RELIABILITY GENERALIZATION OF SCORES ON THE SPIELBERGER STATE-TRAIT ANXIETY INVENTORY

LAURA L. B. BARNES
Oklahoma State University

DIANE HARP
Tulsa Community College

WOO SIK JUNG
Bluffton College

A reliability generalization study for Spielberger's State-Trait Anxiety Inventory (STAI) was conducted. A total of 816 research articles utilizing the STAI between 1990 and 2000 were reviewed and classified as having (a) ignored reliability (73%), (b) mentioned reliability or reported reliability coefficients from another source (21%), or (c) computed reliability for the data at hand (6%). Articles in medically oriented journals were shorter and somewhat less likely to mention or compute reliability than nonmedically oriented articles, perhaps due to paradigm differences. Average reliability coefficients were acceptable for both internal consistency and test-retest, but variation was present among the estimates. State test-retest coefficients were lower than internal consistency coefficients. Score variability was predictive of internal consistency reliability for scores on both scales. Other predictors were the age of research participants, the form of the STAI, and the type of research design.

Many researchers erroneously believe that reliability is a property of a particular instrument. By referring to the reliability of a test or saying an instrument is reliable, what is generally implied is that once an instrument has been found to be reliable or unreliable, the status it is given is immutable; that reliability does not change (Henson, 2001; Vacha-Haase, Kogan, & Thompson, 2000). This assumption is incorrect. An instrument in a single study can produce scores that are reliable and then in a different study with different participants can produce scores that are unreliable. Reliability is not a prop-

Educational and Psychological Measurement, Vol. 62 No. 4, August 2002 603-618

erty of a test; rather, it is a property of the scores on a test for a particular sample of examinees (Vacha-Haase et al., 2000; Vacha-Haase, Ness, Nilsson, & Reetz, 1999). Reliability is influenced by the instrument itself, the composition and variability of the sample, sample size, and the objectivity of administration and scoring of the instrument. Sources of variability in the reliability of scores from a single instrument could include factors such as age, motivation, socioeconomic status, gender, and education. Group heterogeneity influences reliability. Essentially, anything that can affect scores can affect reliability. Therefore, because of the multiple influences on reliability, it is important for researchers to assess reliability for their own data. Score reliability is important to include in research reports as it directly affects both the results and the interpretation of the results of a particular study (e.g., attenuation of effect sizes; see Henson, 2001; Henson & Thompson, in press; Thompson, 1994; Wilkinson & American Psychological Association Task Force on Statistical Inference, 1999).

Vacha-Haase (1998) proposed a meta-analytic method, known as reliability generalization, to study score reliability. In her study of the Bem Sex Role Inventory, she found that approximately 13% of the 638 articles published during a 14-year period reported reliability for the data at hand. Study variables such as sample size, type of reliability coefficient (temporal stability vs. internal consistency), and test form accounted for variation in reliability coefficients across studies. Since then, other reliability generalization studies have investigated sources of variability in score reliability for instruments such as the Mathematics Anxiety Rating Scale (Capraro, Capraro, & Henson, 2001), the NEO personality scales (Caruso, 2000), and the Teacher Efficacy Scale (Henson, Kogan, & Vacha-Haase, 2001).

In this study we explored reliability generalization for Spielberger's State-Trait Anxiety Inventory (STAI) (Spielberger, Gorsuch, & Lushene, 1970). The STAI is a brief self-report assessment designed to measure and differentiate between anxiety as a trait and a state. Trait anxiety refers to individual differences in the frequency and intensity with which anxiety manifests itself over time. Trait anxiety consists of feelings of apprehension, tension, and increased activity of the autonomic nervous system, and is seen as a relatively stable personality trait (Spielberger, 1972). People who are high in trait anxiety tend to perceive more situations as threatening or dangerous than people who have lower trait anxiety scores, and persons with higher trait anxiety scores also tend to have higher state anxiety scores (Spielberger, 1972). State anxiety, on the other hand, fluctuates and is a function of the stressors on an individual. Levels of state anxiety are high in circumstances perceived by an individual to be threatening and irrespective of the objective danger. State anxiety should be low in nonstressful situations or in situations where an existing danger is not perceived as threatening.

The STAI state scale consists of 20 statements that ask people to describe how they feel at a particular moment in time (e.g., calm, tense) rated on a 4-

point intensity scale ranging from *not at all* to *very much so*. The STAI trait scale consists of 20 statements describing how people generally feel (e.g., confident) rated on a 4-point frequency scale ranging from *almost never* to *almost always*. The STAI has been translated into many languages, including Spanish, Turkish, Japanese, Arabic, and Dutch. The first version of the STAI is known as Form X and is frequently used in clinical research. Form Y was developed in 1983. Although the two versions are highly correlated, several items were changed and scores on Form Y are said to have a more replicable factor structure and improved psychometric properties (Oei, Evans, & Crook, 1990). An extension of the STAI for children (STAI-C) also has been developed.

In the 1970 STAI manual, the authors reported test-retest coefficients that were higher for scores on the trait scale (.84 for men, .76 for women) than the state scale (.33 for men, .16 for women). Internal consistencies for the state scale scores ranged from .83 to .92 for male and female high school and college students; for the trait scale scores, coefficients of internal consistency ranged from .86 to .92. For Form Y, median alpha coefficients were .90 and .93 for scores on the trait and state scales, respectively. Test-retest coefficients ranged from .73 to .86 and .16 to .62 for scores on the trait and state scales, respectively (Spielberger, 1983).

Method

Sample

A search of the PsycINFO database using the keywords *STAI* or *State-Trait Anxiety Inventory* (and spelling variants) yielded 454 journal citations in the English language spanning the years 1995 to 2000 and an additional 488 citations from 1990 to 1994. Of the initial 942 articles, 93 were not reviewed because they were not able to be located. (Article collection was carried out over a 9-month period.) Thirty-three articles were reviewed and excluded from this study because they either involved the children's version of the STAI ($n = 20$) or a short-form and/or otherwise modified version of the STAI ($n = 13$). One additional article was excluded because the traditional 4-point response format was changed to a dichotomous yes/no format. To be included in this study, articles needed to indicate that the original STAI, the 1983 STAI Form Y, or a foreign-language translation of one of these two forms was administered to participants. It was not necessary for both the state and trait scales to be administered—only that one of the scales was administered. A total of 816 articles met the inclusion criteria.

Each of these articles was read and classified by at least two of the authors. Articles were classified into one of three groups: (a) articles in which no mention of reliability was made ($n = 595$, 73%); (b) articles in which the authors

acknowledged reliability of the STAI scores, either through citing specific coefficients from other studies (typically normative studies), referencing other studies that had reported reliability although not providing specific coefficients, or stating (inappropriately) something to the effect that the reliability of the instrument was well-established without providing specific citations ($n = 175$, 21%); or (c) articles in which the authors reported reliability coefficients computed on the data analyzed in the article ($n = 46$, 6%).

Of the articles in which the authors provided reliability for their own data, for consistency one was excluded from analysis because the reliability was based on factor scores rather than conventional scoring. Several of the remaining articles that reported reliability coefficients for the data at hand provided more than one reliability coefficient (e.g., separate coefficients reported for subgroups of participants). Each subgroup coefficient was treated as an observation. A total of 117 reliability coefficients were obtained from 45 articles. Fifty-eight were trait coefficients (51 internal consistency, 7 test-retest) and 59 were state coefficients (52 internal consistency, 7 test-retest).

The 816 articles were further classified according to whether the journal in which they appeared was indexed in MedLine. This classification was conducted because during the review process it appeared to us that articles with a medical context—anecdotally identified by subject matter (e.g., drug comparisons in treatment of anxiety), authorship/institutional affiliation (e.g., M.D./hospital), and/or journal (e.g., *Neuropsychobiology*)—mentioned reliability less frequently than did articles in a nonmedical context. We thought this might be related to level of paradigm development among the subfields whose journals were indexed in PsycINFO. According to the literature on disciplinary differences (e.g., Biglan, 1973a, 1973b; for a review of the literature on disciplinary differences, see Braxton & Hargens, 1996), some disciplines—such as the physical and life sciences, engineering, and mathematics (i.e., the paradigmatic or hard science disciplines)—have a stronger codified body of knowledge and display a higher degree of consensus regarding their subject matter and methods of research. Operational definitions of key variables tend to be agreed on so, contrasted with disciplines having less-well-developed paradigms, there is less need to elaborate on research methods, including instrumentation.

In the so-called preparadigmatic (soft science) disciplines, which include the social/behavioral sciences and humanities, there may be less consensus so more journal space is devoted to describing, explaining, and justifying the objects and methods of study. Consequently, scholarly writings in the paradigmatic disciplines tend to be shorter than in the preparadigmatic disciplines (Biglan, 1973b; Creswell & Bean, 1981). Although psychology is typically classified as a preparadigmatic discipline, there is subfield variation (Braxton & Hargens, 1996; Lewis, 1980). Those subfields that publish in

medically oriented journals may share characteristics of the paradigmatic disciplines with respect to scholarly writings (Drees, 1982; Stoecker, 1991). It seemed reasonable to hypothesize that researchers writing in medical fields would be socialized to produce less verbiage and would be less accustomed to justifying or reestablishing the quality of their instrumentation than those in nonmedical fields; therefore, they may be less likely to report reliability for the data at hand.

MedLine, the basis for the above classification, is the National Library of Medicine's premier bibliographic database covering the fields of medicine, nursing, veterinary medicine, the health care system, and the preclinical sciences (U.S. National Library of Medicine, 2001). According to the U.S. National Library of Medicine, journals reviewed for inclusion in MedLine should contain articles predominantly on core biomedical subjects. This suggests that the journals indexed by PsycINFO and Medline may have more medical relevance than those indexed by PsycINFO alone. Of the articles, 532 (65%) were classified as medical; 284 (35%) were classified as nonmedical (listed in PsycINFO alone). Those from journals indexed in MedLine included psychiatric studies and studies in medical settings; journals not included in MedLine included articles of an educational and nonclinical nature. However, it is recognized that many journals indexed by MedLine are psychology oriented. Therefore, we would expect some overlap from our operational classification criteria. Nevertheless, the classification can provide a rough examination of possible differences in paradigms.

Procedures

Features of each study, sample characteristics, type of reliability coefficient, and the coefficients themselves were recorded. Study features and sample characteristics were selected based on the following criteria: (a) They were variables that reasonably might be expected to affect reliability or (b) they had previously been demonstrated to have an association with reliability, and (c) there were a sufficient number of studies that reported data on these variables. Table 1 lists the study features and respective codes.

Examples of contexts that were coded as high stimulus included surveying mothers of hospitalized children during hospitalization, HIV/AIDS patients completing neuropsychiatric assessments, clinical intakes of pain center patients, hospitalized children having veinipuncture, athletes' retrospective recall of competition-related anxiety, and students giving a speech in class. Low stimulus contexts generally involved completion of the STAI as part of a packet of instruments in a low-stress setting (e.g., classroom, mailed survey) with participants who were not specifically presumed to be experiencing stress or elevated anxiety levels (e.g., college students, high school students, and employees). We hypothesized that state anxiety would be more reliably

Table 1
Coded Study Features for Reliability Generalization

Feature/Characteristic	Coded Variable	Code
Form (new form, old form, translation, or unknown)	Form—new/other	1983 form = 1, else = 0
	Form—old/other	1970 form = 1, else = 0
	Form—translation/other	Translation = 1, else = 0
Context elicits anxiety	Context	1 = high stimulus, 0 = low
Medical or nonmedical study	MedContext	1 = medical, 0 = nonmedical
Design (experimental, psychometric, correlational)	Psychometric/other	1 = psychometric, 0 = other
	Experimental/other	1 = experimental, 0 = other
Participant age (kids, adults, seniors)	Sample—younger than 16/other	1 = younger than 16, 0 = 16 +
	Sample—older than 65/other	1 = 65+, 0 = younger than 65
Reliability type (internal consistency or test-retest)	Reliability type	1 = internal consistency, 0 = test-retest
Sample size	Sample size	Continuous values
STAI scale *M*s	State score *M*	Continuous values
	Trait score *M*	
STAI scale *SD*s	State score *SD*	Continuous values
	Trait score *SD*	
STAI scale score reliabilities	State coefficient	Continuous values
	Trait coefficient	

Note. STAI = State-Trait Anxiety Inventory.

measured in a high stimulus context because the construct would be more salient for those participants, thus increasing the signal-to-noise ratio. The reliability of trait anxiety, a more enduring personal characteristic, should be affected less by the context of the measurement. Therefore, trait test-retest coefficients should be higher than state test-retest coefficients. Mean state anxiety scores were higher in the high stimulus condition than in the low (47.62 vs. 36.56; Cohen's $d = 1.26$). The difference for trait scores, although in the same direction, was less (45.10 for high stimulus vs. 39.19 for low stimulus; Cohen's $d = .91$), thus providing evidence for the validity of this classification.

Results

Medical/Nonmedical Comparison

For this study, an average page length for articles from nonmedical journals was 11.43 ($SD = 4.71$) compared to 6.93 pages per article for medical

Table 2
Articles Classified by Journal Type and Reliability Reporting Status

	Reliability Ignored		Reliability Reported		Reliability Computed		
	n	%	*n*	%	*n*	%	Total
1990 to 1994							
Medical	196	74	60	23	10	4	266
Nonmedical	107	68	41	26	10	6	158
Total	303	71	101	24	20	5	424
1995 to 2000							
Medical	205	78	47	18	14	5	266
Nonmedical	87	69	27	21	12	10	126
Total	292	74	74	19	26	7	392

journals (SD = 2.75) for a random sample of 30 articles from each list. An independent groups t test showed this difference to be statistically significant at the .05 level, $t(58) = 4.54$ (Cohen's $d = 1.17$). This difference in page length is consistent with the literature on disciplinary differences in article length and thus provides some evidence for the validity of the procedure used to classify journals.

Table 2 shows the cross-classification of the 816 articles by journal type (medical or nonmedical) and reliability reporting status separately for the years 1990 to 1994 and 1995 to 2000. Authors of articles in the nonmedical journals increased their reporting of reliability for the data at hand from 6% to 10% during the two time periods and were somewhat more likely than authors in the medical journals, particularly in the latter years, to report reliability for the data at hand. Authors of articles in the medical journals were more likely than those in the nonmedical journals to ignore reliability altogether. The differences in the percentages, although small, were consistent and in the expected directions, suggesting that articles appearing in medical journals report reliability for the data at hand less often and ignore reliability more often than those appearing in nonmedical journals.

Descriptive Statistics for Reliability

It should be noted that some authors who reported reliability coefficients for the data at hand neglected to report basic descriptive statistics. Although 117 reliability coefficients were obtained, means and standard deviations of the scales were available for only 75 of the coefficients (38 trait and 37 state).

Table 3 and Figures 1 and 2 show similar average internal consistency coefficients for scores on both scales. The test-retest coefficients were noticeably lower for the state scale than for the trait scale, as displayed in Table 3

Table 3
Descriptive Statistics for Reliability Coefficients

	State		Trait	
	Internal Consistency (n = 52)	Test-Retest (n = 7)	Internal Consistency (n = 51)	Test-Retest (n = 7)
M	.91	.70	.89	.88
Median	.92	.68	.90	.88
SD	.05	.20	.05	.05
Minimum	.65	.34	.72	.82
Maximum	.96	.96	.96	.94

and Figures 3 and 4. This is in keeping with the theoretical model that distinguishes between an enduring trait and a temporary state, the latter being more susceptible to temporal fluctuation (Spielberger, 1972). Standard deviations among the coefficients were similar except for the test-retest coefficients on the state scale, which were considerably more varied. The histograms in Figures 1 through 4 show that the state scale had a greater number of low coefficients than did the trait scale; of the six reported coefficients below .70, four were state scale test-retest coefficients and two were state scale internal consistency coefficients.

Correlates of Internal Consistency Reliability

Because internal consistency and stability coefficients represent different sources of measurement error, their correlates may be different. Table 4 shows the correlations of study variables with internal consistency coefficients. Because there were only seven test-retest coefficients, their correlates were not analyzed. Studies with generally higher state score reliabilities tended to have somewhat higher state score standard deviations. These studies tended to have participants older than 16 years of age, to be psychometric in nature, and to have specified that they used Form Y (mean α = .92) rather than the old form, a translation, or an unspecified form (mean α = .89, Cohen's d = .60). Although the pattern of correlations was generally the same for trait score reliabilities, they were generally lower except for the correlation with the standard deviations of the scale scores. The form effect, in particular, seemed to be absent for the trait scale. For both scales, the study context (high or low anxiety stimulus) was not correlated with internal consistency reliability. The medical context of the journal in which the article appeared was not included in correlational analyses because the interest in

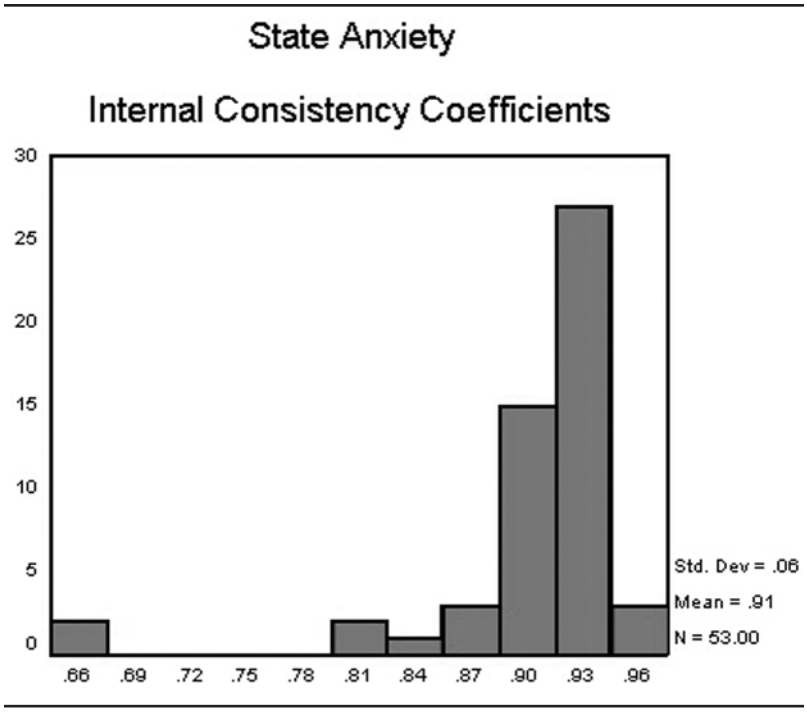

Figure 1. Distribution of state internal consistency reliabilities.

this variable was only its relationship to the incidence of reporting reliability, not its relationship to levels of reliability.

Discussion

For the STAI, internal consistency reliabilities were relatively stable across studies for both scales, especially the state anxiety scale. As expected with a characteristic that is state dependent, stability reliability was lower for scores on the state than the trait scale. State scale test-retest coefficients were lower than internal consistency coefficients, although for the trait scale the two types of coefficients were very similar. For the trait scale, in particular, test score variability was rather strongly related to the magnitude of the internal consistency reliability coefficient. This is certainly not a surprise, given that variability is a necessary although not sufficient condition for reliability. Neither is it a unique finding (e.g., Henson et al., 2001). That the internal consistency of the scores, particularly the state scores, was lower when the STAI scales were used with participants younger than the age of 16 suggests that

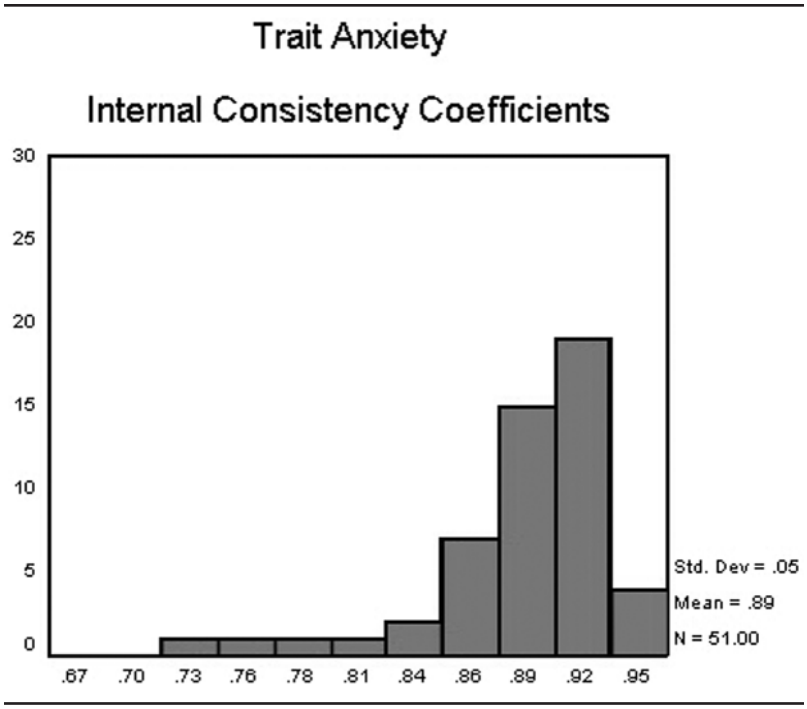

Figure 2. Distribution of trait internal consistency reliabilities.

STAI-C (Spielberger, 1973) perhaps should be considered, even for children older than elementary school (for a review of the STAI-C, see Walker & Kaufman, 1984). The use of Form Y resulted in somewhat higher internal consistency reliabilities than the use of some other form, reinforcing that researchers need to be very specific about which form of an instrument they use (six authors did not specify the form of the STAI used). The salience of the context for measuring state anxiety did not appear to affect the reliability of measurement, that is, state anxiety was not measured more reliably among participants in anxiety-provoking situations than in anxiety-neutral situations.

The results of this study suggest that internal consistency reliability estimates obtained from STAI state and trait scores, although somewhat variable, are generally satisfactory for a broad range of studies involving various populations. Researchers may anticipate that trait score test-retest reliabilities will be within acceptable ranges and that state score temporal stability will be lower. However, this good news should not lull researchers into complacency with regard to the reliability of scores in their own research. The correlates of reliability suggest a few things that researchers can do to enhance the internal consistency reliability of STAI scores in their own research with the STAI,

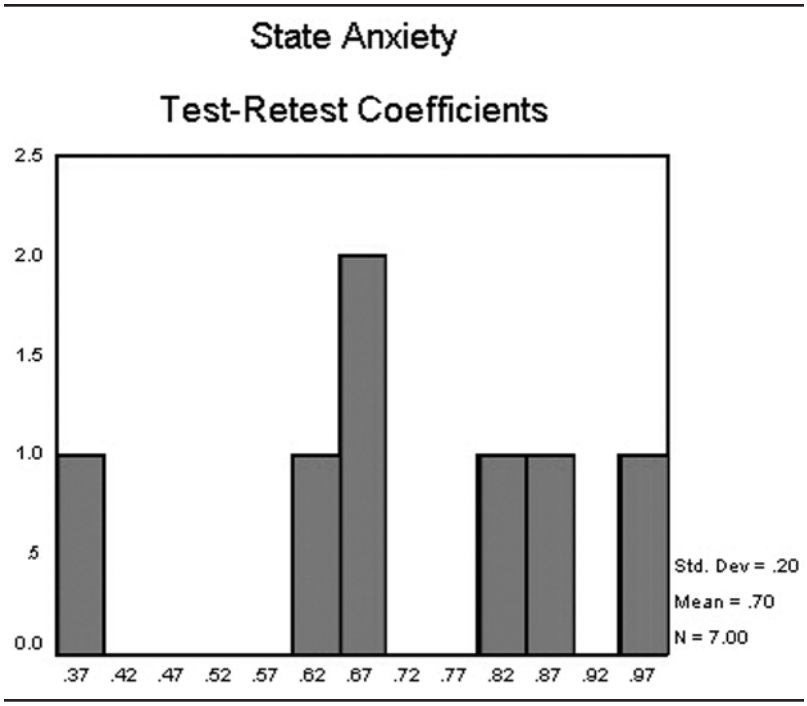

Figure 3. Distribution of state test-retest reliabilities.

and thus potentially increase effect sizes and statistical power. With regard to the form of the instrument, the use of an age-appropriate form is recommended, although the results do not suggest that Form Y is necessarily preferable to Form X. Because range restriction is associated with lower reliabilities, reasonable steps may be taken to increase the range of talent when it is feasible to do so. This may simply involve making wise choices about the target population (e.g., to investigate interventions for decreasing public speaking anxiety, one would probably not choose members of a winning debate team). If research is being done on selected groups (e.g., psychiatric patients admitted for extreme anxiety disorders), variability may be less and reliability lower. In the latter case, researchers should take this into consideration in their reporting and interpretation of results.

The comparison of reliability reporting rates for journals published in more medically oriented fields with those less medically oriented was consistent with our theory-based observation that the former were less likely to elaborate on characteristics of instrumentation, although the MedLine classification of journals was not always consistent with our intuitive sense of medical/nonmedical. However, the effect appeared to be strong enough to withstand the imperfection of our operational definition. It is apparent that

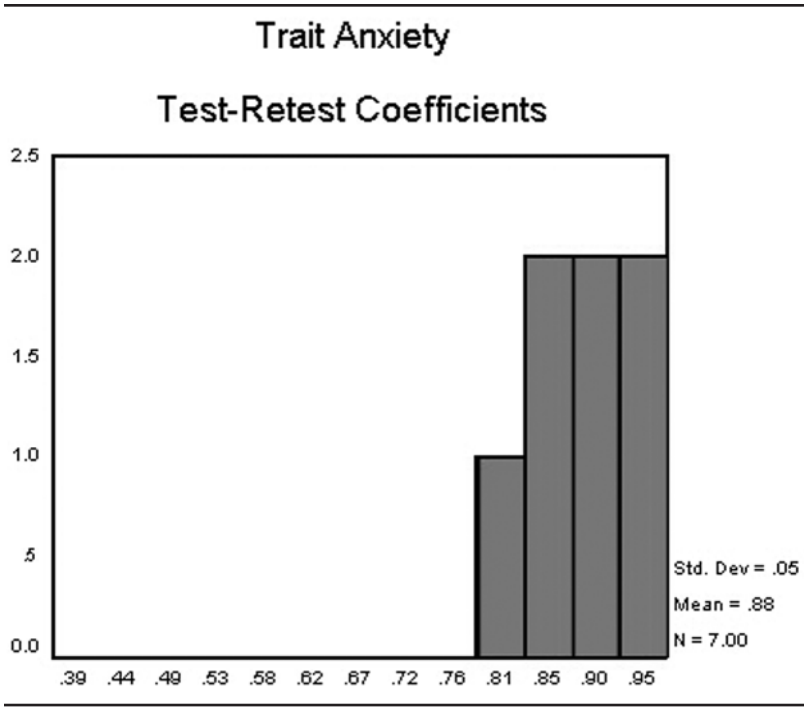

Figure 4. Distribution of trait test-retest reliabilities.

Table 4
Correlations Among Study Variables and Internal Consistency Coefficients

Variable	State	*n*	Trait	*n*
State score *M*	.15	31	—	—
State score *SD*	.22	31	—	—
Trait score *M*	—	—	.11	32
Trait score *SD*	—	—	.59	31
Sample size	.09	53	.10	51
Sample—younger than 16 years/other	–.31	53	–.21	51
Sample—older than 65 years/other	.04	53	.01	51
Study—psychometric/other	.24	47	.13	45
Study—experimental/other	.04	47	.03	45
Form—new/other	.26	53	.07	51
Form—old/other	–.09	53	–.11	51
Form—translation/other	–.18	53	–.00	51
Context	–.02	53	–.03	51

researchers in the social and behavioral sciences tend to lack awareness of the implications of score unreliability; this same ignorance may account for the dearth of reliability reporting in the MedLine-indexed journals. Added to this is the fact that the MedLine-indexed journal articles were, on average, 4.5 pages shorter; therefore, in these journals the reporting of psychometric considerations may be seen as unnecessary (perhaps assumed) detail. Vacha-Haase et al. (1999) noted that "journal author guidelines can be important for shaping practice and thinking" (p. 340) and argued for guidelines that clearly identify the need for reporting sample reliability coefficients.

References

**Arrindell, W. A., & Gerlsma, C. (1990). The validity of the μ index for differentiation of state and trait scales. *Psychological Reports*, *67*, 528-530.

**Bieling, P. J., Antony, M. M., & Swinson, R. P. (1998). The State-Trait Anxiety Inventory, trait version: Structure and content re-examined. *Behaviour, Research and Therapy*, *36*, 777-788.

Biglan, A. (1973a). The characteristics of subject matter in different academic areas. *Journal of Applied Psychology*, *57*, 195-203.

Biglan, A. (1973b). Relationship between subject matter characteristics and the structure and output of university departments. *Journal of Applied Psychology*, *57*, 204-213.

*Bouchard, S., Ivers, H., Gauthier, J. G., Pelletier, M.-H., & Savard, J. (1998). Psychometric properties of the French version of the State-Trait Anxiety Inventory (Form Y) adapted for older adults. *Canadian Journal on Aging*, *17*(4), 440-453.

Braxton, J. M., & Hargens, L. L. (1996). Variation among academic disciplines: Analytical frameworks and research. In J. C. Smart (Ed.), *Higher education: Handbook of theory and research* (Vol. 11, pp. 1-46). New York: Agathon Press.

Capraro, M. M., Capraro, R. M., & Henson, R. K. (2001). Measurement error of scores on the Mathematics Anxiety Rating Scale across studies. *Educational and Psychological Measurement*, *61*, 373-386.

**Carey, M. P., & Faulstich, M. E. (1994). Assessment of anxiety in adolescents: Concurrent and factorial validities of the Trait Anxiety scale of Spielberger's State-Trait Anxiety Inventory for Children. *Psychological Reports*, *75*, 331-338.

Caruso, J. C. (2000). Reliability generalization of the NEO personality scales. *Educational and Psychological Measurement*, *60*, 236-254.

**Creamer, M., Foran, J., & Bell, R. (1995). The Beck Anxiety Inventory in a non-clinical sample. *Behaviour, Research, and Therapy*, *33*(4), 477-485.

Creswell, J. W., & Bean, J. P. (1981). Research output, socialization, and the Biglan model. *Research in Higher Education*, *15*, 69-92.

Drees, L. A. (1982). *The Biglan model: An augmentation*. Unpublished doctoral dissertation, The University of Nebraska–Lincoln.

*Dunbar, E. (1993). The relationship of subjective distress and emergency response experience to the effective use of protective equipment. *Work & Stress*, *7*(4), 365-373.

**Fogel, C. I., & Martin, S. L. (1992). The mental health of incarcerated women. *Western Journal of Nursing Research*, *14*(1), 30-47.

*Gidycz, C. A., & Koss, M. P. (1990). A comparison of group and individual sexual assault victims. *Psychology of Women Quarterly*, *14*, 325-342.

*Gloria, A. M., & Hird, J. S. (1999). Influences of ethnic and nonethnic variables on the career decision-making self-efficacy of college students. *Career Development Quarterly*, *48*, 157-174.

**Goertzel, L., & Goertzel, T. (1991). Health locus of control, self-concept, and anxiety in pediatric cancer patients. *Psychological Reports*, *68*, 531-540.

*Hankin, B. L., Roberts, J., & Gotlib, I. H. (1997). Elevated self-standards and emotional distress during adolescence: Emotional specificity and gender differences. *Cognitive Therapy and Research*, *21*(6), 663-679.

Henson, R. K. (2001). Understanding internal consistency reliability estimates: A conceptual primer on coefficient alpha. *Measurement and Evaluation in Counseling and Development*, *34*, 177-189.

Henson, R. K., Kogan, L. R., & Vacha-Haase, T. (2001). A reliability generalization study of the Teacher Efficacy Scale and related instruments. *Educational and Psychological Measurement*, *61*, 404-420.

Henson, R. K., & Thompson, B. (in press). Characterizing measurement error in scores across studies: Some recommendations for conducting "Reliability Generalization" (RG) studies. *Measurement and Evaluation in Counseling and Development*.

**Hishinumna, E. S., et al. (2000). Psychometric properties of the State-Trait Anxiety Inventory for Asian/Pacific-Islander adolescents. *Assessment*, *7*(1), 17-36.

**Kabacoff, R. I., Segal, D. L., Hersen, M., & Van Hasselt, V. B. (1997). Psychometric properties and diagnostic utility of the Beck Anxiety Inventory and the State-Trait Anxiety Inventory with older adult psychiatric outpatients. *Journal of Anxiety Disorders*, *11*(1), 33-47.

*Kavussanu, M., & McAuley, E. (1995). Exercise and optimism: Are highly active individuals more optimistic? *Journal of Sport & Exercise Psychology*, *17*, 246-258.

**Kohn, P. M., & Gurevich, M. (1993). On the adequacy of the indirect method of measuring the primary appraisal of hassles-based stress. *Personality and Individual Differences*, *14*, 679-684.

Lewis, G. L. (1980). The relationship of conceptual development to consensus: An exploratory analysis of three subfields. *Social Studies of Science*, *10*, 285-308.

**MacFarlane, M. E., & Sony, S. D. (1992). Women, breast lump discovery, and associated stress. *Health Care for Women International*, *13*, 23-32.

**Marteau, T. M., & Bekker, H. (1992). The development of a six-item short-form of the state scale of the Spielberger State-Trait Anxiety Inventory (STAI). *British Journal of Clinical Psychology*, *31*, 301-306.

*McCaffrey, R. J., et al. (1995). Practice effects with the NIMH AIDS Abbreviated Neuropsychological Battery. *Archives of Clinical Neuropsychology*, *10*, 241-250.

*Melnyk, B. M. (1995). Coping with unplanned childhood hospitalization: The mediating functions of parental beliefs. *Journal of Pediatric Psychology*, *20*, 299-312.

*Miles, M. S., Funk, S. G., & Kasper, M. A. (1992). The stress response of mothers and fathers of preterm infants. *Research in Nursing & Health*, *15*, 261-269.

*Miller, F., & Varma, N. (1994). The effects of psychosocial factors on Indian children's attitudes toward computers. *Journal of Educational Computing Research*, *10*, 223-238.

*Muris, P., Mayer, B., & Merckelbach, H. (1998). Trait anxiety as a predictor of behaviour therapy outcome in spider phobia. *Behavioural and Cognitive Psychotherapy*, *26*, 89-91.

*Novy, D. M., Nelson, D. V., Smith, K. G., Rogers, P. A., & Rowzee, R. D. (1995). Psychometric comparability of the English- and Spanish-language versions of the State-Trait Anxiety Inventory. *Hispanic Journal of Behavioral Sciences*, *17*, 209-224.

Oei, T., Evans, L., & Crook, G. (1990). Utility and validity of the STAI with anxiety disorder patients. *British Journal of Clinical Psychology*, *29*, 429-432.

*Orbach, I., Shopen-Kofman, R., & Mikulincer, M. (1994). The impact of subliminal symbiotic vs. identification messages in reducing anxiety. *Journal of Research in Personality*, *28*, 492-504.

**Patterson, I. (1996). Participation in leisure activities by older adults after a stressful life event: The loss of a spouse. *International Journal of Aging and Human Development*, *42*(2), 123-142.

**Pond, E. F., & Kemp, V. H. (1992). A comparison between adolescent and adult women on prenatal anxiety and self-confidence. *Maternal-Child Nursing Journal*, *20*(1), 11-20.

*Reed, M. A., & Derryberry, D. (1995). Temperament and response processing: Facilitatory and inhibitory consequences of positive and negative motivational stress. *Journal of Research in Personality*, *29*, 59-84.

*Renkl, A. (1995). Learning for later teaching: An exploration of mediational links between teaching expectancy and learning results. *Learning and Instruction*, *5*, 21-36.

**Rocha-Singh, I. (1994). Perceived stress among graduate students: Development and validation of the graduate stress inventory. *Psychological Reports*, *54*, 714-727.

*Roy, B. D., & Roy, D. D. (1994). Mathematics preference, anxiety and achievement in mathematics. *Psychological Studies*, *39*(1), 34-36.

**Salminen, S., Liukkonen, J., Hanin, Y., & Hyvonon, A. (1995). *Personality and Individual Differences*, *19*, 725-729.

*Sawyer, C. R., & Behnke, R. R. (1999). State anxiety patterns for public speaking and the behavior inhibition system. *Communication Reports*, *12*(1), 33-41.

*Schaubroeck, J., Ganster, D. C., & Kemmerer, B. (1996). Does trait affect promote job attitude stability? *Journal of Organizational Behavior*, *17*, 191-196.

**Schisler, T., Lander, J., & Fowler-Kerry, S. (1998). Assessing children's state anxiety. *Journal of Pain and Symptom Management*, *16*(2), 80-86.

**Schotte, C.K.W., Maes, M., Cluydts, R., & Cosyns, P. (1996). Effects of affective-semantic mode of item presentation in balanced self-report scales: Biased construct validity of the Zung Self-Rating Depression Scale. *Psychological Medicine*, *26*, 1161-1168.

**Scott, B., & Melin, L. (1998). Psychometric properties and standardised data for questionnaires measuring negative affect, dispositional style and daily hassles: A nation-wide sample. *Scandinavian Journal of Psychology*, *39*, 301-307.

Spielberger, C. (1972). *Anxiety: Current trends in research*. London: Academic Press.

Spielberger, C. (1973). *STAIC preliminary manual*. Palo Alto, CA: Consulting Psychologists Press.

Spielberger, C. D. (1983). *Manual for the State-Trait Anxiety Inventory.* Palo Alto, CA: Consulting Psychologists Press.

Spielberger, C. D., Gorsuch, R. L., & Lushene, R. E. (1970). *Manual for the State-Trait Anxiety Inventory.* Palo Alto, CA: Consulting Psychologists Press.

**Stanlay, M. A., Beck, J. G., & Zebb, B. J. (1996). Psychometric properties of four anxiety measures in older adults. *Behaviour Research and Therapy*, *34*(10), 827-838.

Stoecker, J. L. (1991, April). *The Biglan classification revisited*. Paper presented at the annual meeting of the American Educational Research Association, Chicago.

**Subkoviak, M. J., et al. (1995). Measuring interpersonal forgiveness in late adolescence and middle adulthood. *Journal of Adolescence*, *18*, 641-655.

**Tangney, J. P., Wagner, P., & Gramzow, R. (1992). Proneness to shame, proneness to guilt, and psychopathology. *Journal of Abnormal Psychology*, *101*(3), 469-478.

Thompson, B. (1994). Guidelines for authors. *Educational and Psychological Measurement*, *54*, 837-847.

U.S. National Library of Medicine. (2001). *Fact sheet: Journal selection for Index Medicus/ MEDLINE*. Retrieved from www.nlm.nih.gov/pubs/factsheets/jsel.html

Vacha-Haase, T. (1998). Reliability generalization: Exploring variance in measurement error affecting score reliability across studies. *Educational and Psychological Measurement*, *58*, 6-20.

Vacha-Haase, T., Kogan, L. R., & Thompson, B. (2000). Sample compositions and variabilities in published studies versus those in test manuals: Validity of score reliability inductions. *Educational and Psychological Measurement*, *60*, 509-522.

Vacha-Haase, T., Ness, C., Nilsson, J., & Reetz, D. (1999). Practices regarding reporting of reliability coefficients: A review of three journals. *Journal of Experimental Education*, *67*, 335-341.

*Vealey, R. S., Udry, E. M., Zimmerman, V., & Soliday, J. (1992). Intrapersonal and situational predictors of coaching burnout. *Journal of Sport & Exercise Psychology*, *14*, 40-58.

**Vernon, S. W., et al. (1997). Correlates of psychologic distress in colorectal cancer patients undergoing genetic testing for hereditary colon cancer. *Health Psychology*, *16*, 73-86.

**Virella, B., Arbona, C., & Novy, D. M. (1994). Psychometric properties and factor structure of the Spanish version of the State-Trait Anxiety Inventory. *Journal of Personality Assessment*, *63*, 401-412.

Walker, C. E., & Kaufman, K. (1984). State-Trait Anxiety Inventory for Children. In D. J. Keyser & R. C. Sweetland (Eds.), *Test critiques* (Vol. 1, pp. 633-640). Kansas City, MO: Westport.

Wilkinson, L., & American Psychological Association Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, *54*, 594-604. (Reprint available through the APA home page: http://www.apa.org/journals/amp/amp548594.html)

**Youngstrom, E., Izard, C., & Ackerman, B. (1999). Dysphoria-related bias in maternal ratings of children. *Journal of Consulting and Clinical Psychology*, *67*(6), 905-916.

*Zoller, U., & Ben-Chaim, D. (1990). Gender differences in examination-type preferences, test anxiety, and academic achievements in college science education—A case study. *Science Education*, *74*(6), 597-608.

---

*Journals not indexed in MedLine as of September 2000.

**Journals indexed in MedLine as of September 2000.