. In this installment, Bryant illustrates the application and value of measurement modeling in construct definition, in this case, defining the construct or constructs measured by the CES-D. He reports on a study in which "distress" as measured using the CES-D is modeled as a unidimensional construct, a three-factor construct, a four-factor construct, and a second-order construct consisting of four first-order constructs. Bryant's piece further illustrates the centrality of construct definition in the social and behavioral sciences, underscoring points made in Shaffer's discussion of construct definition and measurement in research on pathologic gambling.

As an old song says, "Little things mean a lot." In research, as in day-to-day living, simple things, "obvious" things, are often overlooked or taken for granted. Time and again we ignore the simple or obvious, much to our peril, and often with ill consequences. Similarly, negative consequences can result from overlooking the nonobvious, the subtleties. Ora Strickland's column, "The Practical Side of Measurement," reminds us that in research as in life, one must attend to the obvious and the nonobvious details, that good measurement involves more than just selecting instruments based primarily or solely on psychometric considerations and then administering them with little or no regard to subjects' capabilities and condition. One size does *not* fit all, that is, as Strickland points out, "instruments that have been reliable and valid in one situation may not function well under other circumstances," instruments that are valid for

(continued on next page)

Introduction *(continued from page 1)*

one age-group of children may not be valid for others, instruments that can be administered easily to able, middle-aged adults may be difficult or impossible to administer to severely debilitated or elderly persons. Strickland's piece demonstrates that good measurement involves thoughtful attention to detail as well as technical matters.

• • •

Address comments and suggestions to The Editor, *The* Behavioral Measurement *Letter*, Behavioral Measurement Database Services, PO Box 110287, Pittsburgh, PA 15232-0787. If warranted and as space permits, your communication may appear as a letter to the editor. Whether published or not, your feedback will be attended to and appreciated.

We also accept short manuscripts for The BML. Submit, at any time, a brief article, opinion piece or book review on a BML-relevant topic to The Editor at the above address. Each submission will be given careful consideration for possible publication.

HaPI reading . . .

Al K. DeRoy, Editor

Phone: (412) 687-6850 Fax: (412) 687-5213

E-mail: bmdshapi@aol.com

One thought fills immensity.
William Bleke

Addiction and Gambling Disorders: On Matters of Measurement and Validity

Howard J. Shaffer

The purpose of this article is to examine some of the validity and measurement issues associated with a contemporary perspective on addictive behaviors in general and disordered gambling in particular. Scientists have tried to understand addictive behaviors (e.g., alcoholism, drug abuse and dependence, pathological gambling) from a variety of viewpoints. For example, neurobiologists have tended to see addiction as a problem of neurotransmitters, receptor sites and transporter mechanisms. Psychologists have gravitated toward psychodynamics, learning, trauma, and other intrapersonal and environmental influences as vital to an understanding of addiction. Sociologists often see addiction as a result of family, community, social network, cultural and other social setting factors. Public policy makers often consider addiction to result from the laws that permit or discourage certain patterns of behavior. With some notable exceptions, there has been a tendency to view addictive behavior from the limited ideological perspectives associated with a particular discipline.

When clinicians examine people who are struggling with addiction, they often observe a range of concurrent or sequential problems. These apparently coexisting or overlapping difficulties suggest there may be an underlying addiction "factor." This matter complicates the task of psychometricians. For example, consider a measure of gambling that correctly identifies cases of alcoholism, drug abuse, and depression even though these behaviors were not the target of the gambling screen. What can we say about the screen? In their analysis of a very similar problem (i.e., the validity of a measure of psychological "adjustment" and its positive correlation with a measure of "intelligence"), Rosenthal and Rosnow (1975) conclude: "That question is difficult to answer, but we could not claim on the basis of these results to understand our [gambling] test very well. It was not

(continued on next page)

intended, after all, to be a measure of [alcoholism or depression]. In short, our test has good concurrent validity but poor differential validity. It does not correlate differentially with criteria for different types of observation" (p. 70). Rosenthal and Rosnow remind us that we must be sensitive to the relationship between appropriate and inappropriate criteria during the process of establishing construct validity.

What can we say about the gambling problems that this screen identifies? Gambling problems may be the cause of excessive alcohol or drug use. Gambling problems also can stimulate depression. Depression also may result indirectly from the confluence of gambling with co-existing alcoholism or drug abuse. Alternatively, alcohol or drug use may stimulate depression that, in turn, encourages intemperate gambling–as an anodyne for emotional discomfort. To determine whether a pattern of excessive gambling, drinking, or drug using represents a distinct disorder, scientists must establish construct validity. While these and other possible permutations of these problems can explain disordered gambling behavior, it also is possible that all of these behaviors reflect the expression of an underlying addiction syndrome.

A syndrome is a group of symptoms that collectively indicate or characterize an undesirable condition or quality, a disease, a psychological disorder, or another abnormal condition.[1] Not all elements of the syndrome need be present all of the time. It is the *pattern* of elements rather than the intensity of any one component that is critical to identifying a syndrome. Establishing criteria and instruments to identify a syndrome represent a daunting measurement task. This challenge has led to the variety of problems associated with psychometric attempts to distinguish non-pathologic drug use, excessive drug use, and drug abuse from drug dependence and addiction (e.g., Shaffer, 1996; 1997) or "heavy" and problem gambling from pathological gambling.

[1] *The American Heritage Dictionary of the English Language, Third Edition* (1992). New York: Houghton Mifflin.

Given that the field of psychiatry in general, and disordered gambling in particular, has no gold standards (Faraone & Tsuang, 1994), there is little rationale for placing one instrument, or even clinical diagnosis, as the gold standard against which another instrument is measured. As Faraone and Tsuang note, "many studies of psychiatric diagnosis compute accuracy statistics. However, these assess the accuracy of one diagnostic approach (e.g., DSM-IV) with respect to another (e.g., expert clinical diagnoses). They do not assess the accuracy of diagnostic procedures with reference to the 'true' but unobservable latent state of illness" (p. 651). Determining the validity of a latent state is complicated further when we consider this circumstance to reflect a syndrome rather than a more coherent unidimensional disorder.

The problems associated with determining the validity of gambling or addictive behaviors are not limited to the matters of syndromal behaviors, but actually begin with the very definition of validity. Validity is the capacity of an instrument to measure what it purports to measure. Validity is neither a static nor an inherent characteristic of a screening instrument. As Goldstein and Simpson (1995) suggest, "validity refers to the questions 'for what purpose is the indicator being used?' . . . and 'how accurate is it for that purpose?'" (pp. 229-230). Determining instrument validity is an unending and dynamic investigative process. We cannot simply conclude that an instrument has been shown to be valid for all purposes and all settings. "An indicator (e.g., an instrument, such as a test, a rating, or an interview) can be valid for one purpose, but not for another" (Goldstein & Simpson, 1995, p. 230).

Directed by theoretical and ultimately practical purposes, validity is the dynamic consequence of applying an instrument to a measurement task. In the field of gambling studies, however, theory is conspicuously absent from most prevalence research (Shaffer, Hall, & Vander Bilt, 1997). When conventional wisdom and scientific theory shift or change, the validity of an instrument can be nullified abruptly. The history of the South Oaks Gambling Screen (Lesieur & Blume, 1987) provides an example of the relative nature of validity. Although for

(continued on next page)

some time researchers considered that the SOGS lifetime measure had "been found valid and reliable" (Volberg, 1994, p. 238), some investigators now suggest that the SOGS lifetime measures ". . . over-state the actual prevalence of pathological gambling" (Volberg, 1997, p. 41). Assessments of validity–even by the same scientist–can wax and wane depending upon the objectives established for an instrument's use and the shifting cultural values that energize those purposes. For example, in her study of Washington state adolescents, Volberg states that "Our approach, while conservative, is intended to focus as clearly as possible on those adolescents who show incontrovertible signs of problematic involvement in gambling" (Volberg, 1993, p. 17). Volberg uses this conservative approach, in part, because she is ". . . uncomfortable with a method [i.e., the SOGS] that classifies 8% of adolescents as problem or probable pathological gamblers" (1993a, p. 34). In other words, the criteria for identifying pathological gambling are not fixed, but vary depending on whether a given set of criteria yields an "acceptable" prevalence rate among a particular group.

Absent a gold standard for determining pathological gambling, we do not know whether the SOGS over-estimates the prevalence rate of gambling disorders or whether clinical assessment and DSM-based instruments under-estimate the prevalence rate. This problem of anchoring reveals itself often as scientists attempt to determine how best to frame prevalence estimates.

Ultimately, the field of gambling studies is in need of research that can provide additional evidence about the construct validity of the notions that represent disordered gambling (e.g., problem, pathological, probable pathological, etc.). While new evidence (Shaffer, Hall, & Vander Bilt, 1997) provides considerable support for a recognizable and identifiable pattern of behaviors that can be considered disordered gambling, there are important conceptual questions that still remain. A primary construct validity question requires scientists to focus on whether disordered gambling is a primary or a secondary disorder. For example, according to the DSM-IV, a person meeting all of the criteria for pathological gambling is *not* considered a pathological gambler if he or she also concurrently meets the criteria for a Manic Episode, and the Manic Episode is responsible for the excessive gambling (APA, 1994). In this case, pathological gambling is not considered a unique disorder, but rather a cluster of symptoms associated with another disorder. If pathological gambling represents a primary disorder, then it can emerge in the absence of other comorbidity and cause sequelae independent of any other condition. However, if it is a secondary disorder, subordinate to other dysfunctional behavior, then pathological gambling will only exist as a consequent of another condition (e.g., manic episode, anti-social personality, alcohol abuse, obsessive-compulsive disorder, or adolescence; Jessor & Jessor, 1977). Although Shaffer et al. (1997) suggest that researchers of disordered gambling have measured a relatively stable and robust phenomenon, the field of gambling studies has not yet established with ample certainty that this phenomenon represents a *unique* construct.

Since gambling researchers have paid very little attention to this important conceptual issue of discrete and comorbid phenomena[2] – and the associated matter of differential validity – the possibility remains that pathological gambling is not a discrete primary disorder. Alternatively, it is possible that pathological gambling is a discrete and primary disorder (i.e., it will exist independent of any other disorder). Current research is beginning to suggest that, in many instances, other primary psychiatric disorders provide better explanations of excessive gambling (i.e., disordered gambling is considered to be subordinate to other primary

[2] Briggs, Goodin, and Nelson (1996) have conducted preliminary research on this issue. Their "findings indicate no significant crossover of addictions between the two samples. This would seem to indicate that alcoholism and pathological gambling are independent addictions. One might infer from this that they also involve independent processes" (Briggs et al., pp. 517-518). In spite of these conclusions, the Briggs et al. study employed a unique subject sample (e.g., use of self-help group members) that likely represents the tails of two special self-selected distributions; they also employ a small sample size. Taken collectively, these factors encourage us to view their results as tentative and their conclusions as uncertain.

(continued on next page)

disorders). According to the current version of the DSM-IV (American Psychiatric Association, 1994), pathological gambling can have either primary or secondary status. In some cases it is considered to be a primary disorder (i.e., independent of other diagnoses), and in other cases it is considered to be the sequelae of another disorder. The implications of this issue are potentially significant for the development of both treatment prescriptions and social policy initiatives designed to ameliorate or regulate gambling-related problems.

Whether we view disordered gambling as primary or secondary, intemperate gambling inflicts human suffering. Whether a primary or secondary problem, pathological gambling often warrants professional assistance. If it is a disorder secondary to another problem, clinicians should focus on providing attention to the gambling issues in addition to the problems related to the primary disorder. Future research will help clarify these theoretical, research, and clinical issues. The scientific measurement of these problems will demand careful conceptual guidance. It is difficult indeed to distinguish the "cart" from the "horse" without some systematic experience with carts and horses.

The author thanks Matthew Hall and Joni Vander Bilt for their important contributions and thoughtful comments on earlier versions of this article. This article was supported in part by grants from the National Center for Responsible Gaming *(Kansas City, MO) and the* Center for Substance Abuse Treatment *(#1U98TI00846).*
Correspondence regarding this article should be sent to Dr. Shaffer at Harvard Medical School, The Division on Addictions, 220 Longwood Avenue, Boston, MA 02115.

References

American Psychiatric Association. (1994). *DSM-IV: Diagnostic and statistical manual of mental disorders.* (Fourth ed.). Washington, DC: American Psychiatric Association.

Briggs, J.R., Goodin, B.J., & Nelson, T. (1996). Pathological gamblers and alcoholics: Do they share the same addiction? *Addictive Behaviors, 21*, 515-519.

Faraone, S.V., & Tsuang, M.T. (1994). Measuring diagnostic accuracy in the absence of a "gold standard". *American Journal of Psychiatry, 151*, 650-657.

Goldstein, J.M., & Simpson, J.C. (1995). Validity: Definitions and applications to psychiatric research. In M. T. Tsuang, M. Tohen, & G. E. Zahner (Eds.), *Textbook in psychiatric epidemiology* (pp. 229-242). New York: Wiley-Liss.

Lesieur, H.R., & Blume, S.B. (1987). The South Oaks Gambling Screen (SOGS): A new instrument for the identification of pathological gamblers. *American Journal of Psychiatry, 144*, 1184-1188.

Rosenthal, R., & Rosnow, R.L. (1975). *Primer of methods for the behavioral sciences.* New York: Wiley & Sons.

Shaffer, H.J. (1996). Understanding the means and objects of addiction: Technology, the Internet, and gambling. *Journal of Gambling Studies, 12*, 461-469.

Shaffer, H.J. (1997). The most important unresolved issue in the addictions: Conceptual chaos. *Substance Use & Misuse, 32*, 1573-1580.

Shaffer, H.J., Hall, M.N., & Vander Bilt, J. (1997). *Estimating the prevalence of disordered gambling behavior in the United States and Canada: A meta-analysis.* Boston: Presidents and Fellows of Harvard College.

Volberg, R. (1993). Gambling and problem gambling among adolescents in Washington state. Albany, NY: Gemini Research.

Volberg, R.A. (1994). The prevalence and demographics of pathological gamblers: Implications for public health. *American Journal of Public Health, 84*, 237-241.

Volberg, R.A. (1997). *Gambling and problem gambling in Mississippi: A report to the Mississippi Council on Compulsive Gambling* (Social Research Report Series 97-1). Mississippi State: Mississippi State University, Social Science Research Center.

Addiction and Gambling *(continued from page 5)*

Howard J. Shaffer, PhD, CAS, is Associate Professor, Director of the Division on Addictions, and founding Director of the Norman E. Zinberg Center for Addiction Studies at Harvard Medical School. He is also founder and currently President of the Board of Trustees of the American Academy of Health Care Providers in the Addictive Disorders, the first international credentialing body for clinicians working that field, Editor of The Journal of Gambling Studies, Associate Editor of The Journal of Substance Abuse Treatment, on the editorial boards of five other journals, a reviewer for The New England Journal of Medicine, and author of six books. In addition to measurement methods in gambling and other addictions, his interests include the natural history of addiction, impulse control regulation, and social perceptions of addiction.

Measurement Modeling: Identifying the Constructs Underlying the Center for Epidemiologic Studies Depression Scale (CES-D)

Fred B. Bryant, PhD

In a previous column (Bryant, 1998), I described a powerful, new data-analytic approach to construct validation known as "measurement modeling." This approach uses state-of-the-art multivariate statistical tools to systematically compare alternative ways of conceptualizing the constructs that a particular research instrument taps. By fine-tuning our understanding of what instruments actually measure, measurement modeling: (a) better enables one to choose the most appropriate instruments for the intended purpose; (b) improves conceptual clarity by identifying constructs that are truly unitary and by decomposing multidimensional constructs into their constituent parts; (c) highlights gaps in measurement coverage for instrument development; (d) often leads to refinements in existing instruments, creating modified measures with improved conceptual and statistical precision; and (e) identifies how to score responses to instruments so as to capture the underlying construct(s) with maximum reliability. In the present column, I describe a published example of measurement modeling (Sheehan, Fifield, Reisine, & Tennen, 1995) that was conducted on one of the most popular measures of distress used in the behavioral sciences, the Center for Epidemiologic Studies Depression Scale (the CES-D; National Institute of Mental Health, 1977; Radloff, 1977). Here I describe how the researchers used measurement modeling to improve understanding and future use of the CES-D.

To review the basics as I covered them in previous columns (Bryant, 1997, 1998), measurement modeling, also known as confirmatory factor analysis, is a special form of structural equation modeling that investigates the "structure" underlying a set of measures collected from a group of people. "Structure" refers to the ways in which responses to the individual measures interrelate (if they do) to define one or more underlying constructs

(continued on next page)

(factors). Questions that strongly reflect a particular factor are said to have strong "loadings" on that factor or to "load" highly on that factor. In other words, each question's loading on a particular factor indicates how strongly that question defines the underlying construct that that factor taps. Thus, through measurement modeling, researchers can: (a) determine the appropriate number of constructs or factors underlying responses to a set of measures; (b) determine how these factors relate to one another; (c) quantify how strongly each measure characterizes each underlying factor, thereby pinpointing the specific subsets of questions that define the constructs; and (d) interpret the meaning of each factor in order to label each construct in theory-relevant terms. Moreover, measurement modeling enables researchers to compare competing models of the constructs that an instrument assesses, determine which model best explains responses to the instrument, and test whether the structure of responses to an instrument is stable across multiple groups or multiple time points.

The following summary of research on the CES-D illustrates measurement modeling and its value. The CES-D consists of 20 questions that reflect various symptoms of depression, including negative mood, feelings of guilt, worthlessness and helplessness, immobilization, poor appetite, and sleep disturbance. Respondents report the frequency with which each symptom occurred during the previous week using a four-point scale as follows: 0 ("rarely, that is, less than 1 day"); 1 ("some of the time, 1 to 2 days"); 2 ("a moderate amount of the time, 3 to 4 days"); and 3 ("most or all of the time, 5 to 7 days"). Although the CES-D is widely used to measure distress in the general population and among patients with chronic disease, there is no agreement about the underlying construct(s) that it taps. On the one hand, many investigators have treated the CES-D as measuring a single underlying construct (i.e., a unidimensional measurement model), and when coding responses to the instrument have simply summed responses to all questions to obtain a global "total score." However, in initially developing the CES-D, Radloff (1977) identified four constructs (i.e., a multi-dimensional model) that underlie responses to the instrument which she labeled Depressed Affect, Low Positive Affect, Somatic Symptoms and Retarded Activity, and Impaired Interpersonal Relations. Other researchers have argued that the Depressed Affect and Low Positive Affect factors really reflect a single underlying construct, and therefore treated the CES-D as measuring three underlying constructs consisting of affect (combining Depressed Affect and Low Positive Affect) and the somatic/vegetative symptoms and interpersonal deficits that Radloff originally found.

Which of these measurement models best explains responses to the CES-D, and is this measurement model equally reliable in assessing the same individuals over time? Sheehan, Fifield, Reisine, and Tennen (1995) set out to answer these questions. They began with data from a sample of 813 rheumatoid arthritis patients who completed the CES-D as part of a telephone interview once a year for three years. To analyze these data, Sheehan, et al. decided to use the most popular software program for conducting measurement modeling – LISREL 8, the LInear Structural RELationships computer analysis package, version 8. They used LISREL 8 to impose four alternative measurement models on the CES-D data: (a) a single-factor (total score) model that assumes depression is unidimensional (Model 1); (b) a three-factor model that assumes that the CES-D taps interrelated dimensions of affect, somatic/vegetative symptoms, and interpersonal deficits (Model 2); (c) a four-factor model that assumes the CES-D taps correlated dimensions of depressed affect, low positive affect, somatic/vegetative symptoms, and interpersonal deficits (Model 3); and (d) a second-order factor model consisting of the same four factors as in Model 3 but with a single higher-order depression construct that explains the relationships among the four lower-order factors (Model 4). This latter model hypothesizes that the four first-order factors are correlated with one another and thus their interrelationships can be explained by a second-order "super" factor that influences each dimension.

(continued on next page)

Measurement Modeling *(continued from page 7)*

Sheehan et al. used LISREL 8 to compare how well each of these four measurement models explains responses to the CES-D. They began by analyzing the data from the first of the three data collection points. LISREL 8 revealed that although the single-factor model provided a reasonable fit to time 1 data, the three-factor, four-factor, and second-order factor models all fit time 1 data significantly better. Furthermore, both the four-factor model and second-order factor model provided a significant improvement in fit over the three-factor model, though the former two models were not significantly different from one another in their goodness-of-fit.

Having compared the fit of the four alternative models at time 1, Sheehan, et al. (1995) next conducted analyses to cross-validate the four-factor and second-order factor models across all three time points. Examining the four-factor model, they found that it had equivalent factor loadings and equivalent correlations among its factors across all three time points. In addition, the amount of variance that the four-factor model explained in each CES-D question was equivalent over time. An identical pattern of results emerged for the second-order factor model. As Sheehan, et al. emphasized, the stability of these models over time in no way implies that people's responses to the CES-D questions will necessarily be stable over time. Rather, it means that the measurement structure underlying responses to the CES-D is equally reliable across time. In other words, the meaning of distress is stable over time. This finding is important for researchers who use the CES-D in longitudinal research, because it demonstrates the validity of using the same four factors to score the instrument at each point in time.

Sheehan, et al.'s (1995) results indicate that the four-factor model is the most appropriate way of scoring the CES-D to measure depressive symptoms because it is psychometrically superior to the commonly used "total score" method. Their work further indicates that although a single "total score" provides a fairly reliable summary of global distress, it ignores the fact that respondents use the CES-D questions to describe four separate, though correlated, forms of negative experience: depressed affect, low positive affect, somatic/vegetative symptoms, and interpersonal deficits. People may not simply report more or less intensity in their distress, but may have more or less of particular types of depressive symptoms. Moreover, and of potential clinical significance, summarizing the CES-D in terms of a total score equates people who are higher on one dimension and lower on another dimension with people who are lower on the former and higher on the latter. In addition, it is possible for CES-D total score to decrease over time while distress along one particular dimension might actually increase; for example, a person might experience less negative affect over time but also experience less positive affect as well. Simply lumping together these multiple dimensions is not the most informative index of distress and may, in fact, produce misleading results.

The work of Sheehan, et al. (1995) illustrates how measurement modeling helps us better understand the constructs that instruments measure, determines the most reliable and informative method of scoring instruments so as to maximize conceptual clarity and statistical accuracy, and thus enables us to use instruments more effectively. Given its demonstrated value, measurement modeling is an indispensable psychometric tool in the behavioral sciences.

References

Bryant, F.B. (1997). The comparative anatomy of related instruments: An emerging specialty. *The Behavioral Measurement Letter, 4 (2)*, 7-9.

Bryant, F.B. (1998). Measurement Modeling: A tool for investigating the comparative anatomy of related instruments. *The Behavioral Measurement Letter, 5 (2)*, 14-17.

National Institute of Mental Health. (1977). Center for Epidemiologic Studies Depression Scale (CES-D). Rockville, MD: Author.

Radloff, L.S. (1997), The CES-D Scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1*, 385-401.

Sheehan, T.J., Fifield, J., Reisine, S., & Tennen, H. (1995). The measurement structure of the Center for Epidemiologic Studies Depression Scale. *Journal of Personality Assessment, 64*, 507-521.

Measurement Modeling *(continued from page 8)*

Fred Bryant is Professor of Psychology at Loyola University, Chicago. He has roughly 80 professional publications in the areas of social psychology, personality psychology, measurement, and behavioral medicine. In addition, he has coedited 5 books, including **Methodological Issues in Applied Social Psychology** *(New York, Plenum Press, 1993). Dr. Bryant has extensive consulting experience in a wide variety of applied settings, including work as a research consultant for numerous marketing firms, medical schools, and public school systems; a methodological expert for the U.S. Government Accounting Office; and an expert witness in several federal court cases involving social science research evidence. He is currently on the Editorial Board of the journal* **Basic and Applied Social Psychology**. *His current research interests include happiness, psychological well-being, Type A behavior, the measurement of cognition and emotion, and structural equation modeling.*

The Practical Side of Measurement

Ora L. Strickland

Using instruments that are reliable and valid to measure concepts of interest is a central concern of researchers. It is a well-accepted fact that one can not study well what can not be measured well. Therefore, a lot of time and energy are invested in developing instruments that seem to measure what they purport to measure. However, a reality of measurement that must never be forgotten is that instruments that have been reliable and valid in one situation may not function well under other circumstances. In the course of selecting or developing questionnaires or designing the measurement component of studies, one may become so bogged down with selecting instruments that have been previously shown to be psychometrically sound that practical aspects of measurement may be forgotten. When the practical side of measurement is not given adequate attention, data that result can be compromised. Measurement instruments need to be practical for respondents or subjects, and for the researcher as well. The more practical a measurement instrument or protocol is for the subject and the researcher, the more likely data generated will be reliable and valid.

Instruments are practical for subjects when they are appropriate for the population studied, easy to understand, simple to complete, and not too demanding of energy and time. When subjects are children, ill or frail, or have low literacy, the challenges of measurement become more pronounced. For example, it may be difficult to find an instrument to measure particular variables in children. When such an instrument is identified, care needs to be taken to assure that it is usable for the age group of children that the investigator will be studying. Since the reading and oral comprehension levels of children vary so much when there is even a couple of years difference in age, an instrument may be useful for one age group but not for another. Although available instruments should be used when they exist, it may be necessary to revise an instrument to make it usable for a specific age group.

(continued on next page)

The Practical Side of Measurement
(continued from page 9)

When subjects are quite ill, a long and cumbersome instrument may be too taxing for them, and thus its use with such subjects may result in questionable data. In this case, the investigator should consider using a shortened version of the instrument and consider administering it as an interview. If a shorter version of the instrument is not available, it may be worth the time required to shorten it. It also may be necessary to use a proxy respondent (such as a parent or caretaker) who is intimately familiar with and aware of the subject's views and situation. However, proxy respondents should be used only when absolutely necessary because there is no real substitute for the subject. When proxy respondents are used, they should be explicitly told that they need to answer based on how the subject would respond and not based on their own responses. When study results are reported, the use of proxy respondents should be noted, of course. If, however, an instrument addresses the variable(s) of interest and is designed to be used by raters or observers, it may be a good alternative to the use of proxies.

Aging or disabled subjects also offer a measurement challenge. When questionnaires are used with older subjects, they should be printed in a large typeface so that items will be easier to see. However, if the sample includes subjects who are blind, have arthritis, or have difficulty using their hands for writing, questionnaires should be administered via interview. When questionnaires are mailed to older subjects or subjects who lack sufficient literacy, it is useful to interview the subjects by phone, or at least a sample thereof, to serve as a reliability check. The phone interviewer can have the subject simultaneously read the mailed questionnaire. Having respondents read their mailed questionnaire as they are being interviewed by phone also improves the response rate.

The total measurement protocol for each data collection point also needs to be considered. I have found this to be of concern even in healthy subjects. When too much time and energy are required of subjects to complete data collection, subject recruitment will be negatively affected and the subject dropout rate will increase. In this regard, one should consider the following: Will the combination of data collection methods and instruments employed be so demanding that subjects could become so fatigued that data will be compromised? How much time and energy will subjects have to expend in each data collection session? Do subjects have physical or mental limitations that compromise the usefulness of specific instruments or data collection approaches? Sometimes it may be necessary to break a data collection point into two or more sessions in order to keep from overly fatiguing subjects. Also, using a variety of data collection approaches can break up the monotony of a data collection session and help prevent subject fatigue.

Instruments are practical for researchers when they are not only available but accessible, easy to administer and score, and not too demanding of time and other resources. It is important to remember that the costs associated with a measurement protocol not only include the financial cost of purchasing the instrument, but also costs for administration and scoring. If it is impractical to interview all subjects due to cost constraints, then a representative sample may be interviewed as a reliability check. If special training is required to administer or score the instrument, the time expended and costs associated with training data collectors or scorers are added expenses. Instruments that require computers for administration and scoring may be more financially costly but less costly as far as time expended in performing these tasks. Other factors that can increase the cost of instruments include needs for special equipment or a special setting for administration.

The practical side of measurement requires striking the right balance between measurement principles and practices and common sense. Highly complicated measurement protocols that are not practical to subjects or investigators are not useful in the long run.

(continued on next page)

(continued from page 10)

Ora L. Strickland, PhD, RN, FAAN, is Professor and the first to occupy the Independence Foundation Endowed Research Chair at the Nell Hodgson Woodruff School of Nursing at Emory University. She is founding Editor of the Journal of Nursing Measurement, on the editorial board/review panel of Advances in Nursing Science, Research in Nursing and Health, Scholarly Inquiry for Nursing Practice: An International Journal, and Nursing Leadership Forum. She co-edited the four-volume Measurement of Nursing Outcomes, has written or contributed to 18 books, and received two awards in health journalism (1998) for her column in the Baltimore Sun, "Nurses Station," and seven American Journal of Nursing Book of the Year awards. She was the youngest person ever elected to the American Academy of Nursing.

HaPInings

There are three items of interest relevant to BMDS and HaPI to report here – a letter to be published in the *Archives of Family Medicine*, a presentation to be made at the 1999 APA Convention, and an article in the quarterly magazine *Eye on Psi Chi*.

Letter on Identifying Instruments for Use in Family Practice

A letter written by Drs. Stephen Zyzanski, Professor of Family Medicine at Case Western Reserve University, and Evelyn Perloff, Director of BMDS, on identifying instruments for measuring behavioral variables found in family practice was accepted for publication by the *Archives of Family Medicine*. The letter was produced in response to an article published in the July/August 1998 issue of the *Archives* that emphasized the important roles of psychosocial factors in health and illness, but discussed a relatively small number of factors and instruments used to measure them. In their letter, Zyzanski and Perloff allude to the large number and broad spectrum of behavioral and psychosocial factors relevant to health, and point to HaPI as an excellent source for identifying and learning about instruments to assess these factors in clinical practice.

Presentation on Using HaPI as an Instructional Tool at 1999 APA Convention

BMDS staff will be presenting an interactive demonstration on the use of HaPI as an instructional tool on August 20 and 21, 1999 at the American Psychological Association's Annual Convention, being held this year in Boston. The presentation, tentatively titled "Using the Health and Psychosocial Instruments (HaPI) Database to Select Measures for Test Validation," is part of the APA's Miniconvention on Education and Technology. The interactive demonstration will employ exercises in which students (in this case APA members) use the HaPI database to identify, learn about, and compare and contrast instruments to measure "grief," "perceived control," and "empathy." The presenters will be Fred B. Bryant, Rebecca Guilbault, and Evelyn Perloff.

Article on Measurement, Instruments, and HaPI in Eye on Psi Chi

Daniel Moore, MA, Fred Bryant, PhD, and Evelyn Perloff, PhD, wrote an article titled "Measurement Instruments at Your Fingertips" that was published in the Winter 1999 issue of *Eye on Psi Chi*, the quarterly magazine of Psi Chi, the National Honor Society in Psychology. The article emphasizes the importance and ubiquity of measurement, measurement instruments, and means to identify appropriate measurement instruments, including HaPI, in the context of an undergraduate psychology class. The "teaser" on the magazine's front cover humorously summarized the article's content in two short sentences: "Looking for the right instrument? Don't worry, be HaPI!"

The BEHAVIORAL MEASUREMENT *Letter*

Behavioral Measurement Database Services

In This Issue:

A newsletter – touching base with the Who, What, Why, How, When, and Where of behavioral measurement

The Behavioral Measurement *Letter*
Behavioral Measurement Database Services
BMDS
PO Box 110287 • Pittsburgh, PA 15232-0787

DATED MATERIAL

Bulk Rate
U.S. Postage
PAID
Permit No. 1235
Pittsburgh, PA