Journal of Collaborative Leadership (2017) Volume 1, pps. 22-47
ISSN # 2472-9248

More on the Construct Validity of the Authentic Leadership Questionnaire

Malcolm James Ree

Our Lady of the Lake University

San Antonio, TX

Thomas R. Carretta

Air Force Research Laboratory

Wright-Patterson AFB, OH

The views expressed are those of the authors and not necessarily those of the US Government, US Department of Defense, or the US Air Force. Send correspondence to Dr. Malcolm James Ree at malcolmree@att.net.

Abstract

Several questionnaires are in widespread use to assess various aspects of leadership. These include the *Multifactor Leadership Questionnaire* (MLQ; Bass & Avolio, 2000), the *Leader Behavior Description Questionnaire* (LBDQ, Stogdill, 1963), and the *Authentic Leadership Questionnaire* (ALQ; Walumbwa, Avolio, Gardner, Wernsing & Peterson, 2008). For both theoretical and practical reasons, it is important to determine the extent to which the constructs measured by these instruments overlap with one another or are unique. The current study examined the construct validity of the ALQ. Participants were rated by colleagues on the ALQ, MLQ and LBDQ. The total number of raters was 1,007 and each person was rated anonymously by multiple raters including supervisors, peers, and subordinates. The ALQ score was regressed on scores from the MLQ and LBDQ to assess its construct validity. The observed correlation between transformational leadership style and authentic leadership, $r = .79$, was consistent with meta-analytic results. Corrected for unreliability of both variables, the correlation was .89. Results for the regression analyses indicated that there is little difference in the traits measured by the MLQ and ALQ. Given these regression findings, the LBDQ was removed from the data and confirmatory factor analyses were conducted and evaluated. The CFA showed a two factor solution representing the MLQ and ALQ. The strong correlation ($r = .89$) between these factors suggests a common source of variance for these two measures. Clear understanding of constructs facilitates collaboration among practitioners and researchers.

Key Words: authentic leadership, construct validity, leadership behavior, leadership style

More on the Construct Validity of the Authentic Leadership Questionnaire

Constructs are abstractions, often thought of as latent variables, that are reflected in the scores on a measure such as a questionnaire or test. Construct validity can be defined as the extent that an instrument or test assesses what it is designed to measure (Brown, 1996; Cronbach & Meehl, 1955). This type of validity is vitally important as it directs the appropriateness of making and using inferences from measures of the construct. Also, it is important that all researchers and practitioners ascribe the same meaning to a construct. This facilitates collaborative communication and directs research and practice.

This study examined the construct validity of leadership measures designed to assess different constructs. Convergent and discriminant validity are ways to assess the construct validity of a measure. Convergent validity helps establish construct validity when more than one measurement method, measurement instrument (e.g. a different test of the same construct) or procedure is used to collect data about a construct. Discriminant validity helps establish construct validity by showing that the construct being measured differs from other constructs. Both convergent validity (correspondence or convergence among similar constructs) and discriminant validity (discrimination among dissimilar constructs) are necessary to establish construct validity. New constructs and measures often are proposed without adequate examination of their convergent and discriminant validity with other established constructs and measures.

Cronbach and Meehl (1955) proposed a three-step process for examining construct validity. These steps involve 1) describing a set of theoretical concepts and their expected relationships to one another, 2) developing methods to measure these constructs, and 3) empirically testing the hypothesized relations among the measures. Although the concept of construct validity has existed for over 60 years, it is ignored all too often when new theoretical

constructs are proposed. A danger is that well-known theoretical constructs might be repackaged as a new construct or that two or more well-known theoretical constructs might be combined and proposed as a unique new construct. This wastes resources, adds confusion to the literature, and muddles thinking. An illustrative example of this is found in studies of leadership.

An Example of Construct Confusion: Authentic Leadership and Transformational Leadership

Authentic (Walumbwa, Avolio, Gardner, Wernsing, & Peterson, 2008) and transformational leadership (Bass, 1985) are two popular leadership constructs. Authentic leadership is an approach that emphasizes building the leader's legitimacy through fostering ethical, honest relations with followers Authentic leaders generally have a positive outlook, truthful self-concept, and promote openness. Transformational leadership is an approach that emphasizes influencing followers to transcend self-interests for the good of the organization. Transformational leaders provide a vision, set an example, communicate high performance expectations, demonstrate sensitivity to the needs of individual followers, encourage a team attitude, and provide intellectual stimulation.

Although these two constructs were specified as differing, scores on the instruments designed to assess them did not support this distinction. A meta-analysis of authentic leadership (AL) and transformational leadership (TL) examined the extent to which these constructs were related (Banks, McCauley, Gardner, & Guler, 2016). The data included 25,452 individuals from 100 studies. The meta-analytic correlation between authentic leadership and transformational leadership was .72. Also, results for predictive validation studies were mixed. AL was more predictive than TL and showed dominance (Budescu, 1993) over TL for organizational citizenship behaviors and group or organizational performance criteria. TL was superior to AL

when predicting follower satisfaction, leader effectiveness, and task performance. However, Banks et al. (date) did not correct for differential range restriction across the studies, suggesting that the meta-analytic correlation may be an underestimate. Further, the lack of correction for differential range restriction and for study-specific reliability reduces the interpretability of the validity differences.

Purpose

This paper extends the examination of the construct validity of authentic leadership by including additional widely-used measures of leadership. Leadership style was broadened to include both transformational and transactional leadership, given the moderate to strong correlation between them. Also, leadership behavior was assessed because behavior is frequently a manifestation of internal states as measured by the ALQ. These constructs were chosen because the theoretical foundations of the constructs were specified as different and therefore the scores from tests of the constructs should not show a strong correlation. A confirmatory factor analysis was conducted to disclose the relationships among the latent constructs measured by the instruments.

Methods

Participants

Participants were 1,007 subordinates, peers, or supervisors who rated their leader. The leaders were enrolled in one of several graduate programs at a south Texas university. They were rated on the *Authentic Leadership Questionnaire* (Walumbwa, Avolio, Gardner, Wernsing, & Peterson, 2008), *Leader Behavior Description Questionnaire XII* (Stogdill, 1963), and *Multidimensional Leadership Questionnaire Form 5X* (Bass, 1985).

Measures

Authentic Leadership Questionnaire (ALQ)

The ALQ (Walumbwa, Avolio, Gardner, Wernsing, & Peterson, 2008) was designed to measure and provide scores for relational transparency (TRANS), internalized moral perspective (MORAL), balanced processing (BAL_P), and self-awareness (SELF). The AL total score composite is the average of the four scores. The four AL component scores are intended as a diagnostic tool to identify strengths and weaknesses. The following reliability estimates are from the developmental study. Internal consistency reliabilities were: relational transparency, .77, internalized moral perspective, .73, balanced processing, .70, and self-awareness, .73 (Walumbwa et al., 2008).

Leadership Behavior Description Questionnaire (LBDQ)

The most current version of the *Leadership Behavior Description Questionnaire Form XII* (Stogdill, 1963) has 100 items which are combined into 12 subscales. Only the two subscales that assess leadership behavior, initiation of structure (INIST) and consideration (CONS), were administered in this study. Both of these subscales have 10 items. The LBDQ manual provides a historical review of how the scales were developed and identifies the two leadership behavior subscales (Stogdill, 1963). Stogdill provided internal consistency reliabilities for the scales for nine groups of leaders (e.g., army division, highway patrol, aircraft executives, ministers, community leaders). Median reliabilities for these groups were .77 for initiation of structure and .78 for consideration.

Multidimensional Leadership Questionnaire (MLQ)

The MLQ (Bass, 1985) was designed to assess the "full range" of leadership styles (Bass, 1990; Bass & Avolio, 1995, 2000), including transactional (TA), transformational (TL), and non-leadership (laissez-faire, passive avoidant). Forty-five items are combined into 12 scales: idealized influence – attributed (IIA), idealized influence – behavior (IIB), inspirational motivation (IM), intellectual stimulation (IS), individual consideration (IC), contingent reward (CR), management-by-exception – active (MBEA), management-by-exception – passive (MEP), laissez-faire leadership (LFL), extra effort (EE), effectiveness (EFF), and satisfaction (SAT). The first nine scales are created by averaging the responses to the items. These scales are then combined to create three leadership style composites representing transactional (CR, MBEA) transformational (IIA, IIB, IM, IS, IC), and non-leadership (MEP, LFL). The non-leadership scales (MEP, LFL) and the outcome scales (EE, EFF, and SAT) were not -used in the regression or confirmatory factor analyses and are described here only for completeness. Reliability estimates for each leadership scale ranged from .74 to .94 in the normative sample (Bass & Avolio, 1995, 2000). These estimates are based on ratings of a leader evaluated by others.

Composites versus Factor Scores

The term factor is sometimes used in an offhand manner. A factor includes all the variables in a factor analysis and is computed by multiplying scores on variables by factor weights. If a factor influences five variables, the factor score must be a weighted sum of the products of the five variables multiplied by their factor weights. Composites are defined as a simple weighted sum of scores and are frequently more appropriate. These weights are not derived from a statistical estimation procedure. For example, if the scores on the five variables

are added together and a sum is computed, a composite score has been created. Factor scores and composite scores are not necessarily equivalent. In this study variables are composite scores except in the confirmatory factor analyses.

Analyses

Analyses began with an examination of the descriptive statistics (means, standard deviations) for the study variables. Also, the correlations of the variables were estimated. Following Le, Schmidt, Harter, and Lauver (2010) the correlations were investigated concerning construct redundancy. Redundancy means that the construct is already measured by another instrument and the redundant instrument is not needed. This was tested by using the LBDQ CONS score as an external criterion to be predicted by AL and TL. Similarity of correlations of AL and CONS and TL and CONS would add to the evidence for construct redundancy.

Each of the four *Authentic Leadership Questionnaire* (ALQ) scores; relational transparency (TRANS), internalized moral perspective (MORAL), balanced processing (BAL_P), and self-awareness (SELF), and the total ALQ score (AL), was regressed on MLQ transformational leadership (TL), MLQ transactional leadership (TA), LBDQ consideration (CONS), and LBDQ initiation of structure (INIST) scores to assess construct validity.

Regression equations were estimated for complete models using TL, TA, CONS, and INIST, while reduced models contained only TL and TA. This was done to evaluate the contribution of the LBDQ to the prediction of AL beyond that provided by the MLQ. All of the regression equations were tested to determine if they were statistically significant. For each criterion variable (the four subtests of the ALQ and the total AL scores), the complete and

reduced models were tested against each other using the usual *F*-test. All statistical tests used a $p < .05$ Type I error rate.

Finally, confirmatory factor analyses (CFAs) were conducted to disclose factor structure. Model fit was evaluated. Model 1 (M1) was a single factor model. M1 was estimated to determine if a parsimonious model described the data. Model 2 (M2), a measurement model, had a factor for each instrument. Subtests from one instrument were not allowed to cross-load on the factor representing the other instrument and the factors were allowed to correlate.

Results

Correlational Statistics

As shown in Table 1, the observed correlations between the MLQ transformational leadership scales (IIA, IIB, IM, IS, and IC) ranged from .48 to .68 with the four ALQ scores. The correlations for the MLQ transactional leadership scales (CR and MBEA) were mixed with low correlations for MBEA and the ALQ scales (.04 to .13). Correlations between the LBDQ and ALQ scores were stronger for consideration (.47 to .54) than for initiation of structure (.24 to .36). Correlations among the ALQ scores ranged from .53 to .73, suggesting a general factor. These are lower than the meta-analytic correlations reported by Banks et al. (2016). Their values were between .84 and .92 and more strongly suggest a general factor.

Table 1.

Means, Standard Deviations, and Correlations for the Test Scores

Score	IIA	IIB	IM	IS	IC	CR	MBEA	CONS	INTST	TRANS	MORAL	BAL_P	SELF
Mean	3.32	3.16	3.39	3.13	3.30	3.26	1.92	38.16	36.48	3.16	3.36	2.99	3.12
SD	0.64	0.61	0.61	0.67	0.67	0.66	1.40	3.95	6.05	0.63	0.65	0.82	0.78
IIA	1.00												
IIB	0.64	1.00											
IM	0.70	0.71	1.00										
IS	0.66	0.65	0.68	1.00									
IC	0.70	0.62	0.69	0.73	1.00								
CR	0.64	0.59	0.63	0.68	0.67	1.00							
MBEA	0.09	0.17	0.07	0.12	0.07	0.22	1.00						
CONS	0.53	0.48	0.49	0.54	0.57	0.46	0.05	1.00					
INIST	0,39	0.38	0.39	0.39	0.37	0.41	0.22	0.46	1.00				
TRANS	0.56	0.53	0.54	0.55	0.56	0.51	0.08	0.54	0.38	1.00			
MORAL	0.54	0.54	0.48	0.50	0.52	0.46	0.04	0.47	0.24	0.59	1.00		
BAL_P	0.55	0.54	0.55	0.62	0.58	0.51	0.11	0.50	0.35	0.57	0.53	1.00	
SELF	0.66	0.60	0.64	0.68	0.68	0.62	0.13	0.54	0.37	0.65	0.57	0.73	1.00

Note. The following scores are composites, not factor scores as suggested by the authors for the scoring of the instruments. IIA = Idealized Influence – Attributed, IIB = Idealized Influence – Behavior, IM = Inspirational Motivation, IS = Intellectual Stimulation, IC = Individual Consideration, CR = Contingent Reward, MBEA = Management-by-Exception – Active,

CONS = Consideration, INIST = Initiation of Structure, TRANS = Relational Transparency, MORAL = Internalized Moral Perspective, BAL_P = Balanced Processing, and SELF = Self-Awareness.

N = 1,007

The entries in Table 2 show the correlations of the ALQ total (AL) score and transformational leadership (TL) score with the LBDQ consideration (CONS) score as an external criterion to be the same at .61. This is partial evidence of construct redundancy. Table 2 also shows the correlations among the composites used in the regressions. The strongest correlation was between AL and TL (r = .79), which was consistent with the meta-analytic value of .72 reported by Banks et al. (2016). Reliabilities of AL and TL were computed by the Wherry-Gaylord (Wherry & Gaylord, 1943) method to be .861 and .912, respectively. These values were then used in Spearman's correction for attenuation (unreliability) (Spearman, 1904). The result indicated a corrected correlation of .892 between AL and TL, showing little difference between the underlying constructs measured by these two questionnaires. We agree with Banks et al. (2016, p. 639) that "This finding is contradictory to scale development studies that argue that authentic and transformational leadership are empirically distinct constructs (Needier & Schriesheim, 2011; Walumbwa et al., 2008)."

Table 2

Means, Standard Deviations, and Correlations of the Composites used in the Regressions

Score	TL	TA	CONS	INIST	TRANS	MORAL	BAL_P	SELF	AL
Mean	3.26	2.62	38.16	36.48	3.16	3.36	2.99	3.12	3.17
SD	0.55	0.69	3.95	6.05	0.63	0.65	0.82	0.78	0.60
TL	1.00								
TA	0.47	1.00							
CONS	0.61	0.27	1.00						
INIST	0.44	0.36	0.46	1.00					
TRANS	0.63	0.33	0.54	0.38	1.00				
MORAL	0.60	0.27	0.47	0.24	0.59	1.00			
BAL_P	0.66	0.25	0.50	0.35	0.57	0.53	1.00		
SELF	0.76	0.42	0.54	0.37	0.65	0.57	0.75	1.00	
AL	0.79	0.41	0.61	0.40	0.86	0.79	0.83	0.99	1.00

Note. TL = Transformational Leadership, TA = Transactional Leadership. CONS = Consideration, INIST = Initiation of Structure, TRANS = Relational Transparency, MORAL = Internalized Moral Perspective, BAL_P = Balanced Processing, and SELF = Self-Awareness, and AL = Authentic Leadership.

N = 1,007

Regression Analyses

Linear regressions[1] were conducted to predict the overall ALQ score (AL) and each of the four ALQ sub-scores - relational transparency (TRANS), internalized moral perspective (MORAL), balanced processing (BAL_P), and self-awareness (SELF). To begin, each of the ALQ scores was regressed on the MLQ transformational leadership (TL) and transactional leadership (TA) scores and the LBDQ consideration (CONS) and initiation of structure (INIST) scores. The statistical significance of the regression was examined for each model and then the LBDQ scores, INIST and CONS were removed to determine their statistical effects. The statistical difference between the full and reduced regressions was determined by the usual *F* test.

As shown in Table 3, the multiple *R* for the full regression models were statistically significant for each of the ALQ scores and ranged from .618 (MORAL) to .807 (AL). The MLQ TL score was the strongest predictor for each of the ALQ scores, with correlations of .634 (TRANS), .601 (MORAL), .681 (BAL_P), .757 (SELF), and .791 (AL). Regression analyses revealed that the MLQ TL and TA scores were strong predictors. The LBDQ CONS and INIST scores contributed little to prediction with a range of differences (i.e., incremental validity over MLQ TL and TA scores) from .007 to .029.

[1] All of the correlations could be corrected for unreliability and used to estimate regressions. However, there is no generally accepted statistical test for the significance of the corrected correlations or regressions using corrected correlations. Therefore, no corrected corrections were used in the regressions.

Table 3

Regression Analyses Results

Criterion	Model	R	R Change
TRANS	TL, TA, CONS, INIST	.665	
	TL, TA	.636	.029
MORAL	TL, TA, CONS, INIST	.618	
	TL, TA	.602	.016
BAL_P	TL, TA, CONS, INIST	.674	
	TL, TA	.662	.012
SELF	TL, TA, CONS, INIST	.767	
	TL, TA	.760	.007
AL	TL. TA. CONS. INIST	.809	
	TL, TA	.792	.017

Note. All regression models were statistically significant at $p \leq .05$. All the tests of the differences between pairs of the regression equations with the same criterion were statistically significant at $p \leq .05$.
N = 1,007

Confirmatory Factor Analyses

Using LISREL 9.1 (Jöreskog & Sörbom, 2012), confirmatory factor analyses were conducted. Maximum likelihood estimation was used and, following the guidance of Brown (2005) and Hu and Bentler (1999), fit indices evaluated were CFI, GFI, RMSEA, SRMR, and Critical N. The fit criteria were CFI > .95, GFI > .90, RMSEA < .08, SRMR < .08, and Critical N > 200.

The models specified were for a single factor model (M1) and a two factor measurement model (M2), one each for the MLQ and ALQ. Initially, we planned to examine a three factor model with factors reflecting the MLQ, LBDQ, and ALQ. This could have been a hierarchical model with a just-specified (three indicators) second-level factor. However, based on results of the regression analyses, where the LBDQ provided little incremental validity for predicting AL scores when used with the MLQ, we decided to omit the LBDQ from the CFA. As a result, a hierarchical model was not tested. Both models (M1 and M2) converged quickly and yielded the fit statistics presented in Table 4. Fit for M1 was poor, whereas, all fit indices for M2 indicated a good or close fit. M1 allowed for the inspection of a first principal factor as a general factor. A general factor has to influence all of the observed variables and all in the same direction (i.e., all positive constructs are loaded in the positive direction). The general factor showed a range of loadings from .14 (MBEA) to .83 (IS, IC, and SELF). The average loading was .69 with a standard deviation of .19 and the median loading was .77.

Table 4.

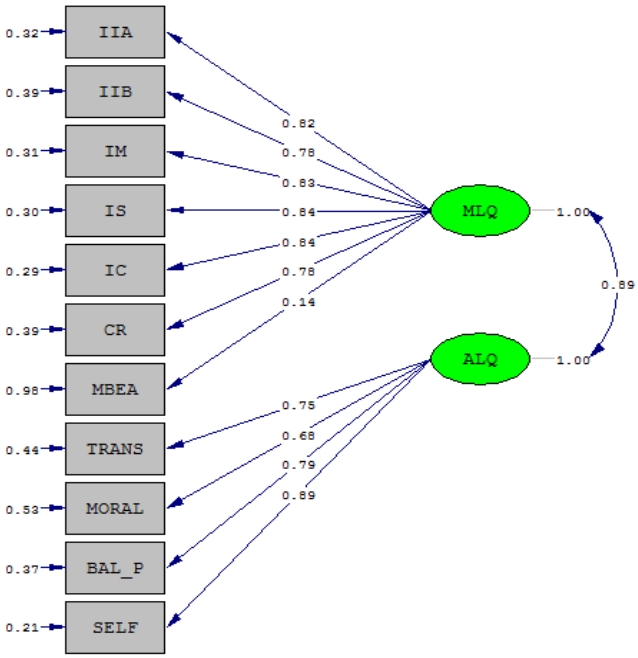
Fit Statistics for the Confirmatory Factor Analyses

	Fit Statistic				
Model	CFI	GFI	RMSEA	SRMR	Critical N
M1: One Factor	.970	.910	.102 (.094 .110)	.039	126.463
M2: Two Factor	.987	.955	.072 (.063 .079)	.028	259.100

The model on the following page shows a joint confirmatory factor analysis of the *Multidimensional Leadership Questionnaire* (MLQ) and *Authentic Leadership Questionnaire* (ALQ). The observed scores are Idealized Influence – Attributed (IIA), Idealized Influence – Behavior (IIB), Inspirational Motivation (IM), Intellectual Stimulation (IS), Individual Consideration (IC), Contingent Reward (CR), Management-by-Exception – Active (MBEA), Relational Transparency (TRANS), Internalized Moral Perspective (MORAL), Balanced Processing (BAL_P), and Self-Awareness (SELF). See figure on next page.

Figure 1.

Model of Joint Confirmatory Factor Analysis



Discussion

The results of the current analyses inspired by Le et al., (2010) and Banks et al. (2016) suggested that authentic leadership and transformational leadership, although defined as differing constructs, were empirically indistinguishable. The two constructs were highly convergent, thereby failing the test of construct uniqueness (Cronbach & Meehl, 1955). The proliferation of the same constructs with differing names hampers cumulative knowledge of organizational research and leadership.

Messick (1995) described two major threats to construct validity. The first, construct underrepresentation, occurs when "… the assessment is too narrow and fails to include important dimensions or facets of the construct." (p. 742). The second, construct contamination, occurs when "…construct-irrelevant variance [is included] that affects responses in a manner irrelevant to the interpreted construct" (p. 742). Both of these issues may affect the interpretation of scores from the ALQ.

Walumbwa et al. (2008) began their development of the authentic leadership model by considering the extant literature from multiple sources such as philosophy, social psychology, and positive psychology. After refining their model, they constructed the ALQ around four content themes and conducted a CFA of the instrument. Discounting all but hierarchical models of the four content factors, they proceeded to conduct construct validation studies. Walumbwa et al. presented the factor loadings of the 16 items for the four authentic leadership content factors. The correlations among the factors were missing, but the existence of a hierarchical factor necessitates the existence of these non-zero correlations. Meta-analytically estimated correlations among the ALQ scores (Banks et al., 2016) ranged from .84 to .92. Our current

results show that the correlations among the scores for the MLQ, LBDQ, and ALQ are similarly positive and large. Further, the correlations of the MLQ and LBDQ with the ALQ suggested that there are common sources of variance (Ree, Carretta, & Teachout, 2015) for the measures, especially between the MLQ and ALQ. Reasons for the existence of the common sources cannot be determined from the current study; however, the strength of the correlation between the MLQ and ALQ factors (.89) supports the existence of a common source. Common method variance may play some small part (Spector, 2006). Both the MLQ and LBDQ measure behavior while the ALQ measures behavior and asks the rater to infer internal states.

Inspection of the standard deviations of the subtests in the normative samples and in our sample showed our sample to be less variable. This had the effect of reducing reliability (Gulliksen, 1950, p 124, eq. 5). Considering the effects of attenuation due to unreliability (Spearman, 1904) it is likely that the correlations in our study underestimated the true population values represented by the normative samples. More accurate estimates of the population values would likely yield larger correlations.

To determine if the construct(s) measured by the ALQ are appropriately defined it would be necessary to extract general and specific factors to assess the magnitude of the influences of the factors on the scores. If the ALQ has a large general factor, then its subscale scores are measuring little of the four constructs that are unique. Similarly, if there is a large general factor across measures such as the ALQ, and MLQ, these measures provide little unique information from one another. Evaluation of the factor loadings in the CFA suggested that the ALQ measures a large general factor and that specific factors generate little variance in the score. To make the ALQ more construct valid the contribution of the general factor to the scores would

have to be reduced and the contribution of the specific factors increased. This would serve to make the ALQ a more construct-valid measure of authentic leadership.

If the ALQ has sources of construct-irrelevant variance these could be detected by additional construct-construct correlations. For example, if the ALQ is not meant to measure behavioral pathology but correlates substantially with measures of behavioral pathology, this could be a source of construct-irrelevant variance. Efforts must be made to remove any source of reliable, but construct-irrelevant variance.

The task of creating items for a questionnaire that measures more specific variance and less general variance is very difficult. Despite efforts to develop specific content based on input from subject matter experts or job/task analyses, often the content areas are related through the influence of a general factor (Ree et al., 2015). Ree et al. provided several methodological and psychological reasons which enhance the relations among the scores (produces positive manifold) and lead to the occurrence of general factors.

Construct validity is the extent to which a test measures what it is intended to measure. It is concerned with whether the test scores behave like the underlying theory predicts measures of that construct should behave. Although the ALQ content areas were developed following a review of pertinent literature, the constructs measured substantially overlap with those assessed by other widely-used measures of leadership. This does not mean per se that the AL theory lacks construct validity. However, the way content of the ALQ is specified probably causes it to lack discriminability from other leadership constructs, an aspect of external validity. This is an important component of construct validity (Messick, 1989).

Messick's (1989) united theory of construct validity proposes six aspects of construct validity – consequential, content, substantive, structural, external, and generalizability. These involve assessing the potential risks if the scores are either invalid or interpreted inappropriately (consequential) and determining whether the content measures the construct of interest (content), the formulation of the underlying construct is sound (substantive), the interrelationships among the test dimensions correlate with the construct and test scores (structural), the test has convergent, discriminant, and predictive validity (external), and the test generalizes across different groups, settings, and tasks (generalizability).

The ALQ fails the consequential test. To interpret its score as differing from measures of transformational and transactional leadership is inappropriate. The content and substantive tests can be considered together in this case. If the underlying construct is not soundly formulated then the content may be irrelevant or worse, misleading. Given the literature and the data, it is currently impossible to disentangle the content and substantive issues. Taking into consideration the correlations among the subtest scores (structural test) it is difficult to assert that each ALQ subscale measures a different construct (Murphy, 2009). The external validity and generalizability tests have been conducted only on a small scale and the results are not yet definitive.

What could have been done differently to reduce the overlap of the ALQ with other constructs, yet yield a reliable measure of authentic leadership? Its developers could have adjusted the definition and formulation of the construct. This would change the content taxonomy. Subject matter experts could have compared the ALQ content taxonomy with those for other measures of leadership to identify possible areas of overlap and examined the content of existing measures such as the LBDQ and MLQ. In addition, developers of the ALQ could

have conducted studies to examine its external validity for similar and different constructs. If it were found that the ALQ was highly redundant with other measures, the developers could repeat the process until a non-redundant measure of authentic leadership was found.

References

Banks, G. C., McCauley, K. D., Gardner, W. L., & Guler, C. E. (2016). A meta-analytic review of authentic and transformational leadership: A test for redundancy. *The Leadership Quarterly, 27*, 634-652.

Bass, B. M. (1985). *Leadership and performance beyond expectations*. NY: Free Press.

Bass, B. M. (1990). *Bass & Stogdill's handbook of leadership: Theory, research and managerial applications*. NY: Free Press.

Bass, B. M., & Avolio, B. (1995, 2000). *MLQ Multifactor Leadership Questionnaire (2nd ed.) Permission Set.* Redwood City, CA: Mind Garden, Inc.

Brown, J. D. (1996). *Testing in language programs.* Upper Saddle River, NJ: Prentice Hall Regents.

Brown, T. A. (2005). *Confirmatory factor analysis for applied research.* NY: Guilford Press.

Budescu, D. V. (1993). Dominance Analysis: A New Approach to the Problem of Relative Importance of Predictors in Multiple Regression, *Psychological Bulletin, 114*, 542-551.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52,* 281–302.

Gulliksen, H. (1950). *Theory of mental tests*. NY: Wiley.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1-55.

Jöreskog, K. G., & Sörbom, D. (2012). *LISREL 9.10 for Windows* [Computer software]. Skokie, IL: Scientific Software International, Inc.

Le, H., Schmidt, F. L., Harter, J. K., & Lauver, K. J. (2010). The problem of empirical redundancy of constructs in organizational research. *Organizational Behavior and*

Human Decision Processes, 112, 112-125.

Messick, S. (1989). Validity. In R. L. Linn (Ed.). *Educational Measurement* (3rd edition), pp. 13-103. NY: American Council of Education/Macmillan.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50,* 741–749.

Murphy, K. R. (2009). Content validation is useful for many things, but validity isn't one of them. *Industrial and Organizational Psychology: Perspectives on science and practice, 2,*453-464.

Neider, L. L., & Schriesheim, C. A. (2011). The authentic leadership inventory (ALI): Development and empirical test. *The Leadership Quarterly, 22,* 1146-1164.

Ree, M. J., Carretta, T. R., & Teachout, M. S. (2015). Pervasiveness of dominant general factors in organizational measurement. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 8,* 409-427.

Spearman, C. (1904). The proof and measurement of association of two things. *American Journal of Psychology, 15,* 72-101.

Spector, P. E. (2006). Method variance in organizational research: Truth or urban legend? *Organizational Research Methods, 9,* 221-232.

Stogdill, R. M. (1963). *Leadership Behavior Description Questionnaire - Form XII: An experimental revision* [Computer manual]. Retrieved March 1, 2012. Available from Cyfernet.

Walumbwa, F. O., Avolio, B. J., Gardner, W. L., Wernsing, T. S., & Peterson, S. J. (2008). Authentic leadership: Development and validation of a theory-based measure. *Journal of Management, 34,* 89-126.

Wherry, R. J., & Gaylord, R. H. (1943). The concept of test and item reliability in relation to factor pattern. *Psychometrika, 8, 247-264*