New wine in a new bottle: Refining the assessment of authentic leadership using Exploratory Structural Equation Modeling (ESEM)

Levesque-Côté, Julie[1], Fernet, Claude[1], Austin, Stéphanie[1], Morin, Alexandre J.S.[2]

[1] Université du Québec à Trois-Rivières, 3351 boul. des Forges. C.P. 500, Trois-Rivières (QC), Canada, G9A 5H7

[2] Substantive-Methodological Synergy Research Laboratory, Department of Psychology, Concordia University, 7141 Sherbrooke W., Montréal (QC), Canada, H3B 1R6

Julie.levesque-cote@uqtr.ca; Claude.fernet@uqtr.ca; Stephanie.austin@uqtr.ca; alexandre.morin@concordia.ca

Corresponding author: Julie Levesque-Côté (Julie.levesque-cote@uqtr.ca)

Author Note

This work was financially supported by the UQTR Research Chair on Motivation and Occupational Health and Fonds de recherche du Québec – Société et culture (FRQSC)

This is the final prepublication version of:
Levesque-Côté, J., Fernet, C., Austin, S., & Morin, A.J.S. (2017, in press). New wine in a new bottle: Refining the assessment of authentic leadership using Exploratory Structural Equation Modeling (ESEM). *Journal of Business and Psychology*. Early view. DOI: 10.1007/s10869-017-9512-y

Abstract

Purpose – To investigate the multidimensional nature of authentic leadership (AL) through the re-evaluation of the factor structure of the most commonly used scales in AL research, the *Authentic Leadership Questionnaire* (ALQ) and the *Authentic Leadership Inventory* (ALI).

Design/Methodology/Approach – Data were collected on 538 individuals working in a broad cross-section of jobs in private (57.81%) and public (42.19%) organizations.

Findings – Results from *Exploratory Structural Equation Modeling* (ESEM) provide support for the multidimensional nature of AL but also reveal excessive content overlap across subscales and an inability to properly reflect the a priori factor structure of both instruments. Further analyses enabled the identification of key items from both instruments providing a cleaner depiction of the a priori dimensionality of AL, leading to the development of an integrated optimized measure, the Authentic Leadership Integrated Questionnaire (AL-IQ).

Implications – This study proposes an alternative and optimized measure of authentic leadership that sheds light on the distinct theoretical facets of this positive leadership style. Results suggest that the AL-IQ is a useful tool to identify the strengths and weaknesses of leaders and managers' AL practices.

Originality/value – To our knowledge, this is the first study to use a combination of classical Confirmatory Factor Analyses (CFA) and newly developed ESEM framework to assess the construct validity (factor structure, reliability and criterion-related validity) of the two main AL measures: The ALQ and the ALI.

Keywords: authentic leadership; Exploratory Structural Equation Modeling (ESEM); factor structure, criterion-related validity; measurement invariance.

Inspired by Seligman and Csikszentmihalyi's (2000) work on *positive psychology* and based on the principles of *positive organizational behavior* (POB; Luthans, 2002), authentic leadership (AL) is considered to be the root construct of positive leadership (Luthans & Avolio, 2003). It refers to "*a pattern of leader behavior that draws upon and promotes both positive psychological capacities and a positive ethical climate"* (Walumbwa, Avolio, Gardner, Wernsing, & Peterson, 2008, p.94). AL theory emphasizes the importance of ethics and positive role modeling in the leader–follower relationship and proposes that leaders can help their followers achieve positive outcomes (and avoid negative ones) by going beyond the relatively limited notion of being true to oneself.

Extending this logic, Walumbwa et al. (2008) proposed a multidimensional AL framework comprising four distinct dimensions: Self-awareness, relational transparency, balanced processing, and internalized moral/ethical perspective. *Self-awareness* (SA) refers to the understanding of what one is (strengths, weaknesses, values, and beliefs) and of how one impacts other people. *Relational transparency* (RT) corresponds to the honest presentation of one's authentic self to others. *Balanced processing* (BP) represents the degree to which one is inclined to objectively analyze relevant data and explore others' opinions before coming to a decision (Avolio & Wernsing, 2008). Finally, *internalized moral/ethical perspective* (MP) refers to the degree to which individuals are guided by their core values and moral standards in their day-to-day practice instead of succumbing to group, organizational, or societal pressure. There is growing empirical evidence to support that, through these four types of behaviors, AL empower followers to better achieve individual and collective goals (Leroy, Anseel, Gardner, & Sels, 2015; Leroy, Palanski, & Simons, 2012; Walumbwa, Avolio, Gardner, Wernsing, & Peterson, 2008). AL has been associated with a variety of individual and organizational outcomes (for a review, see Gardner, Cogliser, Davis, & Dickens, 2011), including higher job satisfaction (Azanza, Leon, & Alonso, 2013; Jensen & Luthans, 2006; Walumbwa et al., 2008) and impaired psychological health (Laschinger & Fida, 2014) among followers.

Although AL is explicitly defined as a multidimensional construct (Neider & Schriesheim, 2011; Walumbwa et al., 2008), the multidimensional nature of AL remains to be empirically established (Avolio & Walumbwa, 2014; Gardner et al., 2011). Specifically, research has not yet adequately and convincingly demonstrated that current AL measures distinctly capture the four a priori dimensions of the AL construct. Given the increased popularity of the AL construct, this is worrisome and may to some extent explain the inconsistent results reported so far in this research area (Gardner et al., 2011).

The primary goal of this study was to conduct an in-depth examination of the multidimensional nature of AL. More precisely, using a combination of classical Confirmatory Factor Analyses (CFA) and the more recently developed *Exploratory Structural Equation Modeling* (ESEM) framework (Asparouhov & Muthén, 2009; Morin, Marsh, & Nagengast, 2013), the present study assesses the construct validity (factor structure, reliability, and criterion-related validity) of the two main AL measures: The *Authentic Leadership Questionnaire* (ALQ; 16 items; Walumbwa et al., 2008) and the *Authentic Leadership Inventory* (ALI; 14 items; Neider & Schriesheim, 2011).

Introducing the Contenders

Built on the same theoretical foundations, the ALQ (Walumbwa et al., 2008) and the ALI (Neider & Schriesheim, 2011) are the most commonly used measures in AL research. The ALQ was developed and validated in three studies using data from different countries (China, United States, Kenya; Walumbwa et al., 2008). Results from confirmatory factor analyses (CFA) supported the a priori four-factor structure, and showed that these four first-order factors could themselves be used to define a second-order factor representing AL. Each of the four dimensions demonstrated satisfactory scale score reliability, with Cronbach's alpha coefficients ranging from .72 to .92. Convergent validity was demonstrated by positive relations between AL perceptions and employees' satisfaction and performance. Other studies provided additional support with positive associations between AL perceptions and indicators of employees' functioning, such as work engagement (Bamford, Wong, & Laschinger, 2013; Giallonardo, Wong, & Iwasiw, 2010; Hassan & Ahmed, 2011; Vogelgesang, Leroy, & Avolio, 2013; Wong, Laschinger, & Cummings, 2010), psychological well-being (Nelson, 2014), trust (Clapp-Smith, Vogelgesang, & Avey, 2009; Hassan & Ahmed, 2011; Peus, Wesche, Streicher, Bruan, & Frey, 2012; Wang & Hsieh, 2013; Wong & Cummings, 2009; Wong & Giallonardo, 2013; Wong et al., 2010), and positive work climate (Nelson, 2014; Woolley, Caza, & Levy, 2011).

Nevertheless, despite these promising findings, some important limitations remain (e.g., Neider & Schriesheim, 2011). First, although Walumbwa et al. (2008) report that their four first-order factors "strongly correlated" (p. 100) in their first study aiming to assess the factor structure of the ALQ, they do not report the strength of these correlations, which are likely to be quite elevated. Indeed, the factor loadings of these four first-order factors on the second-order factor are themselves quite high (.62 to .78), and the zero-order correlations (likely to be deflated by the lack of control for measurement error) among these four dimensions reported in their second and third studies are also quite high (i.e., r = .54 to .69). Taken together, these observations suggest possible redundancy among the ALQ dimensions. Second, because the full measure is commercially copyrighted (only 8 sample items are publically available) access and use of the full version is restricted. Accordingly, researchers may be inclined to use the 8-item short version, for which the psychometric properties are unknown.

Neider and Schriesheim (2011) developed the ALI to address some of these limitations. Participants (undergraduate and MBA students) were asked to rate the authenticity of two politicians, Barack Obama and John McCain, during the 2008 presidential elections. The CFA results showed that the ALI factor structure differed depending on the target. For both targets, the results supported the a priori four-factor structure. However, for Obama, the results showed that these four first-order factors corresponded to a single second-order AL factor, whereas the second-order factor structure could not be supported for McCain. Scale score estimates of reliability proved satisfactory for each of the four a priori dimensions, with Cronbach's alpha coefficients ranging from .74 to .85. In addition, positive relations were found between AL and employees' satisfaction with supervisor, job satisfaction, and organizational commitment, supporting the scale's convergent validity. Additional support was provided by studies that found positive relations between AL and employees' satisfaction with the job (Cerne, Dimovski, Maric, & Penger, 2013; Rahimnia & Sharifirad, 2014) and the leader (Cerne et al., 2013) and between AL and employees' engagement (Cerne et al., 2013) or negative relations with psychological strain (Laschinger & Fida, 2014; Rahiminia & Sharifirad, 2014). Despite these encouraging results, the factor structure for the ALI differs depending on who is described (Obama or McCain), which brings the generalizability of the findings into question. Furthermore, the ALI seems to be similarly characterized by very high factor correlations (i.e., r = .59 to .89; Neider & Schriesheim, 2011), calling into question the true discriminant validity of the factors.

Revisiting the Factor Structure of the ALQ and ALI using ESEM

Although these studies provide initial support to the construct validity of the ALQ and the ALI, the fundamental multidimensional nature of AL has often been empirically neglected in later research based on these instruments. Studies also show clear inconsistencies in how AL was measured, with some studies using a unidimensional model (Fox, Gong, & Attoh, 2015; Hirst, Walumbwa, Aryee, Butarbutar, & Chen, 2016; Liu, Liao, & Wei, 2015; López, Alonso, Morales, & León, 2015; Yagil & Medler-Liraz, 2014; Zubair & Kamal, 2015), a four-factor model (Bamford et al., 2013; Edu Valsania, Moriano Leon, Molero Alonso, & Topa Cantisano, 2012; Shapira-Lishchinsky & Tsemach, 2014; Xiong & Fang, 2014), or a second-order factor model (Rego, Júnior, & Pina Cunha, 2015; Hsieh & Wang, 2015; Lyubovnikova, Legood, Turner, & Mamakouka, 2017; Rego, Sousa, Marques, & Cunha, 2014; Wang et al., 2014). A possible explanation for these discrepant practices could be the high factor correlations among the four a priori factors, which is likely to create multicollinearity, making it impossible to simultaneously consider all four dimensions in multivariate analyses. The inconsistency of the AL factor structure calls into question the ability of the ALQ and the ALI to adequately capture the four distinct dimensions of AL. Even in studies supporting a multifactor structure, the dimensions were so highly correlated as to suggest conceptual redundancy (Banks, McCauley, Gardner & Guler, 2016). This is worrisome as a key aspect of AL theory is the differential ability of these four dimensions to predict specific outcome variables (for a similar argument applied to transformational leadership, see Banks et al., 2016; Van Knippenberg & Sitkin, 2013).

This issue is important for both research and practice on AL. Whereas AL is portrayed as a multidimensional construct, there is little theorizing on the relevant underlying structure (higher-order or four first-order dimensions) and how it relates to work-related outcomes. Indeed, the ability to achieve a more comprehensive theoretical understanding of AL largely depends on the quality of available measures, and adequate multidimensional measures are required to support the investigation of differential relations between desirable work outcomes and the distinct theoretical facets of AL. Greater insight into the underlying structure of AL would also help leaders to understand better how their practices and actions impact themselves and others.

To our knowledge, all studies investigating the factor structure of AL are based on CFA (e.g., Neider & Schriesheim, 2011; Walumbwa et al., 2008). However, CFA relies on very strict independent cluster model (ICM) assumptions, which force each item to correspond to a single factor with all cross-loadings constrained to be exactly zero. The observation that many well established measures failed to meet ICM assumptions recently led Marsh and colleagues (2009, 2010) to question the realism of these restrictive expectations. Nowadays, many researchers recognize that the ICM constraints inherent in CFA are oftentimes not appropriate given the nature of the data (for a review, see Marsh, Morin, Parker, & Kaur, 2014), albeit this recognition has yet to reach mainstream research in the organizational area (Asparouhov, Muthén, & Morin, 2015). As noted by Morin, Arens, and Marsh (2016, also see Morin, Arens, Tran, & Caci, 2016), indicators are rarely, if ever, perfectly and uniquely related to a single construct, and will almost always display some degree of construct-relevant association with non-target factors assessing conceptually-related dimensions. For example, the ALQ item, "*My leader seeks feedback to improve interactions with others*," is designed to assess the SA dimension. However, because it includes a mention of interactions with others, this item may also to a lesser extent tap into RT. Similarly, because of its feedback component, this item may also tap into the BP dimension. These associations with non-target factors are fully consistent with the underlying theory and content of the item, yet forcefully ignored in ICM-CFA models. Worse, the observation of possible cross-loadings through an examination of modification indices even led Neider and Schriesheim (2011) to delete 2 items from the ALI. Similarly, the observation of residual associations among items not explained by the a priori factors led Walumbwa et al. (2008) to incorporate two potentially problematic post hoc correlated uniquenesses (Schweizer, 2012) to their model rather than to consider ways to explicitly model this form of construct-relevant psychometric multidimensionality (Morin, Arens, & Marsh, 2016).

Morin and colleagues (Morin, Arens, & Marsh, 2016; Morin, Arens, Tran, & Caci, 2016) note that construct-relevant psychometric multidimensionality occurs when indicators are associated with more than one source of true score variance. More precisely, they note that classical test theory (Nunnally & Bernstein, 1994) relies on the explicit assumption that any psychometric ratings will reflect a mixture of random measurement error (assessed as part of reliability analyses and controlled for as part of indicators' uniquenesses in latent variable models), construct-relevant true score variance (validity in relation to the a priori construct that is purported to be assessed by this indicator), and construct-irrelevant sources of true score variance (reflecting additional "true score" associations with other constructs). Thus, because this last component of items' variance still reflects true score variance, it does represents some form of validity in relation to the other constructs with which this indicator is associated. Importantly, in the assessment of AL, forcing these additional true-score associations between items and non-target factors to be zero is likely to artificially inflate the observed relations between AL dimensions. In fact, restricting cross-loadings to zero "forces" an artificial increase in the latent factor covariances in order to be able to accurately approximate the relations among items reflected in the observed variance-covariance matrix, which may explain the problematically high factor correlations reported in previous studies.

EFA, which is an important precursor of CFA and structural equation modeling (SEM; Cudeck & MacCallum, 2007), provides insightful information for measures refinement and construct validity evaluation (Conway & Huffcutt, 2003). It represents an easy way to explicitly model construct-relevant multidimensionality related to the assessment of conceptually-adjacent constructs, such as AL dimensions (Morin, Arens, & Marsh, 2016; Morin, Arens, Tran, & Caci, 2016). Unfortunately, EFA has often been criticized for being data-driven (Kahn, 2006; Preacher & MacCallum, 2003), whereas CFA is assumed to be theory-driven. However, as noted by Morin, Marsh, and Nagengast (2013: 396), this criticism relies on a semantically-driven misunderstanding of the nature of EFA (*exploratory* factor analyses) and CFA (*confirmatory* factor analyses) which:

"*still serves to camouflage the fact that the critical difference between EFA and CFA is that all cross-loadings are freely estimated in EFA. Due to this free estimation of all cross-loadings, EFA is clearly more naturally suited to exploration than CFA. However, statistically, nothing precludes the use of EFA for confirmatory purposes, except perhaps the fact that most of the advances associated with CFA/SEM were not, until recently, available with EFA.*"

In fact, the inclusion of post hoc correlated uniquenesses by Walumbwa et al. (2008), as well as the deletion of items presenting possible cross-loadings by Neider and Schriesheim (2011) clearly illustrates that nothing precludes the use of CFA for more exploratory purposes. Furthermore, target rotation makes it possible to estimate an EFA measurement model in a fully confirmatory manner, through the a priori specification of the target loadings, while cross-loadings are freely estimated but "targeted" to be as close to zero as possible (Asparouhov & Muthén, 2009; Browne, 2001).

ESEM (Asparouhov & Muthén, 2009; Marsh et al., 2014; Morin et al., 2013) represents a newly developed analytical approach allowing for the integration of EFA with CFA and SEM into a single overarching framework. In particular, although it remains possible to rely on ESEM for exploratory/inductive purposes, ESEM provides a way to circumvent restrictive ICM assumptions by allowing for the estimation of EFA factors in a fully confirmatory manner (Guay et al., 2015; Morin, Arens, & Marsh, 2016). ESEM makes it possible to implement with EFA factors most analytical possibilities typically reserved for CFA including tests of measurement invariance, predictive models, and goodness-of-fit assessment. An important advantage of ESEM is that it allows for the simultaneous consideration of all cross-loadings in a single step, whereas modification indices are calculated based on the impact of including a single cross-loading at a time (Morin & Maïano, 2011).

A common misunderstanding about EFA/ESEM is that cross-loadings are likely to change the meaning of the latent factors. This flawed criticism neglects the fact that EFA/ESEM corresponds to a reflective measurement model where the factors are assumed to influence the items, rather than the opposite. The clearest demonstration that cross-loadings do not taint the meaning of the latent factors comes from simulation studies (Asparouhov & Muthén, 2009; Sass & Schmitt, 2010; Schmitt & Sass, 2011) and studies of simulated data (Marsh, Lüdtke, Nagengast, Morin, & Von Davier, 2013; Morin, Arens, & Marsh, 2016). Results from these studies show that ESEM provides more exact estimates of true population values for factor correlations when cross-loadings (even small ones as low as .10) are present in the population model, and yet, remain unbiased when the population model corresponds to ICM-CFA assumptions, thus suggesting that the observation of reduced factor correlations associated with the ESEM, relative to CFA, solution represents evidence in favor of the ESEM solution. This led Asparouhov et al. (2015, p.1564) to note that:

"*Overall, these studies clearly show that the inclusion of cross-loadings is neither logically flawed nor logically incorrect but rather empirically supported by statistical research. Going back to the flawed argument that cross-loadings are supported by statistical constructs, these results rather show that it is the exclusion of these cross-loadings that modifies the meaning of the constructs*."

Taken together, inconsistencies in the measurement (or representations) of AL across studies, and repeated observations of high correlations between AL dimensions, suggest that the ALQ and ALI items may not fully capture the multidimensionality of the AL construct. As demonstrated in previous studies, the ESEM approach represents a flexible tool that seems to be tailor-made for a thorough examination of strongly related constructs such as AL dimensions.

Criterion-related Validity and Generalizability

An additional objective of this study was to gain a more nuanced understanding of the multidimensional nature of AL by reassessing the criterion-related validity of the ALQ and the ALI. To do so, we used job satisfaction (i.e., whether the individual is satisfied with the job overall; Fouquereau & Rioux, 2002), work performance (e.g., whether the individual performs behaviors that directly serve the organization's goals; Williams & Anderson, 1991) and psychological distress (i.e., nonspecific symptoms of impaired psychological health; Kessler et al., 2002). Growing empirical supports a positive association between AL perceptions and job satisfaction (i.e., Cerne, Dimovski, Maric, Penger, & Skerlavaj, 2013; Lashinger, Wong, & Grau, 2012; Giallonardo, Wong, & Iwasiw, 2010; Wong & Laschinger, 2013) as well as work performance (i.e., Clapp-Smith, Vogelgesang, & Avey, 2009; Leroy et al., 2015; Peus et al., 2012; Wang & Hsieh, 2013; Wong & Laschinger, 2013). Similarly, we expect AL ratings to be positively associated job satisfaction and work performance. Although no study has addressed the specific relation between AL and psychological distress, some empirical evidence indicate that AL perceptions are negatively associated with employees' ill-being outcomes including burnout (Laschinger et al., 2012; Laschinger & Fida, 2014) and stress symptoms (Rahimnia & Sharifirad, 2015). Accordingly, we expect AL ratings to be negatively associated with psychological distress.

Finally, to evaluate the generalizability of the underlying measurement structure of the retained measures, we examined their measurement invariance across meaningful subgroups of participants defined based on gender, as well as across two independent samples of employees from different work sectors (private and public organizations). In addition to representing a strong test of the generalizability of the measurement model, these tests also address the ability of the scale to be used in the context of gender- or work sector-based comparisons. It is well known that research seeking to provide an improved psychometric representation of measurement instruments need to verify the extent to which the results can generalize to new samples of participants, as we do in the present study. However, the recruitement of a new sample of participants often ends up differing from the original sample in more than one way, as is also the case in the present study (i.e., Sample 1 comes from the private sector, and Sample 2 comes from the public sector), some of which typically remain undocumented. Thus, in order to complement this test of generalizability, it is often usefull to rely on complementary tests of measurement invariance conducted across predetermined, and meaningful, subgroups of participants (Millsap, 2011; Morin, Meyer, Creusier, & Biétry, 2015). Here, given the sample size and composition, the most realistic way to conduct these additional tests were to assess possible gender-differences and similarities in responses to the questionnaires.

METHOD

Samples and Procedures

Sample 1. From May 2014 to March 2015, data were collected on 311 French-Canadian workers (private organizations sample; 64.95% females) working in a broad cross-section of jobs in the retailing and manufacturing sector (20% management, 18% customer services, 20% operations). Mean age was 37.62 years (*SD* = 13.25) and mean years of organizational tenure was 9.36 (*SD* = 9.15). In the retailing sector, we presented our study to multiple store managers and asked permission to place our questionnaires in their staff rooms along with a cover letter describing the purpose of the study. We retrieved completed questionnaires two weeks later. We followed the same procedure in the manufacturing sector, except that a human resources employee distributed the questionnaire to employees. In both cases, participants signed a voluntary consent form that assured them that their participation would remain confidential.

Sample 2. In order to systematically assess the generalizability of our results, a second sample of participants was recruited in the public sector. The decision to move to the public sector for this second sample was based on a desire to pursue a more demanding replication process across slightly different samples. Indeed, previous research suggests that perceptions and effects of leadership practices may slightly differ across the public and private sectors (e.g., Javidan & Waldman, 2003; Pawar & Eastman, 1997). For this second sample, data were collected between July and August 2016 on a total of 227 French-Canadian nurses (public organizations sample; 88.55% female). Mean age was 40.33 years (*SD*=11.55) and mean years of professional tenure was 15.73 (*SD*=11.12). As in sample 1, all participants signed a voluntary consent form that guaranteed their confidentiality.

Measures

All measures were administrated in French to both samples. For the ALI, which was not available in French, we used a classical cross-cultural adaptation method (Hambleton, & Kanjee, 1995; Vallerand & Halliwell, 1983). The items were first translated into French and then back-translated into English by independent bilingual translators. As part of the translation process, all items from the ALQ and ALI were slightly modified to ensure a better consistency across instruments. A panel of bilingual faculty members and graduate students verified the semantic correspondence between back-translated and original items and adjusted the translation, when appropriate.

Authentic leadership. The ALQ (16 items; Walumbwa et al., 2008) and the ALI (14 items; Neider & Schriesheim, 2011) were used to assess AL. For both measures, participants were asked to rate their immediate supervisor's leadership practices on a 5-point scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). The ALQ[1] includes four subscales, assessing self-awareness (4 items; α = .89; *My leader seeks feedback to improve interactions with others*), relational transparency (5 items; α =.81; *My leader is willing to admit mistakes when they are made*), internalized moral perspective (4 items; α =.85; *My leader demonstrates beliefs that are consistent with actions*) and balanced processing (3 items; α =.84; *My leader solicits views that challenge his or her deeply held positions*). The ALI also includes four subscales assessing self-awareness (3 items; α =.88; e.g., *My leader accurately describes the way that others view his/her abilities*), relational transparency (3 items; α =.79; e.g., *My leader clearly states what he/she means*), balanced processing (4 items; α =.90; e.g., *My leader objectively analyzes relevant*

[1] This questionnaire was used with the authorization of Mind Garden. Sample items from the original English instrument can be obtained from Mind Garden.

data before making a decision),and internalized moral perspective (4 items; α =.83; e.g., *My leader is guided in his/her actions by internal moral standards*).

Job satisfaction. The 5-item (α =.90) Satisfaction at Work Scale developed in French (Fouquereau & Rioux, 2002) was used. Derived from the Satisfaction With Life Scale (Diener, Emmons, Larsen, & Griffin, 1985), it assesses overall job satisfaction on a 7-point scale ranging from 1 (*do not agree at all*) to 7 (*completely agree*). A sample item is "*I am satisfied with my work*." Fouquereau and Rioux (2002) demonstrated the scale score reliability, and construct validity by positive associations with the Minnesota Satisfaction Questionnaire (Weiss, Dawis, England, & Lofquist, 1967), a measure of specific work satisfaction. Each item was used as an indicator of a latent construct of job satisfaction.

Work performance. Work performance was assessed using a 4-item self-report scale (α =.94) adapted from William and Anderson's (1991) in-role performance subscale. Participants indicated the extent to which they agreed with statements about their work performance on a 1 (*do not agree at all*) to 7 (*very strongly agree*) scale. A sample item is "*I adequately complete the tasks that are assigned to me*." Results from Fernet, Trépanier, Austin, Gagné, and Forest (2015) supported the scale score reliability and construct validity of the French adaptation of this scale by demonstrating positive associations with organizational commitment and negative associations with burnout. Each item was used as an indicator of a latent construct of work performance.

Psychological distress. Psychological distress was assessed using the French adaptation (Gravel, Connolly, & Bédard, 2003) of the K6 (Kessler et al., 2002). This 6-item instrument (α =.64) measures non-specific symptoms of anxiety and depression experienced during the previous month. Items were rated on a 5-point scale ranging from 1 (*never*) to 5 (*very often*). A sample item is "*During the previous month, I felt that everything was an effort*." Results by Arnaud et al. (2010) supported for the scale score reliability and convergent validity of this French adaptation by demonstrating positive associations with anxio-depressive symptoms. Each item was used as an indicator of a latent construct of psychological distress.

Analyses

All analyses were performed using the Robust Maximum Likelihood (MLR) estimator available in Mplus (version 7.0; Muthén & Muthén, 2012), which is robust to the non-normality of the data and well suited to the analysis of Likert-type scales including five or more answer categories such as those used in the present study (e.g., Finney, & DiStefano, 2013). Given that the objective of the second sample was to systematically test the extent to which the generalizability of the results obtained on Sample 1 would generalize to a new sample, we describe in this section the analytical strategy utilized with Sample 1 and provide additional details of analyses conducted on Sample 2 in the results section.

For comparison purposes, we first estimated a one factor model (which is equivalent for ESEM and CFA). Then, CFA models including two to four first-order factors were estimated according to ICM assumptions, where each item was only allowed to load on the factor it was assumed to measure and cross-loadings were not allowed. The a priori factors included in the AL measures included four correlated factors representing SA, RT, BP, and MP. Alternative two- and three-factor models included all possible combinations involving pairs of factors. The objective of estimating these alternative models was to more clearly ascertain the superiority of the a priori four-factor representation. In the ESEM models, the same sets of factors were represented using a confirmatory oblique target rotation (Asparouhov & Muthén, 2009; Browne, 2001; Morin, Arens, & Marsh, 2016). This rotation procedure allowed us to rely on a priori specification of the main factor loadings (corresponding the the CFA solutions), combined with a free estimation of all cross-loadings which were constrained to be as close to zero as possible (e.g. Morin, Arens, & Marsh, 2016). In second-order CFA and ESEM models, the first-order factors were specified as related to a single second-order factor. To estimate the higher-order ESEM model, we had to rely on the ESEM-within-CFA approach proposed by Morin et al. (2013; Morin, Arens, & Marsh, 2016).

The final retained model was then used to assess the criterion-related validity of the AL factors when used to predict scores on latent CFA factors representing job satisfaction, work performance, and psychological distress, which were integrated in the final model. Invariance of the final retained first-order measurement model across gender was then tested in the following sequence (Millsap, 2011): (1) *configural invariance*, (2) *weak measurement invariance*, (3) *strong measurement invariance* (invariance of loadings and intercepts), (4) *strict measurement invariance* (invariance of loadings, intercepts, and uniquenesses), (5) *invariance of the latent variance-covariance matrix* (invariance of the loadings, intercepts, uniquenesses, and latent variances-covariances), and (6) *latent mean invariance* (invariance of the loadings, intercepts, uniquenesses, latent variances-covariances, and latent means). Should the retained model be a second-order model, the invariance of the second-order structure would then be examined in a similar sequence following Morin et al. (2011; also see Cheung, 2008) recommendations, starting with a baseline first-order factor model specified as invariant according to the conclusions of steps (1) to (4) of the preceding sequence.

We relied on the following goodness-of-fit indices to describe the fit of the alternative models (Hu & Bentler, 1999; Marsh, Hau, & Grayson, 2005): The MLR chi square (χ^2), the comparative fit index (CFI), the Tucker-Lewis index (TLI), the root mean square error of approximation (RMSEA) with its 90% confidence interval. According to typical interpretation guidelines (Hu & Bentler, 1999; Marsh et al., 2005), values greater than .90 and .95 for the CFI and TLI are considered to be respectively indicative of adequate and excellent fit to the data, while values smaller than .08 or .06 for the RMSEA support respectively acceptable and excellent model fit. In tests of measurement invariance, the following guidelines were used (Chen, 2007; Cheung & Rensvold, (2002): a CFI diminution of .01 or less and a RMSEA augmentation of .015 or less between a model and the previous one indicate that the measurement invariance hypothesis should not be rejected.

RESULTS

Factor Structure: CFA versus ESEM

Given the clear superiority of the a priori four-factor measurement models in terms of goodness-of fit relative to the alternative models including one, two, or three first order factors, we only focus on these a priori models in the present manuscript (see Table 1). Note that the results of alternative models are reported in Appendix 1 of the online supplements. We first discuss the results associated with the ALQ, before moving on to the ALI.

The ALQ. As shown in Table 1, the ESEM model provides a far more satisfactory level of fit to the data than the CFA model. More specifically, the CFA solution provided satisfactory level of fit to the data according to the CFI and TLI, but not according to the RMSEA. In contrast, the corresponding ESEM solution provided a much improved, and fully satisfactory, level of fit to the data according to all indicators, supporting the need to account for construct-relevant multidimensionality in items ratings through the incorporation of cross-loadings. Parameter estimates obtained from these models (sample 1) are reported in Table 2, and further support this conclusion. Thus, the CFA solution revealed that all factors were strongly defined by their a priori items (λ = .58–.87), but factor correlations that were so high so as to call into question the discriminant validity of these four dimensions (*r* = .79 to .95) and the ability to use them simultaneously in further predictive models. In contrast, the ESEM factor correlations were much smaller (*r* = .27 to .70), supporting the superiority of this solution based on the fact that ESEM is known to result in lower factor correlations than CFA when cross-loadings are present in the population model (see Asparouhov et al., 2015). However, the ESEM solution also revealed a lack of clear representation of the underlying factor structure, with 7 out of 16 items having their primary loadings on a factor other than their a priori factor, and many items presenting relatively large cross-loadings. This suggests substantial overlap between the ALQ factors, or an improper representation of the a priori dimensions in their a priori items. Since higher-order factors are defined from the first-order factors, the presence of a problematic first-order structure made it unnecessary to investigate a second-order representation.

The ALI. The results obtained for the ALI parallel those obtained for the ALQ. As shown in Table 1, the CFA model provides an acceptable level of fit to the data according to all indicators, whereas the fit of the ESEM model supports its superiority. Parameter estimates from these solutions are reported in Table 3. The CFA solution provided a satisfactory level of fit to the data according to the CFI and TLI (and an acceptable level of fit according to the RMSEA), and revealed well-defined factors (λ =.69–.86). However, the estimated factor correlations remained problematically high (*r* =. 79 to .97). In contrast, the ESEM solution provided a much improved, and fully satisfactory, fit to the data according to all indicators and substantially reduced factor correlations (*r* = .30 to .72). Unfortunately, this solution resulted in a problematic factor structure, with 6 out of 14 items not having their primary loadings on their a priori factor, and many items presenting large cross-loadings. Once again, this unclear factor structure precluded the assessment of the second-order factor structure.

A Combined Measure: The Authentic Leadership-Integrated Questionnaire (AL-IQ)

In summary, by reassessing the factor structure of the ALQ and the ALI, we found support for the multidimensional nature of AL. More importantly, the results showed that ESEM is unequivocally superior to CFA in terms of providing a better level of fit to the data and resulting in reduced estimates of the factor correlations. However, despite these observations, the ESEM factor solutions reveal important problems in terms of correspondence between items and their a priori dimensions. We therefore decided to combine the items from the two measures to explore how they would work together and to attempt to identify a subset of items that would provide a way to achieve a more proper representation of the four underlying dimensions of the AL construct. Using ESEM, we thus estimated an additional four-factor solution, using all 30 items from the ALQ and ALI. This model provided a satisfactory level of fit to the data ($\chi^2(321) = 626.610$ ($p < .01$), CFI = .944, TLI = .924, RMSEA = .055). As shown in Table 4, a substantial number of items again had their main loadings on a factor that was not their a priori factor, and sometimes ALI and ALQ items designed to assess the same dimension even ended up loading on different factors. From this problematic solution, we sought to identify the three to four items that provided the best representation of each of the four AL dimensions. To identify these items, we used the following criteria: (a) Items that presented their highest loading on their primary a priori factor; (b) Items that had no large and unexplainable cross-loadings ($\leq$.30); (c) items that maximally maintained the content coverage of the factor, as assessed independently by two expert judges. As noted by Morin and Maïano (2011), in comparison with CFA in which such more exploratory procedures would be conducted through the use of modification indices in which the added value of a single new parameter would be tested one after the other, ESEM relies on a much more efficient procedure in which all cross-loadings are estimated at once and typically reveals issues impossible to detect with CFA. Through this procedure, we identified a series of 14 suitable items, reported in Appendix 2 of the online supplements. Of those items, 3 correspond to SA (all from the ALQ; $\alpha = .89$), 3 correspond to RT (1 from the ALQ; 2 from the ALI; $\alpha = .83$), 4 correspond to MP (all from the ALQ; $\alpha = .85$), and 4 correspond to BP (all from the ALI; $\alpha = .90$). This new integrated measure was labelled the Authentic Leadership Integrated Questionnaire (AL-IQ).

We performed the same sequence of analyses on the items forming this new instrument using data from Sample 1. The goodness-of-fit results from the a priori models are reported in Table 1. The parameter estimates from the first-order factor models are reported in Table 5. Once again, these results show the superiority of the four factors ESEM model compared to the corresponding four-factor CFA model. In this study, both CFA and ESEM models resulted in well-defined factors (CFA $\lambda = 0.72$ to .85; ESEM λ= .35 to .98), but much lower factor correlations for the ESEM ($r = .44$ to .69) relative to the CFA ($r = .67$ to .88) model. Furthermore, the ESEM solution revealed that all factors were well-defined by their a priori items, that all cross-loadings were smaller than the main loadings, and clearly smaller than those identified for the ALQ and ALI measurement models. Altogether, these results argue in favor of the ESEM representation of the data and allowed the assessment of the second-order factor structure. Results for both CFA and ESEM higher-order models (see Table 1) are similar to those obtained for the corresponding first-order model. Results showed once again, the superiority of ESEM models in every respect with no significant differences between the first- and second-order models ($\Delta\chi^2(2)$= .41, p $\geq$ .05). The AL-IQ second-order model also resulted in well-defined factors (CFA $\lambda = 0.71$ to .86; ESEM λ= .39 to .98), suggesting that both first- and second-order models can be used equivalently.

Generalizability of the Results: Measurement Invariance across Gender

The results from the gender based tests of measurement invariance conducted on the first- and second-order ESEM solutions estimated in sample 1 are reported in Table 6. These results show that all of the models provided a satisfactory level of fit to the data, and that including invariance constraints on the loadings (weak invariance), intercepts (strong invariance), uniquenesses (strict invariance), latent variance-covariance matrices and latent means never resulted in a decrease in fit exceeding the recommended guidelines. Still, statistical studies suggest that common guidelines might be too lax for tests of latent means invariance, so that much lower changes in fit indices may suggest meaningful latent mean differences (Fan & Sivo, 2009). For this reason, and because of their substantive interest, we still examined latent mean differences. This examination revealed that when women's latent means are fixed to zero for identification purposes, men's latent means (expressed in SD units) are significantly lower on perception of RT (M = -.27; SE = .13; $p < .05$), MP (M = -.44; SE = .13; $p < .01$), and BP (M = -.51; SE = .14; $p < .01$), as well as for the higher-order AL factor (M = -.48; SE = .14; $p < .01$). These results suggest that, in comparison to men, women may be more easily convinced of their leaders AL qualities such as relational transparency, moral perspective and balanced processing in their leader (Alimo-

Metcalf, 2010).

Generalizability of the Results to a New Independent Sample

The previous results suggest that the AL-IQ might provide a promising multidimensional measure of AL. However, the only way to ascertain whether the psychometric properties of this new measure (obtained through the capitalization on the specific characteristics of Sample 1 which was used in the selection of the final set of items) can be expected to generalizes to new samples of employees is to systematically assess the extent of this generalization on a new independent sample of employees. To this end, participants from Sample 2 completed the reduced set of items forming the AL-IQ. Given that our objective in this second sample was to systematically assess the psychometric properties of a newly developed instrument, the AL-IQ, further adaptation were conducted on the ALQ items in order to maximize their consistency with the ALI items, without changing the meaning of the items. It is this final set of items that is reported in the appendix. Still, the fact that modifications were brought to the wording of the items made it particularly important to ascertain the measurement invariance of the resulting measure across studies, to ensure that meaning had been preserved.

The goodness-of-fit results from the a priori models are reported in the bottom section of Table 1. These results again support the superiority of the four-factor ESEM solution compared to the corresponding four-factor CFA solution (ΔCFI = +.02; ΔTLI = +.01; ΔRMSEA = -.01). Once again, the four-factor ESEM and CFA models resulted in well-defined factors, although ESEM resulted in smaller factor correlations (ESEM: r = .55 to .80; CFA: r = .74 to .88). As in Sample 1, the results also support the adequacy of the second-order solution, suggesting that it can be used interchangeably with the first-order solution ($\Delta\chi^2(2)$= 1.68, p ≥ .05). To more systematically test the extent to which the parameter estimates obtained in Sample 1 would generalize to sample 1 (i.e., whether any observed difference reflects random sample variation or true sample differences), we also conducted tests of measurement invariance of the retained ESEM solution across employees from Sample 1 (private sector) and 2 (public sector). The results from these tests are reported in the second half of Table 6 and provide strong support for the complete measurement invariance (loadings, intercepts, uniquenesses, latent variances and covariances, latent means) of the first- and second-order measurement models across samples. The fully invariant parameter estimates from the final retained ESEM solution are reported in Table 7. For the second-order ESEM solution, the invariant second-order factor loadings were also fully satisfactory (λ = .76 to .93).

Criterion-related validity

In order to assess the criterion-related validity of the newly developed AL-IQ, we tested the associations between the AL dimensions and job satisfaction, work performance, and psychological distress. These tests were conducted across both samples of participants. Starting from the final retained model of full invariance across samples, latent CFA factors representing these three outcome variables were added to the model and specified as regressed on the AL-IQ factors across samples. In a second model, these regressions were constrained to be equal across samples. The goodness of fit of these two models is reported at the bottom of Table 6, and support the equivalence of the relations between the AL-IQ factors and the outcomes across samples. These regression coefficients are reported in the top section of Table 8. Although almost no statistically significant association could be identified when considering the first-order AL factors, the results show that all three outcomes were significantly associated with employees' levels on the second-order AL construct. More precisely, the results show that higher levels of AL (the second-order factor) significantly predicted higher levels of job satisfaction and work performance, as well as lower levels of psychological distress. In addition, employees' levels on the *BP* first-order factor also predicted higher levels of job satisfaction. Overall, these regressions were able to explain 20% of the variance in employees' levels of job satisfaction, 5% of the variance in their levels of work performance, and 7% of the variance in their levels of psychological distress. For comparison purposes, we also conducted similar analyses relying on a CFA representation of the data. The goodness-of-fit results associated with these alternative models are reported in Appendix 3 of the online supplements, whereas the associated regression coefficients are reported in the bottom section of Table 8. When looking at these alternative results, it is particularly interesting to note that, although the CFA solution revealed slightly more numerous relations involving the first-order factors (BP was found to predict higher levels of job satisfaction and lower levels of psychological distress, and MP was found to predict higher levels of work performance), the second-order association are essentially identical to those observed in the ESEM solution, and these CFA models result in very similar estimates of explained variance (20% for job satisfaction, 2% for work performance, and 6% for psychological distress).

DISCUSSION

The primary objective of this study was to examine the multidimensional nature of authentic leadership (AL) by re-assessing the construct validity of the main measures: the ALQ and the ALI. The comparison of ESEM and CFA results underscored a significant shortcoming with respect to the factor structure of these measures. Although the multidimensional nature of AL is supported by a four-factor solution for both measures, the ESEM results indicated excessive overlap of items in both scales, indicating failure to clearly capture the distinctiveness of the AL dimensions as specified in the theory. Further analysis was therefore conducted to identify the ALQ and ALI items that best captured the theoretical dimensions of AL. Based on the analysis results, an integrated measure of AL is proposed: the AL-IQ. Our results confirm that the AL-IQ shows good psychometric properties and greater coherence with the multidimensional view of AL compared with the ALQ and the ALI.

Theoretical Implications

The main contribution of this study is to provide a measure of AL that accounts for an essential tenet of the theory: The multidimensional nature of AL (Luthans & Avolio, 2003; Walumbwa et al., 2008). First, in line with previous studies (e.g., Neider & Schriesheim, 2011; Walumbwa et al., 2008), our results support the multifactor structure of AL, which in our study can be represented both as four first-order factors, and as a higher-order construct. This factorial versatility suggests that the AL-IQ can be used to represent a global score of AL (higher-order structure) or specific scores corresponding to the four AL dimensions (first-order structure) depending on the objectives of the study. Still, results regarding criterion-related validity support the value of adopting a higher-order representation of AL by showing that most relations between the AL-IQ ratings and the outcomes occurred at the second-order level. Thus, focusing only on the first-order structure might lead to erroneous conclusions regarding an apparent lack of relations between AL-IQ ratings, work performance, and psychological distress. For this reason, we recommend that future research should generally focus on the higher-order representation, or at least rely on a systemtic comparions of higher-order and first-order results.

Our results not only indicated that the ESEM solution provided a better representation of the data, but also a clearer differentiation among the multiple AL dimensions (i.e., weaker intercorrelations), whereas the CFA factor correlations were often so high as to call into question the discriminant validity of the various dimensions. These results thus underscore the relevance of using ESEM over CFA to assess multidimensional constructs when the dimensions are strongly interrelated. Due to its overly restrictive constraints (i.e., no cross-loadings on latent dimensions other than the primary one), CFA tends to results in biased, or inflated, estimates of factor correlations when cross-loadings are present in the population model, whereas ESEM remains unbiased in the presence or absence of cross-loadings (Aspourov et al., 2015). Evidence of the impact of a decision to rely on a CFA, rather than ESEM, representation of the data is illustrated by the analyses of criterion-related validity in which we contrasted results based on an ESEM, relative to CFA, representation of AL-IQ ratings. In these analyses, our results revealed a single direct relation between the AL-IQ first-order factors and the outcomes with ESEM, when compared to three such relations with CFA. Based on statistical evidence showing that ESEM remains unbiased irrespective of the presence, or absence, of cross-loadings in the population model (Asparouhov et al., 2015), these results provide a clear illustration of the risks of relying on a suboptimal measurement model. Overall, our results are in line with previous studies that demonstrated the superiority of ESEM for investigating other multidimensional constructs, including academic motivation (Guay et al., 2015) or passion (Marsh et al., 2013). We believe that this approach not only enables a deeper understanding of AL behaviors, but is also likely to help bring clarification in other areas of leadership theory and research, such as charismatic-transformational leadership theory (see Van Knippenberg & Sitkin, 2013).

Still, for purposes of future AL research , it is noteworthy that the CFA models estimated in Sample 2 on the newly developed AL-IQ did achieve a satisfactory level of fit to the data. In addition, analyses of criterion-related validity based on the second-order AL factor lead to essentially identical conclusions when based on an ESEM, relative to CFA, measurement model. This observation suggests that future research could theoretically decide to rely either on simpler CFA models or on scale scores calculated following the ICM logic underlying CFA when mainly interested in studying the global effects of AL. However, these arguably simpler procedures come at a cost in terms of inflation of the estimated factor correlations (Asparouhov et al., 2015), which were found to have only a limited impact on the estimation of the higher-order factor. Importanly, at least when focusing on work performance and psychological distress, our analyses of criterion-related validity generally supported the superiority of the higher-order representation, showing that key relations would be missed based on a first-order model. However, when job satisfaction was considered, the resulted showed that the effects of global levels of AL seemed to be mainly due to the BP first-order factor, reinforcing that at least some degree of precision would be lost should one decide to rely on CFA. A possibly better alternative would be to rely on factor scores saved from preliminary ESEM model, which, in addition to preserving the underlying structure of the model, also provide a partial control for measurement errors (Morin, Boudrias et al., 2017; Morin, Meyer et al., 2016; Skrondal & Laake, 2001).

Our results also reveal that certain items included in traditional measures of AL failed to capture the distinctiveness of the dimensions proposed by AL theory. Whereas the presence of cross-loadings is a typical feature of instruments assessing conceptually-related constructs (Morin, Arens, & Marsh, 2016; Morin, Arens, Tran, & Caci, 2016), the problems identified here are too severe to be simply controlled using a simple ESEM approach and rather suggest overlap in item content, as well as poor correspondence between items and their a priori subscales. This observation raises significant concerns about the overall ability of measures to distinguish the specific contributions of the four AL dimensions. It is for this reason that we proposed the AL-IQ, as a new integrated multidimensional measure of AL built from the most appropriate items from the ALI and ALQ. Our results demonstrate the superiority of an ESEM (versus CFA) representation of answers to the AL-IQ, supported the factor structure of this revised instrument, as well as its ability in providing a well-differentiated assessment of the four a priori AL dimensions and their correspondence with an single overarching AL construct.

Our results also fully support the complete measurement invariance of the newly created AL-IQ across gender and across two independent samples of private and public sector employees. These results thus provide strong support for the generalizability of the first- and second-order ESEM factor structure of the AL-IQ responses across independent samples and gender, denoting that the set of items retained in the development of this new instrument is equally appropriate for men and women as well as employees from the private and public sector. Our results showed consistent gender-based mean differences in AL dimensions as well as in the higher-factor structure. In our sample, women were more inclined than men to recognize RT, MP, and BP in their leader. These findings are in line with Alimo-Metcalfe's results (2010) showing that leadership skills are perceived differently by men and women. This calls for research that systematically focuses on gender differences in AL perceptions, in other work settings or culture, for which the AL-IQ would be useful.

Despite attesting to the multidimensionality of the AL construct, this study also showed that the effects of AL on work outcomes seemed to be mainly limited to the overarching AL construct, with only limited effects remaining at the dimensional level. However, these relations generalized to both samples, supporting the idea that they did not simply reflect random sampling variations. For instance, work performance and psychological distress were both found to be only related to AL at the second-order factor level. As expected, these relations were respectively positive in the prediction of work performance, and negative in the prediction of psychological distress. In contrast, the results suggest that the effects of the global AL construct on this work outcome (R^2 = .208) might in fact be entirely related to the effects of the BP first-order dimension (R^2 = .200). Balanced processing implies that leaders objectively and transparently request and analyze all relevant information before making a decision, even if this information contradicts their views (Avolio, 2005). Arguably this leadership practice is likely to increase employees' sense of involvement in the decision making process, and ultimately their overall sense of job satisfaction.This suggests that, at least in some situations, the use of AL as a higher-order construct may lead to a less comprehensive understanding of specific relations with particular outcomes. To further the validation of the AL-IQ, future studies could integrate other work-related outcomes, such as organizational commitment and work engagement, to gain a more comprehensive understanding of the distinct relations within the AL dimensions.

Practical Implications

In terms of practical implications, the present study suggests that the AL-IQ is likely to represent a useful tool for leaders and managers who wish to more accurately identify the strengths and weaknesses of their AL practices. Similarly, the AL-IQ could be used in the context of intervention programs aiming to sensitize managers to the relevance of their day-to-day AL practices and the attitudes and behaviors they wish to inspire in their followers. To illustrate, it might be more relevant for leaders to act in accordance with their personal values (i.e., moral perspective) than to be self-aware in order to gain their followers' trust. However, leaders need to be aware of who they are and what they stand for (i.e., self-awareness) in order to know their personal values. Similarly, leaders who want to correctly handle a one-on-one meeting would probably benefit from relying on the authentic self (i.e., relational transparency) due to the proximal nature of the situation. Conversely, soliciting and listening to all points of view (i.e., balanced processing) might be more relevant in a team meeting, so that everyone feels involved and heard. In other words, even though all the AL dimensions are necessary for an individual to be considered an authentic leader (Walumbwa et al., 2008), some dimensions might be more salient than others for specific situations. Therefore, managers could benefit from the contribution of the AL-IQ to the understanding and identification of the multiple facets of their AL practices in order to develop interventions that match the needs of the organization.

Limitations

Despite these theoretical and methodological contributions, some limitations must be considered when interpreting our results. First, only a small number of variables were used to test the construct validity of the AL-IQ. Therefore, to provide additional support for the construct validity of the newly developed AL-IQ, associations with a broader range of covariables should be investigated. In particular, future studies should more systematically test the construct validity using objective measures such as absenteeism, or multi-source data (especially for work performance), to avoid possible biases related to the sole reliance on self-report data. Additionally, this study used a cross-sectional design. To more thoroughly test the validity of the AL-IQ, future studies should include multiple data collection points, which would provide stronger tests of directionality of the associations between AL ratings and covariates, as well as indication of the test-retest reliability of the AL-IQ. Since this study was conducted in a sample of French-Canadians employees, the back-translated English items reported in the Appendix have not been validated. Future studies, conducted in English and other languages, would be necessary to provide empirical evidence for the cross-linguistic validity of these items. Finally, because the current study relied on two convenience samples, future research should devote additional attention to tests of the generalizability of the current results to employees from various occupations, industries, or cultures, including other French-speaking populations.

Conclusion

This study demonstrated the inability of the ALI and ALQ to adequately capture the multiple dimensions of authentic leadership. To solve this problem, we proposed an alternative and optimized measure of authentic leadership to help scholars and practitioners measure, at its fair value, each facet of this positive leadership style. Attesting to the generalizability of this new measures, our results supported the measurement invariance of its factor structure across gender and samples. Altogether, our results provide an interesting first empirical look at the AL-IQ, a combined measure of AL, that could serve as a springboard to help this new leadership construct reach its full potential. This suggests that, contrary to popular belief, two good wines bottled together can result in an improved vintage.

REFERENCES

Arnaud, B., Malet, L., Teissedre, F., Izaute, M., Moustafa, F., Geneste, Schmidt, J., Llorca, P.-M., & Brousse, G. (2010). Validity Study of Kessler's Psychological Distress Scales Conducted Among Patients Admitted to French Emergency Department for Alcohol Consumption–Related Disorders. *Alcoholism: Clinical and Experimental Research, 34*(7),1235-1245. DOI:10.1111/j.1530-0277.2010.01201.x

Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal, 16*(3), 397-438. DOI: 10.1080/10705510903008204

Asparouhov, T., Muthén, B., & Morin, A.J.S. (2015). Bayesian Structural equation modeling with cross-loadings and residual covariances: Comments on Stromeyer et al. *Journal of Management, 41*, 1561-1577. DOI:10.1177/0149206315591075

Avolio, B. J. (2005). *Leadership development in balance: Made/born*. Mahwah, NJ: Lawrence Earlbaum Associates.

Avolio, B. J., & Wernsing, T. S. (2008). Practicing authentic leadership. In S. J. Lopez (Ed.), *Positive psychology: Exploring the best in people* (pp.147-165). Westport, CT: Greenwood Publishing Group.

Azanza, G., León, J. A., & Alonso, F. (2013). Authentic leadership and organizational culture as drivers of employees job satisfaction. *Journal of work and organizational psychology, 29*(2), 45-50. DOI:10.5093/tr2013a7

Bamford, M., Wong, C. A., & Laschinger, H. K. (2013). The influence of authentic leadership and areas of worklife on work engagement of registered nurses. *Journal of Nursing Management, 21*(3), 529-540. DOI:10.1111/j.1365-2834.2012.01399.x

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 Qaurterly, 27*, 634-652. DOI:10.1016/j.leaqua.2016.02.006

Beauducel, A., & Herzberg, P. Y. (2006). On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. *Structural equation modeling, 13*(2), 186-203. DOI:10.1207/s15328007sem1302_2

Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research, 36*, 111–150.

Cerne, M., Dimovski, V., Maric, M., & Penger, S., & Skerlavaj, M. (2013). Congruence of leader self-perceptions and follower perceptions of authentic leadership: Understanding what authentic leadership is and how it enhance employees' job satisfaction. *Australian Journal of Management, 39*(3), 1-19. DOI:10.1177/0312896213503665.

Chen, F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14*(3), 464-504. DOI:10.1080/10705510701301834

Cheung, G. W. (2008). Testing equivalence in the structure, means, and variances of higher-order constructs with structural equation modeling. *Organizational Research Methods, 11*(3), 593-613. DOI:10.1177/1094428106298973

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*(2), 233-255. DOI:10.1207/S15328007SEM0902_5

Clapp-Smith, R., Vogelgesang, G., & Avey, J. (2009). Authentic leadership and positive psychological capital: The mediating role of trust at the group level of analysis. *Organizational Studies, 15*(3), 227-240. DOI:10.1177/1548051808326596

Diener, E. D., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of personality assessment, 49*(1), 71-75. DOI:10.1207/s15327752jpa4901_13

Edu Valsania, S., Moriano Leon, J. A., Molero Alonso, F., & Topa Cantisano, G. (2012). Authentic leadership and its effect on organizational citizenship behaviours. *Psicothema, 24*(4), 561-566.

Fan, X., & Sivo, S. A. (2009). Using goodness-of-fit indexes in assessing mean structure invariance. *Structural Equation Modeling, 16*(1), 54–69. DOI:10.1080/10705510802561311

Fernet, C., Trépanier, S.-G., Austin, S., Gagné, M., & Forest, J. (2015). Transformational leadership and optimal functioning at work: On the mediating role of employees' perceived job characteristics and motivation. *Work & Stress, 29*(1), 11-31. DOI:10.1080/02678373.2014.1003998

Finney, S.J., & DiStefano, C. (2013). Non-normal and categorical data in structural equation modeling. In G.R. Hancock & R.O. Mueller (Eds), *Structural Equation Modeling: A Second Course, 2nd* *edition* (pp. 439-492). Greenwich, CO: IAP.

Ford, K. J., MacCallum, R. C., & Tait, M. (1986). The application of exploratory factor analysis in applied psychology: A critical review and analysis. *Personnel Psychology, 39*, 291-314. DOI:10.1111/j.1744-6570.1986.tb00583.x

Fouquereau, E., & Rioux, L. (2002). Élaboration de l'Échelle de satisfaction de vie professionnelle (ÉSVP) en langue française: Une démarche exploratoire. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement, 34*(3), 210-215. DOI:10.1037/h0087173.

Fox, J., Gong, T., & Attoh, P. (2015). The Impact of Principal as Authentic Leader on Teacher Trust in the K-12 Educational Context. *Journal of Leadership Studies, 8*(4), 6-18. DOI:10.1002/jls.21341

Gardner, W. L., Cogliser, C. C., Davis, K. M., & Dickens, M. P. (2011). Authentic leadership: A review of literature and research agenda. *The Leadership Quarterly, 22*(6), 1120-1145. DOI:10.1016/j.leaqua.2011.09.007

Giallonardo, L. M., Wong, C. A., & Iwasiw, C. L. (2010). Authentic leadership of preceptors: predictor of new graduate nurses' work engagement and job satisfaction. *Journal of Nursing Management, 18*, 993-1003. DOI:10.1111/j.1365-2834.2010.01126.x

Gorsuch, R. L. (1983). *Factor analysis*. Hillsdale, NJ: Erlbaum.

Gravel, R., Connolly, D., & Bédard, M. (2003). Enquête sur la santé dans les collectivités canadiennes (ESCC): santé mentale et bien-être. Retrived from http://www.statcan.gc.ca/cgibin/imdb/p2SV_f.pl?Function=getSurvey&SDDS=501&lang=fr&db=IMDB&dbg=f&adm=8&dis=2.

Guay, F., Morin, A. J., Litalien, D., Valois, P., & Vallerand, R. J. (2015). Application of Exploratory Structural Equation Modeling to Evaluate the Academic Motivation Scale. *The Journal of Experimental Education, 83*(1), 51-82. DOI:10.1080/00220973.2013.876231

Hambleton, R. K., & Kanjee, A. (1995). Increasing the validity of cross-cultural assessments: Use of improved methods for test adaptations. *European Journal of Psychological Assessment, 11*(3), 147-157. DOI:10.1027/1015-5759.11.3.147

Hassan, A., & Ahmed, F. (2011). Authentic leadership, trust and work engagement. *International Journal of Human Sciences, 6*(3), 164-170.

Hirst, G., Walumbwa, F. O., Aryee, S., Butarbutar, I., & Chen, C. J. H. (2016). A multi-level investigation of authentic leadership as an antecedent of helping behavior. *Journal of Business Ethic,* 139(3), 485-499. DOI:10.1007/s10551-015-2580-x.

Hsieh, C.-C., & Wang, D.-S. (2015). Does supervisor-perceived authentic leadership influence employee work engagement through employee-perceived authentic leadership and employee trust? *The International Journal of Human Resource Management, 26*(18), 1-20. DOI:10.1080/09585192.2015.1025234.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1-55. DOI:10.1080/10705519909540118

Javidan, M., & Waldman, D. A. (2003). Exploring charismatic leadership in the public sector: Measurement and consequences. *Public Administration Review, 63*(2), 229-242. DOI:10.1111/1540-6210.00282

Jensen, S. M., & Luthans, F. (2006). Relationship between entrepreneurs' psychological capital and their authentic leadership. *Journal of Managerial Issues, 18*(2), 254-273.

Kahn, J. (2006). Factor analysis in counselling psychology research, training, and practice: Principles, advances, and applications. *The Counseling Psychologist, 34*(5), 684-718. DOI:10.1177/0011000006286347

Kessler, R. C., Andrews, G., Colpe, L. J., Hiripi, E., Mroczek, D. K., Normand, S. L. T., Walters, E. E., & Zaslavsky, A. M. (2002). Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychological Medicine, 32*(6), 959-976. DOI:10.1017/s0033291702006074

Laschinger, H. K., & Fida, R. (2014). A time-lagged analysis of the effect of authentic leadership on workplace bullying, burnout, and occupational turnover intentions. *European Journal of Work and Organizational Psychology, 23*(5), 739-753. DOI:10.1080/1359432x.2013.804646

Laschinger, H. K., Wong, C. A., & Grau, A. L. (2012). The influence of authentic leadership on newly graduated nurses' experience of workplace bullying, burnout and retention outcomes: A cross-sectional study. *International Journal of Nursing Studies, 49*, 1266-1276. DOI: 10.1016/j.ijnurstu.2012.05.012

Leroy, H., Anseel, F., Gardner, W. L., & Sels, L. (2015). Authentic Leadership, Authentic Followership, Basic Need Satisfaction, and Work Role Performance: A Cross-Level Study. *Journal of Management*, *41*(6), 1677-1697. DOI:10.1177/0149206312457822

Leroy, H., Palanski, M. E., & Simons, T. (2012). Authentic leadership and behavioral integrity as drivers of follower commitment and performance. *Journal of Business Ethics, 107*(3), 255-264. DOI:10.1177/0149206312457822

Liu, S.-m., Liao, J.-q., & Wei, H. (2015). Authentic Leadership and Whistleblowing: Mediating Roles of Psychological Safety and Personal Identification. *Journal of Business Ethics*, *131*(1):107-119. DOI:10.1007/s10551-014-2271-z.

López, C. M. G.-G., Alonso, F. M., Morales, M. C, & León, J. A. (2015). Authentic leadership, group cohesion and group identification in security and emergency teams. *Psicothema, 27*(1), 59-64. DOI:10.7334/psicothema2014.161

Lubke, G. H., & Muthén, B. O. (2004). Applying multigroup confirmatory factor models for continuous outcomes to Likert scale data complicates meaningful group comparisons. *Structural equation modeling, 11*(4), 514-534. DOI:10.1207/s15328007sem1104_2

Luthans, F. (2002). The need for and meaning of positive organizational behavior. *Journal of Organizational Behavior, 23*(6), 695-706. DOI:10.1002/job.165

Lyubovnikova, J., Legood, A., Turner, N., & Mamakouka, A. (2017). How authentic leadership influences team performance: The mediating role of team reflexivity. *Journal of Business Ethics, 141*(1), 59-70. DOI:10.1007/s10551-015-2692-3

Marsh, H.W., Hau, K., & Grayson, D. (2005). Goodness of fit in structural equation models. In A. Maydeu-Olivares & J.J. McArdle (Eds.), *Contemporary psychometrics* (pp. 275-340). Mahwah, NJ: Erlbaum.

Marsh, H.W., Liem, G.A.D., Martin, A.J., Morin, A.J.S., & Nagengast, B. (2011). Methodological measurement fruitfulness of exploratory structural equation model: New approaches to issues in motivation and engagement. *Journal of Psychoeducational Assessment, 29*(4), 322-346. DOI:10.1177/0734282911406657

Marsh, H., Lüdtke, O., Muthén, B., Asparouhov, T., Morin, A., Trautwein, U., & Nagengast, B. (2010). A new look at the big five factor structure through exploratory structural equation modeling. *Psychological Assessment, 22*(3), 471–491. DOI:10.1037/a0019227

Marsh, H., Lüdtke, O., Nagengast, B., Morin, A., & Von Davier, M. (2013). Why item parcels are (almost) never appropriate: Two wrongs do not make a right—camouflaging misspecification with item parcels in CFA models. *Psychological Methods, 18*(3), 257-284. DOI:10.1037/a0032773

Marsh, H. W., Muthén, B., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A. J., & Trautwein, U. (2009). Exploratory structural equation modeling, integrating CFA and EFA: Application to students' evaluations of university teaching. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*(3), 439-476. DOI:10.1080/10705510903008220

Marsh, H., Morin, A., Parker, P., & Kaur, G. (2014). Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annual Review of Clinical Psychology*, *10*, 85–110. DOI:10.1146/annurev-clinpsy-032813-153700

Marsh, H.W., Nagengast, B., Morin, A.J.S., Parada, R.H., Craven, R.G., & Hamilton, L.R. (2011). Construct validity of the multidimensional structure of bullying and victimization: An application of exploratory structural equation modeling. *Journal of Educational Psychology*, *103*(3), 701-732. DOI:10.1037/a0024122

Marsh, H. W., Vallerand, R. J., Lafrenière, M.-A. K., Parker, P., Morin, A. J. S., Carbonneau, N., Jowett, S., Bureau, J. S., Fernet, C., Guay, F., Abduljabbar, A. S., & Paquet, Y. (2013). Passion: Does One Scale Fit All? Construct Validity of Two-Factor Passion Scale and Psychometric Invariance Over Different Activities and Languages. *Psychological Assessment*. *25*(3), 796-809. DOI:10.1037/a0032573.

Meier, L. L., Semmer, N. K., & Gross, S. (2014). The effect of conflict at work on well-being: Depressive symptoms as a vulnerability factor. *Work & Stress. 28*(1), 31-48. DOI:10.1080/02678373.2013.876691

Millsap, R.E. (2011). *Statistical approaches to measurement invariance*. New York: Routledge.

Morin, A. J. S., Arens, A., & Marsh, H. (2016). A bifactor exploratory structural equation modeling framework for the identification of distinct sources of construct-relevant psychometric multidimensionality. *Structural Equation Modeling*, *23*(1), 16-139. DOI:10.1080/10705511.2014.961800

Morin, A. J. S., Arens, K., Tran, A., & Caci, H. (2016). Exploring sources of construct-relevant multidimensionality in psychiatric measurement: a tutorial and illustration using the Composite Scale of Morningness. *International Journal of Methods in Psychiatric Research,* 25(4), 277-288. DOI:10.1002/mpr.1485

Morin, A. J. S., Boudrias, J.-S., Marsh, H. W., McInerney, D. M., Dagenais-Desmarais, V., Madore, I., & Litalien, D. (2017). Complementary variable- and person-centered approaches to the dimensionality of psychometric constructs: Application to psychological wellbeing at work. *Journal of Business and Psychology, 32*(4), 395-419. DOI:10.1007/s10869-016-9448-7

Morin, A. J. S., & Maïano, C. (2011). Cross-validation of the short form of the physical self-inventory (PSI-S) using exploratory structural equation modeling (ESEM). *Psychology of Sport & Exercise, 12*(5), 540-554. DOI:10.1016/j.psychsport.2011.04.003

Morin, A. J. S., Marsh, H., Nagengast, B. (2013). Exploratory structural equation modeling: an introduction. In GR Hancock & RO Mueller (Eds.), *Structural Equation Modeling: A Second Course*, 2nd Edition (pp. 395–436). Greenwich, CT: IAP.

Morin, A. J. S., Meyer, J. P., Creusier, J., & Biétry, F. (2016). Multiple-group analysis of similarity in latent profile solutions. *Organizational Research Methods, 19*(2), 231-254. DOI:10.1177/1094428115621148

Morin, A., Moullec, G., Maïano, C., Layet, L., Just, J.-L., & Ninot, G. (2011). Psychometric properties of the Center for Epidemiologic Studies Depression Scale (CES-D) in French clinical and nonclinical adults. *Revue d'épidémiologie et de santé publique, 59*(5), 327-340. DOI:10.1016/j.respe.2011.03.061

Muthén, L. K., & Muthén, B. O. (2012). *Mplus user's guide*. Los Angeles, CA: Muthén & Muthén.

Neider, L. L., & Schriesheim, C. A. (2011). The authentic leadership inventory (ALI): Developement and empirical tests. *The Leadership Quarterly, 22*(6), 1146-1164. DOI:10.1016/j.leaqua.2011.09.008

Nelson, K. (2014). Le leadership authentique: validation d'une mesure et étude de ses effets sur le climat de travail et le bien-être des personnes. Doctoral thesis, Université de Montréal, Canada.

Pawar, B. S., & Eastman, K. K. (1997). The nature and implications of contextual influences on transformational leadership: A conceptual examination. *Academy of Management Review, 22*(1), 80-109. DOI: 10.5465/AMR.1997.9707180260

Peus, C., Wesche, J. S., Streicher, B., Bruan, S., & Frey, D. (2012). Authentic leadership: An empirical test of its antecedents, consequences, and mediating mechanisms. *Journal of Business Ethics, 107*(3), 331-348. DOI:10.1007/s10551-011-1042-3

Preacher, K., & MacCallum, R. (2003). Repairing Tom Swift's electric factor analysis machine. *Understanding Statistics, 2*(1), 13-43. DOI:10.3758/BRM.40.3.879

Rahimnia, F., & Sharifirad, M. S. (2014). Authentic leadership and employee well-being: The mediating role of attachment insecurity. *Journal of Business Ethics*, *132*(2), 363-377. DOI:10.1007/s10551-014-2318-1.

Rego, A., Júnior, D. R., & Pina e Cunha, M. (2015). Authentic leaders promoting store performance: The mediating roles of virtuousness and potency. *Journal of Business Ethics, 128*(3), 617-634. DOI:10.1007/s10551-014-2125-8

Rego, A., Sousa, F., Marques, C., & e Cunha, M. P. (2014). Hope and positive affect mediating the authentic leadership and creativity relationship. *Journal of Business Research, 67*(2), 200-210. DOI:10.1016/j.jbusres.2012.10.003

Sass, D., & Schmitt, T. (2010). A comparative investigation of rotation criteria within exploratory factor analysis. *Multivariate Behavioral Research,* 45, 73-103. DOI:10.1080/00273170903504810

Schmitt, T., & Sass, D. (2011). Rotation criteria and hypothesis testing for exploratory factor analysis: implications for factor pattern loadings and interfactor correlations. *Educational & Psychological Measurement*, *71*, 95-113. DOI:10.1177/0013164410387348

Schweizer, K. (2012). On correlated errors. *European Journal of Psychological Assessment, 28*(1), 1-2. DOI:10.1027/1015-5759/a000094

Seligman, M. E., & Csikszentmihalyi, M. (2000). Positive psychology: An introduction. *American Psychological Association*, *55*(1), 5-14. DOI:10.1037//0003-066X.55.1.5

Shapira-Lishchinsky, O., & Tsemach, S. (2014). Psychological empowerment as a mediator between teachers' perceptions of authentic leadership and their withdrawal and citizenship behaviors. *Educational Administration Quarterly*, *50*(14), 675-712. DOI:10.1177/0013161X13513898.

Skrondal, A., & Laake, P. (2001). Regression among factor scores. *Psychometrika, 66,* 563–576.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate analysis.* Boston: Pearson Education.

Vallerand, R. J., & Halliwell, W. (1983). Toward a methodology of cross-cultural validation of psychological questionnaire. *Canadian Journal of Applied Sport Sciences, 8*, 9-18.

Van Knippenberg, D., & Sitkin, S. B. (2013). A Critical Assessment of Charismatic—Transformational Leadership Research: Back to Drawing Board? *The Academy of Management Annals*, *7*(1), 1-60. DOI:10.1080/19416520.2013.759433

Vogelgesang, G. R., Leroy, H., & Avolio, B. J. (2013). The mediating effects of leader integrity with transparency in communication and work engagement/performance. *The Leadership Quarterly, 24*(3), 405-413. DOI:10.1016/j.leaqua.2013.01.004

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*(1), 89-126. DOI:10.1177/0149206307308913

Wang, D.-S., & Hsieh, C.-C. (2013). The effect of authentic leadership on employee trust and employee engagement. *Social Behavior and Personality*, *41*(4), 613-624. DOI:10.2224/sbp.2013.41.4.613

Wang, H., Sui, Y., Luthans, F., Wang, D. S., & Wu, Y. (2014). Impact of authentic leadership on performance: Role of followers' positive psychological capital and relational processes. *Journal of Organizational Behavior*, *35*(1), 5-21. DOI:10.1002/job.1850

Weiss, D. J., Dawis, R. V., & England, G. W. (1967). Manual for the Minnesota Satisfaction Questionnaire. *Minnesota studies in vocational rehabilitation*. Industrial Relations Center, University of Minnesota, MN.

Williams, L. J., & Anderson, S. E. (1991). Job satisfaction and organizational commitment as predictors of organizational citizenship and in-role behaviors. *Journal of Management*, *17*(3), 601-617. DOI:10.1177/014920639101700305

Wong, C. A., & Cummings, G. G. (2009). The influence of authentic leadership behavior on trust and work outcomes of health care staff. *Journal of Leadership Studies*, *3*(2), 6-23. DOI:10.1002/jls.20104

Wong, C. A., & Giallonardo, L. M. (2013). Authentic leadership and nurse-assessed adverse patient outcomes. *Journal of Nursing Management, 21*(5), 740-752. DOI:10.1111/jonm.12075

Wong, C. A., & Laschinger, H. K. (2013). Authentic leadership, performance, and job satisfaction: the mediating role of empowerment. *Journal of Advanced Nursing, 69*(4), 947-959. DOI:10.1111/j.1365-2648.2012.06089.x.

Wong, C. A., Laschinger, H. K. S., & Cummings, G. G. (2010). Authentic leadership and nurses' voice behaviour and perceptions of care quality. *Journal of Nursing Management, 18*(8), 889-900. DOI:10.1111/j.1365-2834.2010.01113.x

Woolley, L., Caza, A., & Levy, L. (2011). Authentic leadership and follower development: Psychological capital, positive work climate, and gender. *Journal of Leadership and Organizational Studies*, *18*(4), 438-448. DOI:10.1177/1548051810382013

Xiong, H.-B., & Fang, P. (2014). Authentic leadership, collective efficacy, and group performance: An empirical study in China. *Social Behavior and Personality, 42*(6), 921-932. DOI:10.2224/sbp.2014.42.6.921

Yagil, D., & Medler-Liraz, H. (2014). Feel free, be yourself: Authentic leadership, emotional expression, and employee authenticity. *Journal of Leadership & Organizational Studies, 21*(1), 59-70. DOI:10.1177/1548051813483833.

Yu, C.Y. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes*. Los Angeles, CA: University of California.

Zubair, A., & Kamal, A. (2015). Authentic Leadership and Creativity: Mediating Role of Work-Related Flow and Psychological Capital. *Journal of Behavioural Sciences*, *25*(1), 150-171.

Table 1

Alternative Measurement Models Estimated in the Current Study

Model description	MLR χ^2	*df*	CFI	TLI	RMSEA and 90% CI
Sample 1: ALQ Measurement Models					
First-Order Models					
CFA	297.904*****	98	.920	.902	.081 (.070–.091)
ESEM	124.747*****	62	.975	.951	.057 (.042–.071)
Sample 1: ALI Measurement Models					
First-Order Models					
CFA	177.722*****	71	.946	.931	.070 (.057–.082)
ESEM	72.900*****	41	.984	.964	.050 (.031–.068)
Sample 1: AL-IQ measurement models					
First-Order Models					
CFA	195.776*****	71	.941	.924	.075 (.063–.088)
ESEM	81.833*****	41	.981	.957	.057 (.038–.074)
Second-Order Models					
CFA	206.560*****	73	.937	.921	.077 (.064–.089)
ESEM	81.309*****	43	.982	.962	.054 (.035–.071)
Sample 2: AL-IQ measurement models					
First-Order Models					
CFA	144.240*	71	.962	.951	.067 (.052-.083)
ESEM	70.126*	41	.985	.966	.056 (.032-.078)
Second-Order Models					
CFA	161.055*	73	.954	.943	.073 (.058-.088)
ESEM	69.825*	43	.986	.970	.052 (.028-.074)

Notes. ***** $p < .01$; ALQ = Authentic Leadership Questionnaire; ALI: Authentic Leadership Inventory; AL-IQ= Authentic Leadership Integrated Questionnaire; CFA = confirmatory factor analysis; ESEM = exploratory structural equation model; MLR χ^2 = Robust chi square test of exact fit; df = degrees of freedom; CFI = comparative fit index; TLI = Tucker–Lewis index; RMSEA = root-mean-square error approximation; CI = confidence interval.

Table 2

Standardized Parameter Estimates for the Authentic Leadership Questionnaire (ALQ) Measurement Models (Sample 1)

	CFA					ESEM				
Items	F1 (λ)	F2 (λ)	F3 (λ)	F4 (λ)	δ	F1 (λ)	F2 (λ)	F3 (λ)	F4 (λ)	δ
ALQ_SA1	.835**				.304**	**.510****	-.049	.113	.393**	.252**
ALQ_SA2	.729**				.468**	**.899****	-.224	.116	-.053	.231*
ALQ_SA3	.838**				.298**	**.588****	.320**	-.003	.120	.279**
ALQ_SA4	.833**				.305**	**.496***	.349**	.022	.176	.294**
ALQ_RT1		.583**			.660**	.137	**.368****	.258*	-.006	.596**
ALQ_RT2		.785**			.384**	.045	.239	.251*	**.485****	.359**
ALQ_RT3		.775**			.400**	.231*	.125	.140	**.491****	.357**
ALQ_RT4		.641**			.589**	.092	-.296*	**.380****	.047	.545**
ALQ_RT5		.576**			.668**	.096	.137	**.494****	-.035	.605**
ALQ_MP1			.807**		.349**	.048	.316*	**.610****	-.055	.320**
ALQ_MP2			.826**		.318**	-.034	.284**	**.748****	-.087	.253**
ALQ_MP3			.742**		.450**	-.022	-.152	**.759****	.253	.326**
ALQ_MP4			.709**		.497**	**.421****	.127	.361*	-.124	.468**
ALQ_BP1				.694**	.519**	**.494****	-.119	.207*	.199	.452**
ALQ_BP2				.871**	.241**	**.307****	.355	.012	.445	.229**
ALQ_BP3				.867**	.249**	.236*	.139	.121	**.581****	.222**
Standardized Factor Correlations										
F1	-					-				
F2	.933**	-				.421**	-			
F3	.824**	.899**	-			.700**	.495**	-		
F4	.950**	.951**	.791**	-		.582**	.274**	.451**	-	

Notes. * $p < .05$; ** $p < .01$; CFA = confirmatory factor analysis; ESEM = exploratory structural equation model; F1-F4: Factor 1 to 4; λ = standardized factor loading; δ = Standardized item uniquenesses; SA = Self-Awareness items; BP = Balanced Processing items; RT =Relational Transparency items; MP = Moral/Ethical Perspective items. Main factor loadings are indicated in bold.

Table 3

Standardized Parameter Estimates for the Authentic Leadership Inventory (ALI) Measurement Models (Sample 1)

	CFA					ESEM				
Items	F1 (λ)	F2 (λ)	F3 (λ)	F4 (λ)	δ	F1 (λ)	F2 (λ)	F3 (λ)	F4 (λ)	δ
ALI_SA2	.777**				.396**	.235*	.205	**.320****	.212**	.388**
ALI_SA4	.798**				.363**	.155	.184	**.334***	.282*	.376**
ALI_SA5	.788**				.379**	.245	**.305****	-.055	.469	.359**
ALI_RT1		.744**			.446**	-.027	**.692****	.125	-.052	.451**
ALI_RT2		.718**			.484**	.221**	**.376****	.129	.162**	.485**
ALI_RT3		.812**			.341**	-.035	**1.043****	-.061	-.119	.168
ALI_MP1			.692**		.522**	.006	.047	**.702****	-.028	.482**
ALI_MP2			.712**		.493**	-.059	-.132	**.959****	-.070	.346**
ALI_MP3			.700**		.511**	-.009	.205	**.437****	.107	.540**
ALI_MP4			.846**		.285**	.035	.155	**.582****	.121	.352**
ALI_BP1				.797**	.365**	.451*	-.042	.091	**.522***	.290**
ALI_BP2				.855**	.269**	.235*	.049	.142	.566	.314**
ALI_BP3				.851**	.276**	**-.376***	-.028	-.032	1.272	-.312
ALI_BP4				.826**	.319**	**.470***	.081	.064	.468**	.236**
Standardized Factor Correlations										
F1	-					-				
F2	.891**	-				.302*	-			
F3	.865**	.813**	-			.380**	.717**	-		
F4	.973**	.785**	.786**	-		.380*	.706**	.655**	-	

Notes. * $p < .05$; ** $p < .01$; CFA = confirmatory factor analysis; ESEM = exploratory structural equation model; F1-F4: Factor 1 to 4; λ = standardized factor loading; δ = Standardized item uniquenesses; SA = Self-Awareness items; BP = Balanced Processing items; RT =Relational Transparency items; MP = Moral/Ethical Perspective items. Main factor loadings are indicated in bold.

Table 4

Standardized Parameter Estimates from the ESEM Model Simultaneously Conducted on the Authentic Leadership Questionnaire (ALQ) and Authentic Leadership Inventory (ALI) (Sample 1)

Item	F1 (λ)	F2 (λ)	F3 (λ)	F4 (λ)	δ
ALQ_SA1	**.462****	.149*	.009	.440**	.277**
ALQ_SA2	**.352****	.197	.144	.224*	.482**
ALQ_SA3	.248**	**.373****	.165	.241**	.320**
ALQ_SA4	**.342****	.317**	.170	.199**	.337**
ALQ_RT1	.172*	**.739****	.097	-.295**	.445**
ALQ_RT2	.155*	.134	.232**	**.446****	.369**
ALQ_RT3	.349*	.156	.041	**.460****	.349**
ALQ_RT4	.161	.049	**.544****	-.001	.548**
ALQ_RT5	.307**	.207	**.393****	-.203*	.575**
ALQ_MP1	.215*	.097	**.716****	-.166**	.320**
ALQ_MP2	.048	-.040	**.939****	-.125	.245**
ALQ_MP3	.290*	-.082	**.472****	.205**	.455**
ALQ_MP4	.249**	.213*	**.413****	-.061	.508**
ALQ_BP1	**.462****	.041	.099	.306**	.462**
ALQ_BP2	.223**	.270*	.112	**.450****	.272**
ALQ_BP3	.441**	.114	.048	**.474****	.250**
ALI_SA2	-.033	.268*	.268	**.363****	.414**
ALI_SA4	-.013	.317**	.221*	**.389****	.360**
ALI_SA5	-.046	.255*	.115	**.549****	.371**
ALI_RT1	-.034	**.983****	-.132	-.086	.306**
ALI_RT2	.113	**.305***	.157	.287**	.495**
ALI_RT3	.108	**.794****	.080	.003	.325**
ALI_MP1	-.036	.159	**.584****	-.005	.522**
ALI_MP2	-.172	.015	**.756****	.020	.491**
ALI_MP3	-.184	.157	**.523****	.162*	.528**
ALI_MP4	-.215	.008	**.794****	.198*	.291**
ALI_BP1	.211*	-.011	.164*	**.609****	.327**
ALI_BP2	.009	.056	.227**	**.653****	.282**
ALI_BP3	-.144*	.274**	.144	**.631****	.236**
ALI_BP4	.120	.025	.137	**.688****	.279**
Standardized Factor Correlations					
F1	-				
F2	.356**	-			
F3	.473**	.744**	-		
F4	.383**	.595**	.564**	-	

Notes. ***** $p < .05$; ****** $p < .01$; CFA = confirmatory factor analysis; ESEM = exploratory structural equation model; F1-F4: Factor 1 to 4; λ = standardized factor loading; δ = Standardized item uniquenesses; SA = Self-Awareness items; BP = Balanced Processing items; RT =Relational Transparency items; MP = Moral/Ethical Perspective items. Main factor loadings are indicated in bold.

Table 5

Standardized Parameter Estimates for the Authentic Leadership Integrated Questionnaire (AL-IQ) Measurement Models (Sample 1)

	CFA					ESEM				
Items	F1 (λ)	F2 (λ)	F3 (λ)	F4 (λ)	δ	F1 (λ)	F2 (λ)	F3 (λ)	F4 (λ)	δ
AL-IQ_SA1	.849**				.279**	**.623****	.013	.047	.263**	.251**
AL-IQ_SA2	.750**				.437**	**.691****	.120	.073	-.012	.369**
AL-IQ_SA3	.805**				.352**	**.348****	.194**	.222*	.181	.376**
AL-IQ_RT1		.789**			.378**	.084	**.840****	-.149**	.082	.307**
AL-IQ_RT2		.721**			.480**	.105	**.625****	.240**	-.203*	.441**
AL-IQ_RT3		.839**			.296**	-.056	**.700****	.038	.204**	.321**
AL-IQ_MP1			.799**		.361**	.017	.091	**.734****	-.014	.363**
AL-IQ_MP2			.829**		.313**	.168**	-.015	**.980****	.038	.198**
AL-IQ_MP3			.736**		.458**	.189	-.140*	**.568****	.189*	.425**
AL-IQ_MP4			.724**		.475**	.229*	.161*	**.491****	-.078	.484**
AL-IQ_BP1				.818**	.330**	.309**	-.054	.069	**.551****	.337**
AL-IQ_BP2				.854**	.270**	-.044	.013	.110*	**.846****	.228**
AL-IQ_BP3				.823**	.323**	-.212	.213**	.004	**.820****	.239**
AL-IQ_BP4				.837**	.300**	.217	-.009	.016	**.659****	.313**
Standardized Factor Correlations										
F1	-					-				
F2	.684**	-				.435**	-			
F3	.813**	.725**	-			.630**	.642**	-		
F4	.878**	.666**	.727**	-		.688**	.528**	.609**	-	

Notes. ***** p < .05; ****** p < .01; CFA = confirmatory factor analysis; ESEM = exploratory structural equation model; F1-F4: Factor 1 to 4; λ = standardized factor loading; δ = Standardized item uniquenesses; SA = Self-Awareness items; BP = Balanced Processing items; RT =Relational Transparency items; MP = Moral/Ethical Perspective items. Main factor loadings are indicated in bold.

Table 6

Measurement Invariance of the ESEM representation of responses to the Authentic Leadership Integrated Questionnaire (AL-IQ)

Model description	MLR χ^2	*df*	CFI	TLI	RMSEA and 90% CI	$\Delta\chi^2(df)$	ΔCFI	Δ TLI	ΔRMSEA
Invariance Across Gender: First-Order									
Configural invariance	170.394*	82	.960	.912	.083 (.066–.101)				
Weak invariance	191.436*	122	.969	.953	.060 (.043–.077)	38.457 (40)	+.009	+.041	-.023
Strong invariance	200.829*	132	.969	.957	.058 (.041–.074)	8.5646 (10)	.000	+.004	-.002
Strict invariance	233.645*	146	.960	.951	.062 (.047–.077)	33.326 (14)*	-.009	-.006	+.004
Latent variance-covariance invariance	238.924*	156	.963	.957	.058 (.043–.073)	7.781 (10)	+.003	+.006	-.004
Latent means invariance	255.989*	160	.957	.951	.061 (.048-.074)	19.016 (4) *	-.006	-.006	+.003
Invariance Across Gender: Second-Order									
Configural invariance	259.495*	150	.951	.940	.069 (.054-.082)				
Weak invariance	259.314*	153	.952	.943	.067 (.053–.081)	2.869 (3)	+.001	+.003	-.002
Strong invariance	263.460*	156	.952	.943	.067 (.052–.080)	4.175 (3)	.000	.000	.000
Strict invariance	261.198*	160	.954	.948	.064 (.049–.078)	1.308 (4)	+.002	+.005	-.003
Latent variance-covariance invariance	260.974*	161	.955	.949	.063 (.049–.077)	0.233 (1)	+.001	+.001	-.001
Latent means invariance	272.617*	162	.950	.944	.066 (.052–.080)	11.467 (1)*	-.005	-.005	+.003
Invariance Across Samples: First-Order									
Configural invariance	152.386*	82	.982	.961	.056 (.042-.070)				
Weak invariance	209.933*	122	.978	.967	.052 (.040-.063)	60.421 (40)	-.004	+.006	-.004
Strong invariance	255.081*	132	.969	.958	.059 (.048-.070)	57.436 (10)*	-.009	-.009	+.007
Strict invariance	298.800*	146	.962	.953	.062 (.052-.072)	40.098 (14)*	-.007	-.005	+.003
Latent variance-covariance invariance	312.717*	156	.961	.955	.061 (.051-.071)	15.730 (10)	-.001	+.002	-.001
Latent means invariance	322.027*	160	.960	.954	.061 (.052-.071)	9.403 (4)	-.001	-.001	.000
Invariance Across Samples: Second-Order									
Configural invariance	341.861*	150	.952	.942	.069 (.059-.079)				
Weak invariance	343.170*	153	.953	.944	.068 (.058-.078)	0.710 (3)	+.001	+.002	-.001
Strong invariance	344.163*	156	.953	.945	.067 (.057-.077)	0.680 (3)	.000	+.001	-.001
Strict invariance	349.906*	160	.953	.946	.066 (.057-.076)	7.240 (4)	.000	+.001	+.001
Latent variance-covariance invariance	354.517*	161	.952	.946	.067 (.057-.076)	4.878 (1)	-.001	.000	+.001
Latent means invariance	362.758*	162	.950	.944	.068 (.059-.077)	8.864 (1)*	-.002	-.002	+.001
Relations with the outcomes									
First-Order: Relations Free	1473.204*	754	.920	.913	.060 (.055-.064)				
First-Order: Relations Invariant	1486.249*	766	.919	.915	.059 (.055-.064)	14.388 (12)	-.001	+.002	-.001
Second-Order: Relations Free	1536.049*	774	.915	.911	.060 (.056-.065)				
Second-Order: Relations Invariant	1544.623*	777	.914	.910	.061 (.056-.065)	8.112 (3)	-.001	-.001	+.001

Notes: **p* < .01; ESEM = exploratory structural equation model; MLR χ^2 = Robust chi square test of exact fit; df = degrees of freedom; CFI = comparative fit index; TLI = Tucker–Lewis index; RMSEA = root-mean-square error approximation; CI = confidence interval; Δ = change from previous model in the sequence; Scaled chi square difference tests (calculated from models loglikelihoods for greater precision); goodness-of-fit indices including a correction for parsimony (TLI, RMSEA) can improve with the addition of model constraint. Although the remaining indicators (χ^2 and CFI) should be monotonic with model complexity, it is possible for them to improve with added constraints when the MLR scaling correction factors differ importantly across models. These improvements should be considered as random.

Table 7

Fully Invariant (Across Samples) Standardized Parameter Estimates for the final ESEM solution of Authentic Leadership Integrated Questionnaire (AL-IQ)

Items	**F1 (λ)**	**F2 (λ)**	**F3 (λ)**	**F4 (λ)**	**δ**
AL-IQ_SA1	**.653****	.041	.018	.221**	.231**
AL-IQ_SA2	**.794****	.083	.078	-.070	.289**
AL-IQ_SA3	**.444****	.100	.236**	.141	.331**
AL-IQ_RT1	.096*	**.839****	-.153**	.060	.315**
AL-IQ_RT2	.023	**.664****	.275**	-.153*	.360**
AL-IQ_RT3	-.050	**.844****	-.030	.130**	.222**
AL-IQ_MP1	-.035	.116*	**.734****	.014	.343**
AL-IQ_MP2	-.099	-.027	**.929****	.053	.220**
AL-IQ_MP3	.227**	-.069	**.476****	.186*	.432**
AL-IQ_MP4	.166*	.124	**.543****	-.058	.486**
AL-IQ_BP1	.246**	.005	.021	**.623****	.280**
AL-IQ_BP2	-.090	-.036	.085	**.955****	.149**
AL-IQ_BP3	-.170**	.140**	.062	**.864****	.226**
AL-IQ_BP4	.217**	.044	-.009	**.636****	.293**
Standardized Factor Correlations					
F1	-				
F2	.564**	-			
F3	.659**	.738**	-		
F4	.786**	.653**	.700**	-	

Notes. ***** $p < .05$; ****** $p < .01$; CFA = confirmatory factor analysis; ESEM = exploratory structural equation model; F1-F4: Factor 1 to 4; λ = standardized factor loading; δ = Standardized item uniquenesses; SA = Self-Awareness items; BP = Balanced Processing items; RT =Relational Transparency items; MP = Moral/Ethical Perspective items. Main factor loadings are indicated in bold.

Table 8

Latent Regressions between the Dimensions of the Authentic Leadership Integrated Questionnaire (AL-IQ) and the Outcomes

	Job satisfaction		Work performance		Psychological distress	
	b	*β*	*b*	*β*	*b*	*β*
ESEM Representation						
SA	-.026	-.026	-.094	-.095	.083	.071
RT	.059	.057	.142	.143	-.013	-.011
MP	.144	.138	.177	.178	-.113	-.097
BP	.328**	.315	-.057	-.058	-.268	-.229
Second-order factor	.478**	.455	.130**	.135	-.299**	-.256
CFA Representation						
SA	-.109	-.104	-.206	-.206	.373	.314
RT	.045	.044	.139	.140	.031	.026
MP	.178	.172	.297*	.297	-.273	-.229
BP	.374*	.359	-.063	-.063	-.438*	-.369
Second-order factor	.471**	.450	.133**	.138	-.291**	-.250

Notes: * $p < .05$; ****** $p < .01$; CFA = confirmatory factor analysis; ESEM = exploratory structural equation model; SA = Self-Awareness; BP = Balanced Processing; RT =Relational Transparency; MP = Moral/Ethical Perspective; *b* = Unstandardized regression coefficient; *β* = Standardized regression coefficient.

Online Supplemental Materials for:
New wine in a new bottle: Refining the assessment of authentic leadership using Exploratory Structural Equation Modeling (ESEM)

Appendix 1

Table S1
Alternative Measurement Models Estimated on the Authentic Leadership Questionnaire (ALQ) in the First Sample

Model description	MLR χ^2	*df*	CFI	TLI	RMSEA (90% CI)
One Factor Model	395.907*	104	.883	.864	.095 (.085–.105)
CFA					
2 factors (SA+RT; MP+BP)	394.258*	103	.883	.863	.095 (.085-.105)
2 factors (SA+MP; RT+BP)	383.993*	103	.887	.868	.094 (.084-.104)
2 factors (SA+BP; RT+MP)	345.960*	103	.902	.886	.087 (.077-.097)
3 factors (SA+RT; MP; BP)	317.347*	101	.913	.897	.083 (.073-.093)
3 factors (SA+MP; RT; BP)	378.639*	101	.888	.867	.094 (.084-.104)
3 factors (SA+BP; RT; MP)	308.703*	101	.916	.901	.081 (.071-.092)
3 factors (MP+BP; RT; SA)	378.677*	101	.888	.867	.094 (.084-.104)
3 factors (RT+MP; SA; BP)	338.971*	101	.904	.886	.087 (.077-.097)
3 factors (BP+RT; SA; MP)	315.836*	101	.914	.897	.083 (.072-.093)
A priori four factors	297.904*	98	.920	.902	.081 (.070–.091)
ESEM					
2 factors (SA+RT; MP+BP)	219.488*	89	.947	.929	.069 (.057-.080)
2 factors (SA+MP; RT+BP)	219.488*	89	.947	.929	.069 (.057-.080)
2 factors (SA+BP; RT+MP)	219.488*	89	.947	.929	.069 (.057-.080)
3 factors (SA+RT; MP; BP)	161.210*	75	.965	.944	.061 (.048-.074)
3 factors (SA+MP; RT; BP)	161.210*	75	.965	.944	.061 (.048-.074)
3 factors (SA+BP; RT; MP)	161.210*	75	.965	.944	.061 (.048-.074)
3 factors (MP+BP; RT; SA)	161.210*	75	.965	.944	.061 (.048-.074)
3 factors (RT+MP; SA; BP)	161.210*	75	.965	.944	.061 (.048-.074)
3 factors (BP+RT; SA; MP)	161.210*	75	.965	.944	.061 (.048-.074)
A priori four factors	124.747*	62	.975	.951	.057 (.042–.071)

Note. *$p < .01$; CFA = confirmatory factor analysis; ESEM = exploratory structural equation model; MLR χ^2 = Robust chi square test of exact fit; df = degrees of freedom; CFI = comparative fit index; TLI = Tucker–Lewis index; RMSEA = root-mean-square error approximation; CI = confidence interval; SA = Self-Awareness; BP = Balanced Processing; RT =Relational Transparency; MP = Moral/Ethical Perspective.

Table S2

Alternative Measurement Models Estimated on the Authentic Leadership Inventory (ALI) in the First Sample

Model description	MLR χ^2	*df*	CFI	TLI	RMSEA (90% CI)
One Factor Model	293.218*	77	.891	.871	.095 (.084–.107)
CFA					
2 factors (SA+RT; MP+BP)	288.769*	76	.892	.871	.095 (.083-.107)
2 factors (SA+MP; RT+BP)	286.490*	76	.893	.872	.094 (.083-.106)
2 factors (SA+BP; RT+MP)	228.646*	76	.923	.907	.080 (.069-.092)
3 factors (SA+RT; MP; BP)	209.157*	74	.932	.916	.077 (.065-.089)
3 factors (SA+MP; RT; BP)	232.123*	74	.920	.902	.083 (.071-.095)
3 factors (SA+BP; RT; MP)	193.989*	74	.939	.925	.072 (.060-.085)
3 factors (MP+BP; RT; SA)	264.676*	74	.903	.881	.091 (.079-.103)
3 factors (RT+MP; SA; BP)	213.269*	74	.929	.913	.078 (.066-.090)
3 factors (BP+RT; SA; MP)	235.042*	74	.918	.900	.084 (.072-.096)
Four factors model	177.722*	71	.946	.931	.070 (.057–.082)
ESEM					
2 factors (SA+RT; MP+BP)	198.515*	64	.932	.903	.082 (.069-.095)
2 factors (SA+MP; RT+BP)	198.515*	64	.932	.903	.082 (.069-.095)
2 factors (SA+BP; RT+MP)	198.515*	64	.932	.903	.082 (.069-.095)
3 factors (SA+RT; MP; BP)	112.496*	52	.969	.946	.061 (.046-.077)
3 factors (SA+MP; RT; BP)	112.496*	52	.969	.946	.061 (.046-.077)
3 factors (SA+BP; RT; MP)	112.496*	52	.969	.946	.061 (.046-.077)
3 factors (MP+BP; RT; SA)	112.496*	52	.969	.946	.061 (.046-.077)
3 factors (RT+MP; SA; BP)	112.496*	52	.969	.946	.061 (.046-.077)
3 factors (BP+RT; SA; MP)	112.496*	52	.969	.946	.061 (.046-.077)
Four factors model	72.900*	41	.984	.964	.050 (.031–.068)

Note. *$p < .01$; CFA = confirmatory factor analysis; ESEM = exploratory structural equation model; MLR χ^2 = Robust chi square test of exact fit; df = degrees of freedom; CFI = comparative fit index; TLI = Tucker–Lewis index; RMSEA = root-mean-square error approximation; CI = confidence interval; SA = Self-Awareness; BP = Balanced Processing; RT =Relational Transparency; MP = Moral/Ethical Perspective.

Table S3

Alternative Measurement Models Estimated on the Authentic Leadership Integrated Questionnaire (AL-IQ) in the First Sample

Model description	MLR χ^2	*df*	CFI	TLI	RMSEA (90% CI)
One Factor Model	455.362*	77	.821	.788	.126 (.115–.137)
CFA					
2 factors (SA+RT; MP+BP)	442.659*	76	.826	.792	.125 (.113-.136)
2 factors (SA+MP; RT+BP)	266.399*	76	.903	.883	.090 (.078-.102)
2 factors (SA+BP; RT+MP)	268.501*	76	.901	.882	.090 (.079-.102)
3 factors (SA+RT; MP; BP)	253.294*	74	.908	.887	.088 (.076-.100)
3 factors (SA+MP; RT; BP)	257.206*	74	.906	.885	.089 (.077-.101)
3 factors (SA+BP; RT; MP)	261.150*	74	.904	.882	.090 (.078-.102)
3 factors (MP+BP; RT; SA)	255.027*	74	.907	.886	.089 (.077-.101)
3 factors (RT+MP; SA; BP)	265.831*	74	.902	.879	.091 (.080-.103)
3 factors (BP+RT; SA; MP)	263.781*	74	.903	.881	.091 (.079-.103)
Four factors model	195.776*	71	.941	.924	.075 (.063–.088)
ESEM					
2 factors (SA+RT; MP+BP)	302.039*	64	.887	.840	.109 (.097-.122)
2 factors (SA+MP; RT+BP)	302.039*	64	.887	.840	.109 (.097-.122)
2 factors (SA+BP; RT+MP)	302.039*	64	.887	.840	.109 (.097-.122)
3 factors (SA+RT; MP; BP)	153.522*	52	.952	.916	.079 (.065-.094)
3 factors (SA+MP; RT; BP)	153.522*	52	.952	.916	.079 (.065-.094)
3 factors (SA+BP; RT; MP)	153.522*	52	.952	.916	.079 (.065-.094)
3 factors (MP+BP; RT; SA)	153.522*	52	.952	.916	.079 (.065-.094)
3 factors (RT+MP; SA; BP)	153.522*	52	.952	.916	.079 (.065-.094)
3 factors (BP+RT; SA; MP)	153.522*	52	.952	.916	.079 (.065-.094)
Four factors model	81.833*	41	.981	.957	.057 (.038–.074)

Note. *$p < .01$; CFA = confirmatory factor analysis; ESEM = exploratory structural equation model; MLR χ^2 = Robust chi square test of exact fit; df = degrees of freedom; CFI = comparative fit index; TLI = Tucker–Lewis index; RMSEA = root-mean-square error approximation; CI = confidence interval; SA = Self-Awareness; BP = Balanced Processing; RT =Relational Transparency; MP = Moral/Ethical Perspective.

Table S4

Alternative Measurement Models Estimated on the Authentic Leadership Integrated Questionnaire (AL-IQ) in the Second Sample

Model description	MLR χ^2	*df*	CFI	TLI	RMSEA (90% CI)
One Factor Model	363.079*	77	.851	.824	.128 (.115-.141)
CFA					
2 factors (SA+RT; MP+BP)	341.108*	76	.862	.835	.124 (.111-.138)
2 factors (SA+MP; RT+BP)	340.948*	76	.862	.835	.124 (.111-.137)
2 factors (SA+BP; RT+MP)	234.225*	76	.918	.901	.096 (.082-.110)
3 factors (SA+RT; MP; BP)	279.190*	74	.893	.869	.111 (.097-.124)
3 factors (SA+MP; RT; BP)	214.300*	74	.927	.910	.091 (.077-.106)
3 factors (SA+BP; RT; MP)	196.707*	74	.936	.921	.085 (.071-.100)
3 factors (MP+BP; RT; SA)	213.911*	74	.927	.910	.091 (.077-.106)
3 factors (RT+MP; SA; BP)	183.521*	74	.943	.930	.081 (.066-.096)
3 factors (BP+RT; SA; MP)	288.601*	74	.888	.863	.113 (.099-.127)
Four factors model	144.240*	71	.962	.951	.067 (.052-.083)
ESEM					
2 factors (SA+RT; MP+BP)	206.514*	64	.926	.895	.099 (.084-.114)
2 factors (SA+MP; RT+BP)	206.514*	64	.926	.895	.099 (.084-.114)
2 factors (SA+BP; RT+MP)	206.514*	64	.926	.895	.099 (.084-.114)
3 factors (SA+RT; MP; BP)	138.041*	52	.955	.922	.085 (.068-.103)
3 factors (SA+MP; RT; BP)	138.041*	52	.955	.922	.085 (.068-.103)
3 factors (SA+BP; RT; MP)	138.041*	52	.955	.922	.085 (.068-.103)
3 factors (MP+BP; RT; SA)	138.041*	52	.955	.922	.085 (.068-.103)
3 factors (RT+MP; SA; BP)	138.041*	52	.955	.922	.085 (.068-.103)
3 factors (BP+RT; SA; MP)	138.041*	52	.955	.922	.085 (.068-.103)
Four factors model	70.126*	41	.985	.966	.056 (.032-.078)

Note. *$p < .01$; CFA = confirmatory factor analysis; ESEM = exploratory structural equation model; MLR χ^2 = Robust chi square test of exact fit; df = degrees of freedom; CFI = comparative fit index; TLI = Tucker–Lewis index; RMSEA = root-mean-square error approximation; CI = confidence interval; SA = Self-Awareness; BP = Balanced Processing; RT =Relational Transparency; MP = Moral/Ethical Perspective.

Appendix 2

The 14 items of the Authentic Leadership Integrated Questionnaire

My leader…
1. … encourages others to voice opposing points of view. (BP-ALI)
2. … solicits comments to improve his/her way of interacting with others. (SA-ALQ)
3. … clearly states what he/she means. (RT-ALI)
4. … acts in accordance with his/her stated beliefs (MP-ALQ)
5. … asks for ideas that challenge his/her core beliefs. (BP-ALI)
6. … describes precisely how others views his/her abilities. (SA-ALQ)
7. … openly express his/her thoughts. (RT-ALQ)
8. … bases his/her decisions on its fundamental values. (MP-ALQ)
9. … indicates that he/she understands how certain actions can influence other people. (SA-ALQ)
10. … expresses his/her ideas and thoughts clearly to others. (RT-ALI)
11. … encourages me to make decisions that are consistent with my fundamental values. (MP-ALQ)
12. … carefully listens to alternative perspectives before reaching a conclusion. (BP-ALI)
13. … makes decisions based on a rigorous ethical code. (MP-ALQ)
14. … objectively analyzes relevant data before making a decision. (BP-ALI)

Note. For the purpose of this article, we followed the translation back-translation procedure described by Vallerand and Halliwell (1983) to translate the original French Canadian items into English. The selected items from the ALQ have all been adapted from the original items in order to achieve a greater level of consistency with the ALI items; SA = Self-Awareness; BP = Balanced Processing; RT =Relational Transparency; MP = Moral/Ethical Perspective.

Appendix 3

Table S5

Alternative Tests of Predictive Invariance based on CFA representations of responses to the Authentic Leadership Integrated Questionnaire (AL-IQ)

Model description	MLR χ^2	*df*	CFI	TLI	RMSEA and 90% CI	$\Delta\chi^2(df)$	ΔCFI	Δ TLI	ΔRMSEA
First-Order: Relations Free	1622.356*	784	.906	.903	.063 (.059-.067)				
First-Order: Relations Invariant	1636.594*	796	.906	.904	.063 (.058-.067)	14.728 (12)	.000	+.001	.000
Second-Order: Relations Free	1677.094*	804	.902	.901	.064 (.059-.068)				
Second-Order: Relations Invariant	1685.257*	807	.902	.901	.064 (.059-.068)	7.938 (3)	.000	.000	.000

Notes: *$p < .01$; CFA = confirmatory factor analysis; MLR χ^2 = Robust chi square test of exact fit; df = degrees of freedom; CFI = comparative fit index; TLI = Tucker–Lewis index; RMSEA = root-mean-square error approximation; CI = confidence interval; Δ = change from previous model in the sequence; Scaled chi square difference tests (calculated from models loglikelihoods for greater precision); goodness-of-fit indices including a correction for parsimony (TLI, RMSEA) can improve with the addition of model constraint. Although the remaining indicators (χ^2 and CFI) should be monotonic with model complexity, it is possible for them to improve with added constraints when the MLR scaling correction factors differ importantly across models. These improvements should be considered as random