Educational Gerontology, 32: 409–421, 2006

ISSN: 0360-1277 print/1521-0472 online
DOI: 10.1080/03601270600685628

Routledge
Taylor & Francis Group

RELIABILITY AND VALIDITY ISSUES FOR TWO COMMON MEASURES OF MEDICAL STUDENTS' ATTITUDES TOWARD OLDER ADULTS

T. J. Stewart
E. Roberts
P. Eleazer
R. Boland
D. Wieland

School of Medicine, University of South Carolina, Columbia, South Carolina, USA

Results are reported from 2 common measures of medical student attitudes toward older adults: Maxwell-Sullivan Attitude Survey (MSAS); and UCLA Geriatrics Attitude Survey (GAS), with students entering the University of South Carolina School of Medicine (USCSM) in the period 2000–2005. A reliability analysis incorporating item means, Cronbach's alpha, item correlation matrix, and, Spearman-Brown prediction for positively and negatively worded items was conducted. Internal consistency results were unacceptable, revealing reliability and validity problems in this sample of medical students. Reconsideration of the use of these common measures, and a reframing of attitudes of medical students toward older adults seem appropriate.

The attitudes of physicians, residents and medical students toward older adults has been a consistent theme in geriatrics education. Research on the topic of medical students' attitudes toward geriatrics patients or older adults in general has been underway for 40 plus years with more than 60 cited publications (dating to the early

Funding sources: This work was supported by grants from the John A Hartford Foundation (through the American Association of Medical Colleges), the Donald W. Reynolds Foundation, and institutional funding from the University of South Carolina. This research was approved by the University of South Carolina School of Medicine Institutional Review Board.

Address correspondence to Thomas J. Stewart, Division of Geriatrics, School of Medicine, University of South Carolina, 15 Medical Park, Columbia, SC 29203. E-mail: thomas.stewart@palmettohealth.org

1960s; Parker, 1960; Ford, Liske, & Ort, 1962). One of the earliest studies focusing specifically on medical students referred to student attitudes in such strong terms as "hostility" toward older adults, and the authors interpreted their findings about student ageist attitudes to be more pronounced than racism (Spence, Feigenbaum, Fitzgerald, & Roth, 1968). The authors suggested that the attitudes of physicians (residents and students as well) not only reflected the negative attitudes of society, but also the combined impact of training with frequent exposure to frail older patients who had chronic and advanced illnesses and exposure to physician role models with poor attitudes. This type of attitude study became common in geriatrics education in American medical schools during the latter half of the 20th century and into the 21st century.

Most of the published research in this area was grounded in curriculum applications, such as measuring student attitudes prior to and following "geriatric interventions" in courses and clinical rotations/clerkships. Typically curriculum activities incorporated attitude objectives (e.g., measuring and positively affecting student attitudes toward older adults in general, or geriatrics patients specifically). These studies assumed that positive attitudes as a student might influence or predict positive attitudes and behavior toward older patients in clinical practice. In some cases, attitudes were explored to determine the extent to which positive attitudes might predispose students toward selection of geriatrics as a clinical specialty.

In a 1984 review Coccaro and Miles described curriculum-based interventions to change student attitudes about older adults as a "mixed success" (Coccaro & Miles, 1984). A subsequent review in 1987 indicated that it was difficult to compare and evaluate published studies to that historical point because of variability in study design, attitude measures, and types of curriculum interventions (Adelman & Albert, 1987). Although results from studies conducted since these reviews also reported findings without distinct patterns, the imperative to continue to chart or monitor student attitudes has continued.

As geriatric components in the medical school curriculum evolved at the University of South Carolina over the past 15 years, it seemed a natural step to incorporate attitude measures as part of the evaluation of geriatric learning opportunities. It was also natural to participate in collaborative efforts among medical schools with major geriatrics initiatives underway to develop useful measures. In the late 1990s, it was decided to use three measures in an attempt to determine baseline attitudes and chart attitude change over time. The three measures selected were: the Rosencranz and McNevin Aging Semantic Differential (Rosencranz & McNevin, 1969); the Maxwell-Sullivan

Attitude Survey on Aging (Maxwell & Sullivan, 1980); and the UCLA Geriatric Attitude Survey (Reuben et al., 1998).

This paper reports the reliability analysis of two of the measures: the Maxwell-Sullivan Attitude Survey (MSAS), and the UCLA Geriatrics Attitude Survey (GAS). The analysis was conducted to determine the extent to which the measures can be used internally for evaluation and other purposes. The reliability findings for University of South Carolina School of Medicine (USCSM) students are compared with those of similar studies of medical students and residents. Finally, validity issues for the two measures are discussed.

METHODS

The Measures

The Maxwell-Sullivan Attitude Survey (MSAS) was designed for use with family medicine residents, however, it has been applied with medical students as well. It is a 28-item survey which produces scores for five scales: general attitudes about the elderly; cost effectiveness of treating the elderly; time and energy spent in treating the elderly; therapeutic potential of the group; and, physician educational preparation for work with geriatrics patients. Responses are scored using a 5-point Likert scale for each item with a scoring scheme in which lower scores are interpreted as "positive" scores and higher ones as "negative" scores. This measure was adapted for first year USCSM students by dropping the educational preparation scale (6 items) and a single item (item number 21 in the original MSAS) of the therapeutic potential scale, resulting in a 21-item measure. The educational-preparation scale, which focused on the extent to which one's medical education prepared an individual for clinical work with elderly people, was deemed inappropriate for entering medical students. The single item dropped from the therapeutic-potential scale was judged to be redundant to another item in the original scale.

The UCLA Geriatrics Attitude Survey (GAS) is a 14-item survey, also using Likert-scale responses indicating the extent to which the respondent disagrees or agrees with each item. It was first developed with medical residents and later applied to medical students. The GAS uses a 5-point scale with low scores (1–2) indicating negative views, three as the neutral response and 4–5 interpreted as "positive" scores. The GAS was not modified. Both the MSAS and the GAS utilize positively and negatively phrased survey items, with the latter reverse scored to produce a consistent scoring framework.

The Sample

USCSM students from two entering cohorts completed the MSAS during the early months of academic years 2000–2001 and 2001–2002 ($N = 143$ complete forms). Students entering medical school for the academic years 2002–2003 through 2004–2005 completed the GAS during the first geriatrics experience ($N = 196$ complete forms). The measures are administered at other points for each student cohort, but this report uses only data from entry into medical school.

Analytic Procedures

Analytic procedures included calculation of means and standard deviations for selected measures, items and sets of items within those measures; calculation of Cronbach's coefficient alpha for reliability analysis; calculation of a correlation and covariance matrix to consider the strength of relationships between items in measures; and, calculation of Spearman-Brown prediction formula for sets of positively and negatively worded items within measures. Analyses were conducted using the statistical package for the social sciences 12.0 (2004). As a criterion for judging reliability coefficients we applied the range recommended by Streiner and Norman, that is, Cronbach alpha values of .70 as a minimum and .90 as a maximum (Streiner & Norman, 2003). The lower end of the range is based upon the historical cut-point of .70 for "acceptable" levels of Cronbach's alpha values (Nunnally, 1978). That range is also suggested in attitude measurement (DeVellis, 2003) and medical education (Downing, 2004) literature.

RESULTS

Reliability Analysis

The Maxwell Sullivan Attitude Survey (MSAS)
Although the full scale score reliability of the MSAS reported originally by Maxwell and Sullivan (Table 1) was above the minimally acceptable threshold, the internal scale scores were unacceptable by the standards adopted (Maxwell & Sullivan, 1980). The same pattern was evident for the USCSM student scores. One of the scale alpha results was so poor in the USCSM analysis ("cost effectiveness" at .222) that it indicated not only the particular scale was unreliable, but also caused concern about the full measure. The scale performance

Table 1. Maxwell-Sullivan scale reliability using cronbach's alpha for selected studies

Scales	Univ. of South Carolina[1]	Maxwell- Sullivan original[2]	Univ. of Michigan[3]	UCLA[4]
General attitudes	.608	.632	.560	NA
Cost effectiveness	.222	.496	NA	NA
Time and energy	.519	.609	NA	NA
Therapeutic potential	.631	.620	NA	NA
Full measure	.759	.739	NA	.540

[1]Data from 143 students entering University of South Carolina School of Medicine, in two academic years and using 21 items of original 28 item measure.
[2]Maxwell & Sullivan, 1980.
[3]Fitzgerald et al., 2003.
[4]Reuben et al., 1998.
NA: Results for these scales within the measure not provided.

overall was so poor it seemed best to consider that alpha score sensitivity to the number of items included (21 in this case) was a prevailing factor (Schmitt, 1996); that is, the "acceptable" alpha value was a function of the number of items in the measure. As Streiner has pointed out, measures with over 14 or more items tend to show "acceptable" values of coefficient alpha, even under circumstances where the measure taps into two or more different constructs (Streiner, 1996). Additionally, the mean interitem correlation for the USCSM sample was .149 (range of $-.261$ to .593). The correlation matrix for item relationships exhibited a relatively high proportion of zero-order correlations (about 25%). Taken together, these results left us with no confidence in either the scales or the full measure. Reuben et al. commented earlier on poor reliability and validity results for the MSAS based on attempts at UCLA to apply the MSAS as an attitude measure (Reuben et al., 1998). Other medical schools which have reported work with the MSAS have been more interested in the internal scales than in the full measure (Fitzgerald, Wray, Halter, Williams, & Supiano, 2003; Boal et al., 2001). The experience reported here should prompt caution and reservations about use of the MSAS, either as a full measure or its respective scales.

UCLA Geriatrics Attitude Scale (GAS)

The published reliability scores (Cronbach's alpha) for the UCLA GAS at sites away from UCLA have been marginal. Table 2 indicates alpha values for scores from the original use of the GAS, and those of three medical schools that have applied it. Although the original analysis from the development of the instrument (Reuben et al.,

Table 2. Selected findings for the UCLA geriatrics attitude survey (GAS): UCLA and universities of Hawaii, Michigan, and South Carolina

Institution	Population surveyed (*N*)	Means (*SD*)	Cronbach's alpha
UCLA (Reuben et al., 1998)	Residents, fellows & faculty (121)	3.71 (.47)	.760
Univ. of Michigan (Fitzgerald et al., 2003)	Entering medical students, 2001 (174)	3.70 (NA)	.690
Univ. of South Carolina	Entering medical students, 2002–2004 (196)	3.70 (.410)	.697
Univ. of Hawaii (Kishimoto et al., 2005)	Medical students, residents and fellows (211)	3.77 (NA)	.690

1998) reported an acceptable level (.76), the other three projects reported alpha values either at or below the lower threshold. The GAS score alpha value for the USCSM students (.697), coupled with a very poor interitem correlation picture, led us into a more detailed analysis. The mean interitem correlation for all 14 items was .137 with a range of $-.141$ to .467. The appearance of numerous negative correlations prompted a review of the scoring and data entry process to assure that scores had been entered and reverse scored as prescribed. Indeed, no mistakes were made—the data were accurate and not reassuring. In the interitem correlation matrix, approximately 1/3 of the correlations fell into what we interpreted as zero-order correlations (in the range of $-.09$ to $+.09$), indicating that the internal consistency of the GAS with our students was problematic: many of the items seemed to be unrelated to each other.

Positive and Negative Items in the MSAS and GAS

An added methodological concern for both measures was the mixed use of positive and negative items. During the past 40 or so years, it has become common practice to employ a balanced set of positive and negative items in the design of attitude surveys (refer to Nunnally, 1978; Anastasi, 1982; Anderson, Basilevsky, & Hum, 1983). This recommendation grew from the perceived need to deal with response sets on the part of individuals taking attitude measures—response sets such as acquiescence or serial agreement (e.g., consistently marking the "agree" response on the typical Likert scale). The use of positive and negative items interspersed throughout an attitude measure was a strategy to reduce the likelihood that respondents would fall into a patterned response. The negative and positive items were assumed

to be essentially equivalent, and the negative items were reverse-scored to establish a congruent scoring scale. Indeed, the team which developed the UCLA GAS began with 40 items with "about half the statements worded positively and half negatively." But, eventually, the research team adopted an unbalanced set of 14 items (5 positively worded and 9 negatively worded) (Reuben et al., 1998, p. 1426) after analysis of pretesting results.

A growing body of literature has focused upon problems stemming from the mixed use of positively- and negatively-phrased items in attitude measures. The research has questioned the assumption that people respond as expected to negatively-phrased survey items. It demonstrated that the use of mixed positive and negative items usually results in differences between the means of the positive and negative item. That is, if the positive and negative items are viewed as sets, then the means of those sets differ, and the reliability estimates are affected (usually suppressed). It has also been noted that in factor analysis the negative items loaded first on a single factor (Pilotte & Gable, 1990; Schriesheim, Eisenbach, & Hill, 1991). Experimental studies comparing three forms of a given measure (an all positive items version; an all negative items version; and a mixed negative and positive item version) found that the form using a consistent set of all positive or all negative items produced more reliable and accurate results than the mixed form (Schriesheim & Hill, 1981; Benson & Hocevar, 1985). The evidence is growing that respondents seem to process positive and negative items differently. It is now suggested that the practice of using mixed negative and positive items to deal with response sets creates methodological problems, and the practice needs to be re-evaluated (Barnette, 2000; Stewart & Frye, 2004).

Both the MSAS and the GAS used mixed positive and negative attitude items, but neither incorporated a balanced set as recommended in attitude methods texts (12 positive and 16 negative items in the MSAS; and 5 positive and 9 negative items in the GAS). Analysis of the USC data revealed signs of these issues in both the MSAS and the UCLA GAS. Table 3 indicates that the means of the positive and negative items when clustered as sets have mean differences which are statistically significant (.001 level). We viewed this as a "flag" about the assumption that students responded to the positive and negative forms in the same manner—simply comparing the positive and negative items as sets we detected a likely pattern where none should be expected. Further analysis for the UCLA GAS and the MSAS exhibited another sign of the negative/positive items phenomenon, that is, the reliability coefficient (Cronbach's alpha) was affected by the mixed

Table 3. Positive and negative item set comparisons for the maxwell-sullivan attitude survey (MSAS) and UCLA geriatrics attitude survey (GAS), 143 entering students, university of South Carolina school of medicine (means of sets, mean difference between sets and projected coefficient alpha values for scales constructed of all positive and all negative items)

Measure	Positive & negative item sets	Mean (SD)	Significance level	95% confidence intervals	
				Lower	Upper
MSAS	MSAS + items	2.111 (.350)	.001	2.055	2.171
	MSAS − items	2.281 (.418)	.001	2.212	2.350
GAS	GAS + items	3.820 (.468)	.001	3.754	3.886
	GAS − items	3.637 (.478)	.001	3.569	3.704

use of positive and negative items. The Spearman-Brown prediction formula provides a projected coefficient alpha that serves as an estimate for a scale constructed of either all positive or all negative items (the Spearman-Brown formula provided in Spector, 1991). As noted in Table 4, using all negatively phrased items for the MSAS increased the reliability score estimate, while using all positively phrased items decreased it. In the GAS data we found that either all positively phrased or all negatively phrased items "improved" its coefficient alpha score, with greater improvement from an all-negative measure (increasing the alpha value from .697 to .788). Better reliability estimates were produced by "modified" versions of both measures. The results from the means and the coefficient alpha projections led us to conclude that the use of combined positive and negative items in these measures created a method or measurement effect that further reduced the confidence we had in any interpretation of findings.

Table 4. Positive and negative item set comparisons for the maxwell-sullivan attitude survey (MSAS) and UCLA geriatrics attitude survey (GAS): Projected coefficient alpha values for scales constructed of all positive and all negative items applying the spearman-brown prediction formula)

Measure	Original scale alpha value	Projected alpha: all positive items	Projected alpha: all negative items
MSAS	.759	.690	.853
GAS	.697	.719	.788

IMPLICATIONS

Methodological

The findings of this application of two common measures of student attitudes toward older adults highlight serious internal consistency issues for both measures in this sample of students. These findings, when placed alongside the added issue of a likely measurement effect from the mixing of positively and negatively phrased items, point to serious weaknesses. Furthermore, the findings of the use of the two measures at other medical schools have not presented a more favorable picture. Consequently, it was necessary to make an internal decision to drop the use of both measures for evaluation and other purposes. Yet, at the present time, we have not found other measures that are superior for the purpose intended; i.e., measurement of attitudes towards older adults. Additionally, it seems appropriate to echo the growing cautions to potential users of these measures and/or similar attitude measures using mixed negative and positive items: that is, there is growing evidence that this practice needs to be reconsidered and subjected to further research.

Validity Issues

It was not our intent to "validate" these measures, but it seems appropriate to comment on implications for validity. The reliability picture for the MSAS and UCLA GAS with the USCSM students is problematic when we consider the issue of validity. Traditionally, in psychometric evaluation of measures, "good" reliability is a sine qua non for judging validity (Guilford, 1954). Some researchers in education and behavioral sciences have questioned an uncritical attachment to the principle that a measure cannot be valid without first being reliable (Moss, 1994). In the case of student attitudes toward older adults, unfortunately, we have little alternative in determining validity of a measure without reference to reliability of the instrument. There are limitations to approaching validity through the traditional tripartite categories of validity (content, criterion, and construct validity).

Content validity usually focuses upon the adequacy of the selected items in representing the domain of interest (here student attitudes toward older adults). Unfortunately, this domain is poorly established and not well enough defined to determine if the MSAS or UCLA GAS match with what might be considered representative items and related concepts. To a certain extent, face validity may

be an issue with the measures. For instance, consider an item in the GAS, "I tend to pay more attention and have more sympathy towards my elderly patients than my younger patients." On the face of it, a student might view the question from the perspective of either attitudes toward the elderly or attitudes about professionalism (that is, having biased attitudes toward one class of patients over another as opposed to treating all patients equally). This item produced the lowest mean response of all items.

To be effective, criterion validity relies upon comparisons and performance of a measure against gold standards. Again, the formative state of this area of enquiry makes comparisons difficult because such standards require a history of research to establish relationships and criteria. For instance, is enough known about the attitudes of fellows and faculty members in divisions and departments of geriatrics to use the scores of these groups as a "standard"? Probably not! Are there conceptual issues to be resolved in using these groups as standards? Definitely!

Construct validity frames validity by exploring theoretical relationships. The history of medical student attitudes toward older adults arose in the broader exploration of ageist attitudes and stereotypes of the elderly. This accounts for the fact that the Aging Semantic Differential (ASD), as a general population measure of ageism attitudes, has been used in medical schools. However, there has not been a well-articulated theoretical framework for these measures that would link the attitude domains that are measured with general population attitudes towards older adults. Furthermore, we have not found a well-articulated theoretical basis that would show that the domains that are measured by these scales are more or less important when compared with other attitudes thought to be important in modern medical practice (such as empathy, respect for patients and others, patient-centeredness, etc.). For example, are MSAS and GAS scores related to empathy scores? Are positive scores on either instrument associated with humanistic qualities or patient-centered attitudes? We have no framework to hypothesize relationships between student characteristics/behaviors and stereotypical attitudes toward the elderly. The options and circumstances for exploring validity independent of reliability are limited.

Perhaps it is time for researchers in geriatrics education to question whether the measurement of attitudes towards older adults is an appropriate surrogate measure for student and doctor behavior toward older adults. For example, current accreditation guidelines for medical schools emphasize curriculum objectives and competencies that foster the development of physicians who are altruistic,

compassionate, empathic, nonjudgmental, and honest (Association of American Medical Colleges, Medical School Objectives Project, Report 1, 1998). It may well be that these general characteristics—rather than attitudes about aging per se—are of greater value in predicting which students will provide compassionate care to older adults. If that is the case, then specific measures focusing on students' attitudes toward older adults seem less compelling or necessary.

We question the original premise that student attitudes toward older adults are a surrogate measure for subsequent student behavior. Specifically, it is possible for students to have a positive attitude and not render kind, compassionate, and competent care to older adults. The reverse is also true—students may have less positive or negative attitudes, yet behave and care for older adults in a kind, compassionate, and competent fashion. As one colleague commented, "I don't really care if they have a positive attitude as long as they are competent and behave in a professional, kind manner to older adults." Perhaps geriatrics education should focus less on surrogate attitudinal measures that may or may not reflect behavior. It could focus instead on the desired educational outcome: kind, compassionate, and competent care. Students who consistently develop this habit will have the "virtue" of an approach to caring for older people that has long been a major goal in geriatric education.

REFERENCES

Adelman, R. D. & Albert, R. C. (1987). Medical student attitudes toward the elderly: A critical review of the elderly. ***Gerontology and Geriatrics Education****,* ***7****,* **141–155.**

Anastasi, A. (1982). ***Psychological testing*** **(5th ed.). New York: Macmillan.**

Anderson, A. B., Basilevsky, A., & Hum, D. P. J. (1983). Measurement: Theory and techniques. In P. H. Rossi, J. D. Wright, & A. B. Anderson (Eds.), ***Handbook of survey research****. New York: Academic Press.**

Association of American Medical Colleges, Medical School Objectives Writing Group. (1998). Learning objectives for medical student education—guidelines for medical schools. ***Academic Medicine****,* ***74****, 13–18.**

Barnette, J. J. (2000). Effects of stem and Likert response option reversals on survey internal consistency: If you feel the need, there is a better alternative to using those negatively worded stems. ***Educational and Psychological Measurement****,* ***60****(3), 361–370.**

Benson, J. & Hocevar, D. (1985). The impact of item phrasing on the validity of attitude scales for elementary school children. ***Journal of Educational Measurement****,* ***22****, 231–240.**

Boal, J., Fabacher, D., Miller, R., Siu, A., Kantor, B. & Flaherty, J. (2001). Validation of an instrument designed to assess medical student attitudes toward home care. ***Journal of the American Geriatrics Society****,* ***49****, 470–473.**

Coccaro, E. F. & Miles, A. M. (1984). The attitudinal impact of training in gerontology/geriatrics in medical school. *Journal of the American Geriatrics Society*, *32*, 762–768.

DeVellis, R. (2003). *Scale development: Theory and applications*. Thousand Oaks, CA: Sage.

Downing, S. M. (2004). Reliability: On the reproducibility of assessment data. *Medical Education*, *36*, 1006–1012.

Fitzgerald, J. T., Wray, L. A., Halter, J. B., Williams, B. C., & Supiano, M. A. (2003). Relating medical students' knowledge, attitudes and experience to an interest in geriatric medicine. *The Gerontologist*, *43*, 849–855.

Ford, A., Liske, R., & Ort, R. (1962). Reactions of physicians and medical students to chronic illness. *Journal of Chronic Diseases*, *15*, 785–794.

Guilford, J. P. (1954). *Psychometric methods*. New York: McGraw-Hill.

Kishimoto, M., Nagoshi, M., Williams, Masaki, K. H., & Blanchette, P. L. (2005). Knowledge and attitudes about geriatrics of medical students, internal medicine residents, and geriatric medicine fellows. *Journal of the American Geriatrics Society*, *53*, 99–102.

Maxwell, A. J. & Sullivan, N. (1980). Attitudes toward the geriatric patient among family practice residents. *Journal of the American Geriatrics Society*, *28*, 341–345.

Moss, P. A. (1994). Can there be validity without reliability? *Educational Researcher*, *23*(2), 5–12.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Parker, S. (1960). The attitudes of medical students toward their patients. *Journal of Medical Education*, *35*, 849–855.

Pilotte, W. J. & Gable, R. K. (1990). The impact of positive and negative item stems on the validity of a computer anxiety scale. *Educational and Psychological Measurement*, *50*, 603–610.

Reuben, D. B., Lee, M., Davis, J. W., Eslami, M. S., Osterweil, D. G., Melchiore, S., & Weintraub, N. T. (1998). Development and validation of a geriatrics attitudes scale for primary care residents. *Journal of the American Geriatrics Society*, *46*, 1425–1430.

Rosencranz, H. A. & McNevin, T. E. (1969). A factor analysis of attitudes toward the aged. *The Gerontologist*, *9*, 55–59.

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, *8*(4), 350–353.

Schriesheim, C. A., Eisenbach, R. J., & Hill, K. D. (1991). The effect of negation and polar opposite item reversals on questionnaire reliability and validity: An experimental investigation. *Educational and Psychological Measurement*, *51*, 67–78.

Schriesheim, C. A. & Hill, K. D. (1981). Controlling acquiescence response bias by item reversals: The effect on questionnaire validity. *Educational and Psychological Measurement*, *41*, 1101–1114.

Spector, P. E. (1991). *Summated rating scale construction*. New York: Sage Pub.

Spence, D. L., Feigenbaum, E. M., Fitzgerald, F., & Roth, J. (1968). Medical student attitudes toward the geriatric patient. *Journal of the American Geriatrics Society*, *16*(9), 976–983.

Streiner, D. L. & Norman, G. R. (2003). *Health measurement scales* (3rd ed.). New York: Oxford University Press.

Streiner, D. L. (1996). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment*, *80*(1), 99–103.

Stewart, T. J. & Frye, A. W. (2004). Investigating the use of negatively-phrased survey items in medical education settings: Common wisdom or common mistake? *Academic Medicine*, *79*(10 Supplement), S1–S3.

Statistical Package for the Social Sciences, 12.0. (2004). Chicago, Ill: SPSS.