APPLIED MEASUREMENT IN EDUCATION, 8(3), 199–209

Connotatively Inconsistent Test Items

Lei Chang
College of Education
University of Central Florida

An item is defined as connotatively consistent (CC) or connotatively inconsistent (CI) when its connotation agrees with or contradicts that shared by the majority of items on a test. This definition is more accurate than what have been referred to as negative versus positive items. The study examined the equivalence of CC and CI items using convenience samples of college students' responses to the Life Orientation Test (Scheier & Carver, 1985). Confirmatory factor analysis results showed that CC and CI items measured correlated but distinct traits. Practical and theoretical implications of these findings are discussed.

What are negative items? What do they measure? In this article, I (a) clarify the concept of negative versus positive items and (b) examine the psychometric equivalence of these two kinds of items.

WHAT ARE NEGATIVE ITEMS?

Syntax and Semantics

A test item[1] can be made grammatically negative by including, in an affirmative statement, the adverb *not* together with helping verbs, such as *do, have, will, should,* and so forth, or by adding the adverb after the connecting verb, *be*. Thus, test items such as "I do not like computers" or "This manual is not well written," are grammatically negative. Apparently, when researchers talk about negative items, they refer to more than just grammatically negative

Requests for reprints should be sent to Lei Chang, Department of Educational Foundations, University of Central Florida, Orlando, FL 32816–1250.

[1]Test items means items from attitude or personality measures or tests of sentiments. The word *tests* as used in this article does not include achievement tests.

statements. For instance, "I dislike computers," "This manual is badly written," or "He is a difficult person," may also be considered negative even though they are grammatically affirmative statements, whereas a grammatically negative statement (e.g., "I'm not nervous") is not necessarily a negative item. Thus, the so-called negative items do not have to have a negative syntax.

Some researchers call negative items "semantically negative" (Ahlawat, 1985), meaning they have a negative meaning. Such a definition also lacks accuracy. *Semantic* refers to the formal meaning or nature of a statement free from value judgment or sentiment. The value judgment or feeling of a word or statement is represented by its connotation. For example, the words *frugal* and *miserly* have the same formal meaning, whereas one has a positive or neutral connotation and the other has a negative connotation. Similarly, a test item can be connoted as positive or negative, whereas its semantics are free from a positive or negative sentiment. Thus, defining items as semantically negative is incorrect.

Value Implications

Another ambiguity regarding negative items used in the literature is the distinction between what is positive and what is negative. The connotation of an item is positive or negative depending on the context in which the item is presented. Because most social science research addresses issues that reflect a public conception of what is desirable and undesirable in a society at a given time, test items connoted as positive or negative have time and cultural constraints. For example, with the current public obsession of thinness, any *thin* words (e.g., *skinny*) tend to be greeted as compliments, whereas any *fat* words (e.g., *plump*) tend to be perceived as insults even though *skinny* and *plump* used to have negative and positive connotations, respectively. Because values are culturally defined, what is desirable in one culture may not be so in another. Personality traits, such as obedience, compliance, or submission, which are devalued in Western cultures, are sometimes endorsed in Eastern cultures. Cross-cultural studies of such personality traits may become confusing or ethnocentric if items of the related personality instruments are defined as connotatively positive or negative. To avoid ambiguity and to maintain scientific neutrality (research itself should be value free (Borg & Gall, 1989) even though the issues researched in the social sciences may bear timely and culturally defined values), therefore researchers should avoid labeling items negative or positive.

Redefining Negative Items

As previously suggested, items should be defined not in terms of the manifest syntax or semantics but in terms of the underlying connotation. The

opposite connotations of items represent two directions of a latent construct continuum of which items or their semantics are indicators. It was further explained that labeling the two directions of a trait continuum as positive and negative implies a value judgment that may not be generalizable across time or culture. There is a more generalizable way to represent item directionality: A test item is defined as connotatively consistent (CC) or connotatively inconsistent (CI) when the connotation of the item agrees with or contradicts that shared by the majority of the items making up a test or a subscale of a test. If half of the items on a test represent one direction and the other half represent the opposite direction in connotation, all items on the test are considered connotatively inconsistent. The directionality of items can thus be universally defined as CC or CI disregarding the grammatical form—negative or affirmative—or value implication of the items.

ARE CC AND CI ITEMS EQUIVALENT?

Do CC and CI items measure the same thing? In other words, do they differentiate individuals with respect to an underlying trait in the same way but along two opposite ends of the trait continuum? Most researchers seem to assume they do, as is reflected in the common practice of reversing the scale points of the CI items and analyzing the reversed CI items together with the CC items on a test. For instance, such reversals are performed when computing internal consistency reliability estimates, making composite scores, or comparing scale means. However, few researchers have investigated this assumption. Two studies that investigated this issue both raised questions regarding the equivalence of CC and CI items. From an analysis of variance (ANOVA) comparing four kinds of item presentation, Ahlawat (1985) concluded that "semantically negative and positive item contents do not measure essentially the same construct" (p. 98). His study was based on a sample of Jordanian middle-school students using an Arabic translation of the State–Trait Anxiety Inventory (Spielberger, Gorsuch, & Lushene, 1970). Chang (1993) compared CI items and their CC counterparts using generalizability theory and found that the reversed CI and CC items were not fully exchangeable.

Both of these studies used an ANOVA approach, whereas a factor analysis can more directly answer the question of whether CC and CI items measure the same thing or different things. Research has shown that some respondents particularly endorse response options appearing at the beginning (Carp, 1974; Chan, 1991; Payne, 1972) or end (McClendon, 1986) of a scale (e.g., the Likert-type scale). This response-ordering effect was not controlled in either of the two previous studies. Rating a CC and CI item means endorsing the scale points appearing on the two opposite ends of a scale. If there is a primacy or recency effect, for example, a rating of 1 for a CI item on a 7-point Likert-type scale would not truly equal a 7 if the item had been CC.

Findings from these two studies can also be better corroborated by employing a different and smaller number of scale points. The reason for using fewer scale points is that ratings on a larger number of scale options are more likely to be inconsistent because of the increased information-processing demand (Chang, in press). Ahlawat (1985) and Chang (1993) used 5-point and 6-point Likert-type scales, respectively. If their findings can be replicated with a 4-point scale, a finding of inequivalence due to connotative inconsistency will be less likely to be confounded by inconsistency due to the cognitive demand for processing multiple-response options. The remainder of the article presents an empirical investigation of the equivalence between CC and CI items of a 4-point Likert-type scale using confirmatory factor analysis.

METHODS

Instrument

The instrument used to evaluate the equivalence of item connotations is the Life Orientation Test (LOT, Scheier & Carver, 1985). The LOT has eight items measuring a single-trait dispositional optimism. The eight items are of opposite connotations, with four items representing optimism and four indicating pessimism. The eight LOT items follow[2]:

1. In uncertain times, I usually expect the best. (optimism)
2. If something can go wrong for me, it will. (pessimism)
3. I always look on the bright side of things. (optimism)
4. I'm always optimistic about things. (optimism)
5. I hardly ever expect things to go my way. (pessimism)
6. Things never work out the way I want them to. (pessimism)
7. I'm a believer in the idea that "every cloud has a silver lining." (optimism)
8. I rarely count on good things happening to me. (pessimism)

Connotations of the items were reconstructed to make four versions of the LOT. Version 1 was the original LOT as previously listed. Items in this version were connotatively inconsistent. In the second version, the connotations of the eight items were reversed. For example, Item 1 was changed into "In uncertain times, I usually expect the worst" and Item 8 was changed into "I always count on good things happening to me." Items in this version,

[2]The original test also contains four filler items that the authors used to disguise the purpose of the test. The four filler items are not presented here. Words in parentheses were added by myself. From "Optimism, coping, and health: Assessment and implications of generalized outcome expectancies" by M. F. Scheier and C. S. Carver, 1985, *Health Psychology, 4,* p. 225. Copyright 1985 by Lawrence Erlbaum Associates, Inc. Adapted by permission.

which represented the opposites of the first version, were also connotatively inconsistent. In the third and fourth versions, the connotation of the four pessimism items and the connotation of the four optimism items were reversed, respectively. Thus, both of these versions were connotatively consistent, one connoting optimism and the other connoting pessimism. In all four versions the ordering of items was the same as shown above.

The LOT was originally developed and used with a 5-point Likert-type scale ranging from *strongly disagree* (0) to *strongly agree* (4). Four scale points were used in the present study ranging from *strongly disagree* (1) to *strongly agree* (4). The neutral point was eliminated.

To rule out possible primacy or recency effects, the ordering of scale points for the two connotatively consistent versions was reversed. For the optimism version, the scale was presented as, from left to right, *strongly disagree* (1) to *strongly agree* (4). For the pessimism version, the scale appeared as, from left to right, *strongly agree* (4) to *strongly disagree* (1). If respondents had the tendency to particularly endorse either end of the scale, such a primacy or recency effect would be a constant in relation to the opposite connotations of the two versions.

Samples

Subjects were undergraduate students enrolled in different courses in a metropolitan university. About half of the students were administered the two connotatively consistent versions and the other half were administered the two versions in which items had inconsistent connotations. Thus, each student was administered the LOT twice, at a week apart. The order of administrations of the two versions was mixed among students. Students' responses were anonymous. To match the two administrations per respondent, students were asked to put a same nickname, which was only known to them, on their two responses. The resulting samples are as follows: For the students receiving connotatively consistent items, the optimism version had 129 responses, the pessimism version had 149 responses, and there were 107 students with useable responses to both versions. For those receiving the two connotatively inconsistent versions, one version had 211 responses, the other had 204 responses, and there were 139 students with useable responses to both versions. Sample size discrepancies between the number of useable responses for both versions and the number responding to either one of the two versions were because some students were absent or did not provide their nicknames during one of the two test administrations. Prior to data analysis, scale points of the appropriate items were reversed to make all four versions of LOT in the same direction of optimism.

LISREL Issues

Maximum likelihood (ML) estimation in LISREL-7 (Jöreskog & Sörbom, 1988) was used to conduct the confirmatory factor analyses. Lab tests indi-

cated the robustness of ML for ordinal or censored data (Jöreskog & Sörbom, 1988; Muthen & Kaplan, 1985), such as Likert-type scales.

The following goodness-of-fit tests were examined for the confirmatory factor analyses (CFA): The overall χ^2 test indicates the difference of lack of fit between the testing model and a saturated model of perfect fit. The ratio of χ^2 to degrees of freedom (χ^2/df) examines the lack of fit of the testing model in relation to the model's complexity or simplicity. The goodness-of-fit index (GFI) shows the relative amount of variance and covariance jointly explained by the testing model. The adjusted goodness-of-fit test is the same as GFI with adjustment for degrees of freedom of the testing model. The root mean square residual shows average discrepancy between the elements in the testing model and the sample covariance matrix. Finally, the χ^2 difference test compares the difference of lack of fit between two nested models. Two subjective indices of fit were also evaluated. One was the Bentler and Bonett (1980) normed-fit index (BBI). The other was the Tucker and Lewis (1973) nonnormed-fit index (TLI). Both indices are among the most widely used subjective measures of fit in the LISREL literature (e.g., Marsh 1993; Sternberg, 1992). Both BBI and TLI range from 0, indicating total lack of fit, to 1.00 showing perfect fit.

ANALYSES AND RESULTS

Means, standard deviations, and correlations of the four versions of the LOT items are presented in Table 1.

To answer the question of whether CC and CI items measured the same thing or different things, two parameter-nested models were tested within each of the two connotatively consistent and two connotatively inconsistent samples. One model specified a single factor as was intended by the LOT. The competing model parameterized two factors, one factor representing the four items originally having the optimism connotation and the other factor representing the four original pessimism items. In addition, a no-factor or null model was tested within the four samples to set a baseline for evaluating the goodness-of-fit of the two competing models. Results are reported in Table 2.

There was a contrasting difference between the results obtained from the two connotatively consistent and the two connotatively inconsistent samples. When the items were consistent in connotation, either of optimism or pessimism, the 1-factor model represented the best fit to the data. An additional factor did not improve the goodness-of-fit over the 1-factor model for either the optimism or pessimism sample. In contrast, for the other two samples, one factor clearly failed to account for the covariations among the connotatively inconsistent items. Two factors, one representing optimism items and one representing pessimism items fit the data extremely well. These results indicated that CC and CI items did not measure the same trait. When the items were CC, they measured one trait. When the items were CI,

TABLE 1
Means, Standard Deviations, and Correlations of the Life Orientation Test Items

	Optimism[a]		*Pessimism*[b]		
Item	*M*	*SD*	*M*	*SD*	*Correlations*[c]
1	2.87	.89	2.93	.99	.45
2	2.92	.92	3.26	.85	.43
3	3.19	.85	3.49	.70	.32
4	3.12	.86	3.46	.77	.47
5	2.98	.79	3.41	.74	.30
6	2.78	.75	3.44	.72	.33
7	3.00	.84	3.27	.85	.35
8	3.11	.79	3.35	.84	.41
	Optimism–Pessimism[d]		*Pessimism–Optimism*[e]		
Item	*M*	*SD*	*M*	*SD*	*Correlations*[f]
1	2.64	.99	2.80	1.04	.37
2	3.01	.93	2.34	.98	.18
3	2.73	.95	3.29	.87	.48
4	2.68	.93	3.15	.88	.49
5	3.10	.89	2.60	.86	.38
6	3.18	.84	2.42	.78	.26
7	2.79	.89	3.00	.88	.16
8	3.25	.87	2.84	.83	.45

Note. Scale points of related items are reversed to make all items represent the same scale direction of optimism.

[a]$n = 129$. [b]$n = 149$. [c]Correlation between each of the eight optimism items and its corresponding pessimism items, $n = 107$ (e.g., Item 1 optimism with Item 1 pessimism, $r = .45$). [d]Eight optimism–pessimism items where Items 1, 3, 4, and 7 are of optimism connotation and Items 2, 5, 6, and 8 are of pessimism connotation, $n = 211$. [e]Eight pessimism–optimism items where Items 1, 3, 4, and 7 are of pessimism connotation and Items 2, 5, 6, and 8 are of optimism connotation, $n = 204$. [f]Correlation between each of the eight optimism–pessimism and the corresponding pessimism–optimism items, $n = 139$ (e.g., Item 1 optimism–pessimism with Item 1 pessimism–optimism, $r = .37$).

they measured two traits that were in alignment with the inconsistent connotations of the items.

To further test whether connotations of the items changed what was being measured, the same items of the two opposite connotations were tested for equivalence. Three sets of two CFA models, 1-factor versus 2-factor, were conducted. One set was conducted within the sample of 107 cases that had complete responses to the two versions of eight consistent optimism and eight consistent pessimism items. The 1-factor model had the eight items of optimism connotation and the same eight items of pessimism connotation loaded onto one common factor. This model was tested against the alternative 2-factor model in which the eight optimism items and the eight pessimism items loaded onto two related but distinct factors.

The other two sets of two CFA models each, 1-factor versus 2-factor, were conducted within the sample 139 cases that had complete responses to the

two connotatively inconsistent versions of the LOT. To repeat what was stated earlier, in one version, four items had an optimism connotation and the other four items had a pessimism connotation. In the other version, the opposite halves of each set of the four items had these two inconsistent connotations. The two sets of CFA models examined half of the 16 items (from two versions) at a time. In one CFA set, the four items of optimism and the same four items of pessimism were parameterized to load onto either one common factor (1-factor model) or two related factors (2-factor model). In the other set, the four pessimism items and their optimism counterparts were loaded onto either one factor (1-factor model) or two factors (2-factor model).

Results from these three sets of CFA models are reported in Table 3. The 2-factor models fit the data statistically or showed an overwhelming superiority over the 1-factor models. Identical items of inconsistent connotations measured two related traits instead of one common trait. These results from repeated measures analyses rendered stronger indications of the inequivalence between CC and CI items. The two sets of repeated measures from the same persons corresponding to the same items would not converge onto the same trait when the items were presented in different connotations.

TABLE 2
Goodness-of-Fit Indices of 1-Factor and 2-Factor Models

Model	χ^2	*df*	χ^2/df	*GFI*	*AGFI*	*RMR*	*BBI*	*TLI*	$\Delta\chi^{2a}$
Consistent connotation									
Optimism[b]									
Null	829.3	28	29.61	.243	.026	.589			
1-factor	53.4	20	2.67	.903	.826	.036	.935	.941	
2-factor	52.1	19	2.74	.906	.822	.036	.937	.939	1.3
Pessimism[c]									
Null	671.3	28	23.97	.325	.132	.480			
1-factor	40.6	20	2.03	.937	.886	.041	.939	.955	
2-factor	38.5	19	2.02	.943	.891	.043	.942	.955	2.1
Inconsistent connotation									
Optimism–Pessimism[d]									
Null	828.8	28	29.60	.426	.262	.387			
1-factor	289.7	20	14.48	.699	.458	.168	.650	.528	
2-factor	31.5	19	1.65	.964	.931	.053	.962	.977	257.7
Pessimism–Optimism[e]									
Null	482.1	28	17.22	.556	.429	.298			
1-factor	126.1	20	6.30	.845	.721	.107	.738	.673	
2-factor	29.5	19	1.55	.967	.938	.053	.938	.966	96.6

Note. df = degrees of freedom; GFI = goodness-of-fit index; AGFI = adjusted goodness-of-fit test; RMR = root mean square residual; BBI = Bentler and Bonett normed-fit index; TLI = Tucker and Lewis nonnormed fit index; Chi-square and chi-square difference tests are all significant at $\alpha < .001$, except for those that are underlined that are not significant at $\alpha = .05$.

[a]Chi-square difference test between the 2-factor model and the 1-factor model with 1 degree of freedom. [b]$n = 129$. [c]$n = 149$. [d]$n = 211$. [e]$n = 204$.

TABLE 3
Goodness-of-Fit Indices of Models Testing Equivalence of Item Connotations

Model	χ^2	*df*	χ^2/df	*GFI*	*AGFI*	*RMR*	$\Delta\chi^{2a}$
8 identical items of optimism versus pessimism connotation[b]							
1-factor	443.7	104	4.27	.573	.441	.130	
2-factor	203.7	103	1.98	.886	.817	.064	240
4 identical items of optimism versus pessimism connotation[c]							
1-factor	109.9	20	5.49	.821	.677	.090	
2-factor	22.7	19	1.19	.962	.929	.045	87.2
4 identical items of pessimism versus optimism connotation[c]							
1-factor	86.5	20	4.33	.842	.715	.103	
2-factor	23.2	19	1.22	.958	.921	.052	53.3

Note. Chi-square and chi-square difference tests are all significant at $\alpha < .001$, except for those that are underlined, which are not significant at $\alpha = .05$.

[a]Chi-square difference test between the 1-factor model and 2-factor model with 1 df. [b]The analyses were based on 107 cases that had repeated responses to the same eight items that were first connoted as optimism and then connoted as pessimism. [c]Both set of analyses were based on 139 cases that responded to the same eight items of inconsistent connotations over two administrations. In one administration, there were four optimism and four pessimism items. In another administration, the opposite set of the four items were of optimism and pessimism connotations. The analyses compared eight connotatively inconsistent items at a time.

DISCUSSION

With several methodological improvements, this investigation confirms the findings of two earlier studies. Together, these studies offer important implications for measurement practice and theory development. One practical implication is that CI items should not be used. Items on a test or representing a scale of a test should have consistent connotations with respect to the construct being measured. A CC and a CI item are not bipolar indicators of a common trait continuum and, therefore, construct unidimensionality cannot be maintained by simply reversing the scale points associated with the CI item. Consequently, inconsistent connotations are likely to change the intended factor structure of a test. Researchers should not deliberately use CI items, even for the purpose of countering response-set effect. Including filler items in a test may be a more effective way to evaluate and partial out response set. As an aside, the possible belief that certain CI items cannot be replaced by CC counterparts may be unjustifiable. If researchers are willing to be thoughtful about item writing, almost any CI items can be changed to be connotatively consistent with the rest of the items on a test.

Including many CI items in a test may have the impact of altering the purported operational definition of the underlying construct. For example, the LOT, given the way the items are currently connoted, does not measure one concept of dispositional optimism as the authors of the test intended. There are two related concepts reflecting the optimistic and pessimistic connotations of the items. Researchers either have to make the connotation of the items consistent, in one direction or the other, to maintain the intended

1-factor concept or have to redefine the construct to reflect the 2-factor structure. Research may be needed to reexamine the construct validity of other tests that use a large number of CI items. This additional note for measurement practice leads to broader theoretical implications.

The purpose of measurement is to quantify individual differences with respect to some trait so that some people or things can be shown to have more or less of the trait. This purpose is achieved by ordering people or things along a set of test items as indicators of the underlying traits. Individual differences measured by these items and summarized in a variance–covariance matrix can then be factor analyzed to discern the underlying trait patterns and relations. Findings from this study indicate that latent constructs may be imbedded in or inseparable from the connotations of the items that are indicators of the constructs.

In other words, item connotations represent a systematic source of variance independent from the true individual differences along the constructs the test intends to measure. The systematic connotation variance may interfere with the identification of the intended construct structure in varying ways depending on the number of constructs purportedly measured by a test and the number of CI items contained in the test. With the present unidimensional measurement instrument, the intended 1-factor was extracted from CC items, whereas connoting half of the CC items into CI resulted in a 2-factor structure. Replication studies may be conducted to examine the specific factorial changes associated with varying numbers of CI items defining varying numbers of constructs in multidimensional tests.

The literature on multitrait–multimethod observes, more often than not, unsatisfactory rather than satisfactory convergent validity relating multimethods to a common trait (Fiske & Campbell, 1992). The present study expands the scope of the traditional multitrait–multimethod issue and its general notion that the nature of a construct to be measured is in part determined by how it is measured. Extending the interaction and interdependence between a construct and a method purporting to measure it (Campbell & Fiske, 1959; Fiske & Campbell, 1992), one may contend that, within the same paper-and-pencil test, the meaning of a construct is also restricted by the connotations of the items. It makes sense that the presence of a construct often indicated by CC items does not necessarily mean the opposite (reversal of scale points) of the absence of a construct often indicated by the CI counterparts on a test. For example, "feeling happy" is not the same as not "feeling sad" (reversal of scale options). This last point also suggests a cognitive perspective about construct validity that is often regarded as a purely measurement issue. Just as certain concepts are inseparable from the language used to express them, so are the constructs in measurement inseparable from the way their item indicators are connoted. These are some thoughts about measurement theory suggested by the present study that may be interesting and rewarding to ponder.

ACKNOWLEDGMENT

The author thanks James C. Impara and three anonymous reviewers for their constructive comments on an earlier version of this article.

REFERENCES

Ahlawat, K. S. (1985). On the negative valence items in self-report measures. *Journal of General Psychology, 112*(1), 89–99.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88,* 588–606.

Borg, W. R., & Gall, M. D. (1989). *Educational research: An introduction* (5th ed.). New York: Longman.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait–multimethod matrix. *Psychological Bulletin, 56,* 81–105.

Carp, F. M. (1974). Position effects on interview responses. *Journal of Gerontology, 29,* 581–587.

Chan, J. C. (1991). Response-order effects in Likert-type scales. *Educational and Psychological Measurement, 51,* 531–540.

Chang, L. (1993, June). *Connotatively consistent and connotatively inconsistent items are not fully equivalent.* Paper presented at the Annual Convention of the American Psychological Society, Chicago.

Chang, L. (in press). A psychometric evaluation of 4-point and 6-point Likert-type scales in relation to test reliability and validity. *Applied Psychological Measurement.*

Fiske, D. W., & Campbell, D. T. (1992). Citations do not solve problem. *Psychological Bulletin, 12,* 393–395.

Jöreskog, K. G., & Sörbom, D. (1988). *LISREL 7: A guide to the program and application.* Chicago: Spss Inc.

Marsh, H. W. (1993). Stability of individual differences in multiwave panel studies: Comparison of simplex models and one-factor model. *Journal of Educational Measurement, 30,* 157–183.

McClendon, M. J. (1986). Response-order effects for dichotomous questions. *Social Science Quarterly, 67,* 205–211.

Muthen, B., & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology, 38,* 171–189.

Payne, J. D. (1972). The effects of reversing the order of verbal rating scales in a postal survey. *Journal of Market Research Society, 14,* 30–44.

Scheier, M. F., & Carver, C. S. (1985). Optimism, coping, and health: Assessment and implications of generalized outcome expectancies. *Health Psychology, 4,* 219–247.

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

Sternberg, R. J. (1992). Psychological Bulletin's top 10 "Hit Parade." *Psychological Bulletin, 112,* 387–388.

Tucker, L. R., & Lewis, C. (1973). The reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38,* 1–10.