

# ECONSTOR

*Make Your Publications Visible.*

A Service of

 Leibniz-Informationszentrum  
Wirtschaft  
Leibniz Information Centre  
for Economics

Hanges, Paul J.; Shteynberg, Gary

**Article**

## Methodological challenges and solutions for leadership researchers

Zeitschrift für Personalforschung

**Provided in Cooperation with:**

Rainer Hampp Verlag

Suggested Citation: Hanges, Paul J.; Shteynberg, Gary (2004) : Methodological challenges and solutions for leadership researchers, Zeitschrift für Personalforschung, ISSN 1862-0000, Hampp, Mering, Vol. 18, Iss. 3, pp. 346-358

This Version is available at:

<http://hdl.handle.net/10419/105192>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

[WWW.ECONSTOR.EU](http://WWW.ECONSTOR.EU)

Mitglied der  
  
Leibniz-Gemeinschaft

Paul J. Hanges, Gary Shteynberg\*

## **Methodological Challenges and Solutions for Leadership Researchers\*\***

Leadership scholars are faced with pressing methodological issues that challenge our current scholarship. In this paper, we discuss methodological concerns such as multi-level hypothesis testing, construct validity of group-level scales, sampling countries in cross-cultural research, self-report measurement – issues that are endemic to organizational leadership research. We point out the methodological challenges facing modern leadership researchers, and point to possible solutions that can provide opportunities for future progress in our field.

Key words: **Methodology, Statistics, Multilevel, HLM, Levels of Analysis, Cross-cultural Leadership Research, Factor Analysis, Measurement**

---

\* Paul J. Hanges, Department of Psychology, University of Maryland, College Park, MD 20742, USA. E-Mail: hanges@psyc.umd.edu.

Gary Shteynberg, Department of Psychology, University of Maryland, College Park, MD 20742, USA. E-Mail: gshteynberg@psyc.umd.edu.

\*\* Article received: September 30, 2003  
Revised version accepted: July 17, 2004.

The quest for explanation and prediction of effective organizational leadership has occupied social science researchers for the better part of the 20<sup>th</sup> century (Bass 1990; DeMeuse 1986; Mitchell 1979). Now at the dawn of 21<sup>st</sup> century, the growing sophistication of leadership theories has presented organizational researchers with important methodological questions in need of urgent answers. In this paper, we address several methodological issues which, in our estimation, cannot be ignored. In particular, we discuss several issues surrounding level of analysis as well as cross-cultural research. We will also discuss issues surrounding alternatives to self-report measures. It is our hope that this discussion will prove useful to researchers by identifying solutions to these methodological issues.

### **Levels of analysis issues**

There is growing awareness of the importance of level of analysis issues in the broader organizational literature (Klein/Kozlowski 2000; Hoffman 2002). Levels of analysis issues become a concern when empirical data has a nested or multilevel structure. Even though simple random sampling is taught in the typical graduate level statistics course, researchers do not frequently employ this sampling strategy in field research. Rather, researchers often obtain their data by gaining access to one or more organizations and then gathering information from multiple individuals within each organization. In all likelihood, several individuals probably come from the same work team in each organization. The data that results from employing this sampling strategy has a nested/multilevel structure because multiple respondents belong to the same work team (i.e., respondents are nested within work teams) and multiple work teams are sampled from the same organization (i.e., work teams are nested within organizations). The consequence of this nested data structure is that responses among respondents tend to be more similar than would be expected if true random sampling was employed. Specifically, the assumption of independence of errors which is critical assumption in almost all statistical techniques used in the social sciences is violated with nested data. As a result of violating this assumption, statistical analyses will yield biased results and conclusions (Bliese/Hanges, in press).

Leadership researchers have long lead the way by discussing the statistical problems caused by nested/multilevel data (e.g., Dansereau/Yammarino 1998a, 1998b; Dansereau/Yammarino/Markham 1995; Hanges/Dickson 2004; Hanges/Dickson/Sipe 2004; Waldman/Yammarino 1990; Yammarino/Bass 1991). While statistical issues surrounding such data have been discussed in the leadership literature, the majority of this work has focused on procedures justifying aggregation of measures. Unfortunately, fewer studies have discussed how to properly handle multilevel data to test scientific hypotheses. For example, suppose that a researcher is interested in the relationship between leader behavior and team morale. Following the sampling strategy described above, the researcher gets the permission of several organizations to ask their employees to complete a leadership-organizational culture questionnaire. Each respondent describes the behavior of his/her leader and the morale of their team. The researcher's data is nested because multiple respondents are sampled from the same team/leader and multiple teams are sampled from each organization. How should the

researcher properly test the hypothesis that leader behavior and team morale are related?

Historically, researchers have used one of two statistical approaches to test such hypotheses (Bryk/Raudenbushy 1987; Hofmann 1997; Hofmann et al. 2000). The first approach, typically called the “aggregate approach,” involves averaging all variables to team-level of analysis and performing traditional statistical analyses (e.g., correlation, regression) on this aggregated data (Hofmann 1997). The degrees of freedom associated with this approach are based on the number of work teams included in the data. The second approach, typically called the “disaggregated approach,” involves simply analyzing individual level ratings of leader behavior and team morale. The disaggregated approach appears to have the advantage of increased statistical power over the aggregate approach because in the disaggregated approach, degrees of freedom are based on the number of respondents in the entire data set and not on the number of teams (James/Williams 2000; Kreft/Leeuw 1998).

Both of these traditional approaches have limitations. For example, the aggregated approach cannot assess hypotheses in which more than one level of analysis (e.g., both leader behavior and follower personality affect team morale) is hypothesized to operate (Hofmann 1997). The disaggregated approach is problematic because by ignoring the nested nature of data, the “error independence” statistical assumption is violated and misleading conclusions can result.

### ***Levels of analysis and hypothesis testing***

Statistical techniques have been designed that specifically deal with multilevel data. One technique that can be used to test hypotheses at multiple levels of analysis is Within-and-Between Analysis (WABA). It was originally developed and primarily used in the leadership literature (Dansereau/ Alutto/Yammarino 1984). This analytic technique consists of three stages. First, the variability of each variable in a set of variables is examined to determine the level of analysis of each variable (Dansereau et al. 1984; Yammarino/Markham 1992). This step is typically referred to as WABA I. Second, the level of analysis responsible for obtained covariation between 2 variables is determined (Dansereau et al. 1984). This step is typically referred to as WABA II. Finally, the results of WABA I and WABA II are compared so that the researcher reaches a decision about the overall level of analysis for each variable and the level of analysis for each significant correlation found in an empirical study (Yammarino/Markham 1992).

Another technique that is starting to be discussed to test multilevel hypotheses is Hierarchical Linear Modeling (HLM), also known as random coefficient modeling (Hofmann 1997; Hofmann et al. 2000). This technique was originally developed in the educational literature (Bryk/Raudenbush 1992). It can be thought of as a multi-stage regression analysis that tests relationships between independent and dependent variables at multiple levels of analysis. The first stage focuses on the lowest (i.e., within group) level of analysis. Specifically, the significance of a regression equation predicting the within-group portion of some dependent variable by the within-group portion of one or more independent variables is tested (Hanges et al 2004). This stage basically estimates separate regression equations for each group. The second stage of the analy-

sis uses group variables to predict group differences in the intercepts and slopes of the within-group regression equations. This technique overcomes some of the deficiencies of the previous analytic approaches and is becoming an increasingly popular technique in the organizational sciences. Indeed, this statistical procedure was used in the GLOBE project to test its hypotheses (Hanges/Dickson/Sipe 2004).

Of course, these two techniques have their limitations. For example, WABA II can only assess the relationship between two variables if the independent and dependent variables are operating at the same level of analysis (e.g., team leader behavior affects team morale) (Castro 2002). HLM, on the other hand, is problematic when working with small samples as well as its requirement that the dependent variable be measured at the lowest possible level of analysis (Castro 2002; James/Williams 2000). Even though these two techniques have their limitations, they are still better than the more traditional statistical analyses. The question that remains is which of these two analyses provide more accurate conclusions. Researchers are starting to make comparisons between these two techniques (Bliese/Halverson/Schriesheim 2002; Castro 2002; Klein et al. 2000). We hope that more researchers will take advantage of these statistical techniques and use the more appropriate procedures in the future.

### ***Levels of analysis and factor analysis***

Another level of analysis issue that needs to be addressed in the leadership literature is how to properly conduct a factor analysis with items that were developed to measure some group level construct (e.g., team morale, organizational climate, societal culture). Factor analysis refers to a set of statistical techniques used to either explore or confirm the underlying structure among a set of items/variables to determine those items/variables that tap a factor or latent construct (Nunnally/Bernstein 1994). The utility of factor analysis in the identification and confirmation of scales that measure a single construct (i.e., unidimensionality) as well as the utility of factor analysis for establishing the construct validity (i.e., convergent/discriminant validity) of a scale is well known. It is therefore not surprising that leadership researchers rely on this statistical methodology for developing their scales.

It is critical to ensure that a scale measuring a group level (e.g., dyad, team, organization, societal) construct exhibits the desired dimensionality as well as convergent/discriminant validity at the group level of analysis. Unfortunately, researchers have once again either averaged their data to the group level of analysis (i.e., the aforementioned "aggregated" approach) or they ignored the nested structure of their data (i.e., the aforementioned "disaggregated" approach). As indicated previously, both of these approaches are problematic. It has been known since the mid-seventies that factor analysis of means (i.e., the aggregated approach) can produce misleading results (Cronbach 1976; Harnqvist 1978). A scale's factor structure can differ at different levels of analysis. Items may form a single factor at the group level but may split into two or more factors at the lower level (Hanges/Dickson 2004). The factor analysis of means reflects the group-level and lower-level factor structure (Dyer/Hanges/Hall 2004). Analyzing disaggregated data is also problematic. Like the aggregated analysis, factor analysis on disaggregated data can produce results that primarily reflect the factor structure at the wrong level (i.e., it can reflect the within-group as opposed to

group as opposed to group-level factor structure). It can also produce an overall result that combines the different level factor structures (e.g., the scale shows two factors overall by combining the single factor at the group level with the three factors at the within-group level).

Concern regarding the inappropriate application of factor analysis to group level scales was expressed by Chan (1998), who worried that "...despite the existence of broad theoretical frameworks and methodological advances, the fundamental substantive issue of construct validation in multilevel research has not been addressed adequately" (p.234). Importantly, the lack of empirically based studies examining the construct validity of aggregate measures means that we often do not know whether a given construct has an identical structure across different levels of analysis, or whether its structure varies across levels. This is a critical shortcoming in our research base.

Recent work, especially by Bengt Muthen (1990, 1994) provide a solution to this conundrum. Muthen developed a technique known as multilevel confirmatory factor analyses (MCFA) that specifically incorporates the hierarchical structure of data into the analysis and permits assessment of the factor structure across the two levels of analysis. MCFA can now be performed on updated versions of structural equation modeling (SEM) programs (e.g., EQS, MPlus, Lisrel). The analysis proceeds by splitting the data into a group level portion and a within-group level portion. MCFA allows researchers to fit the same or different factor structure at these two levels of analysis (Dyer/Hanges/Hall 2004; Hall/Hanges/Dyer, in press). This technique properly specifies the nature of the data into the analysis and thus, eliminates the problems with the aggregated/disaggregated factor analytic approaches.

Unfortunately, MCFA is not widely known in organizational research (for exceptions, see Hall et al., in press; Hall et al. 1999; Hanges/Dickson 2004). This is probably due to the fact that MCFA is only now becoming widely available as an option in SEM statistical packages. Clearly, more work with this analytic technique is needed in the leadership research literature.

### **Cross-cultural research**

Recent work in the leadership literature has explored the often subtle and hidden interplay between culture and leadership (e.g., Chong/Thomas 1997; Hanges, Lord/Dickson 2000; Hanges, Dorfman, Shteynberg/Bates, in press; Leslie/Van Velsor 1998). For example, Suutari (1996), Maczynski, Jago, Reber, and Boehnisch (1994), Den Hartog, House, Hanges, Ruiz-Qintanilla/Dorfman, et al (1999), and the GLOBE project (House et al. 2004) have found cultural variation in the desirability of particular leadership attributes. Research and theory focusing on cross-cultural leadership has clearly "taken off." However, while cross-cultural research is a step forward, there are several methodological issues that have not been adequately discussed with regard to such research.

#### ***Cross-cultural research and sample selection***

A discussion of an effective strategy to identify countries to include in a study has not appeared in our literature. However, there is a growing empirical based literature on the cultural similarity/differences among nations (e.g., Ronen/Shenkar 1985;

Gupta/Hanges 2004) that can be useful for addressing this issue. Ronen and Shenkar (1985) and Gupta and Hanges (2004) have classified numerous countries into categories based on the similarity/differences of their scores across multiple culture dimensions. These classifications provide a holistic conceptualization of the distance between societies because they are a function of considering multiple cultural dimensions. We believe that these clusters can assist cross-cultural researchers by helping them identify societies that meaningfully vary in terms of cultural distance.

For example, in the recent GLOBE study (House et al. 2004), 61 nations were grouped into 10 distinct clusters (Gupta/Hanges 2004). The ten clusters are *Anglo* (e.g., England, Australia, New Zealand), *Latin Europe* (e.g., Italy, Spain, Israel), *Nordic Europe* (i.e., Finland, Sweden, Denmark), *Germanic Europe* (i.e., Netherlands, Austria, Switzerland, Germany), *Eastern Europe* (e.g., Hungary, Russia, Kazakhstan), *Latin America* (e.g., Costa Rica, Venezuela, Ecuador), *Middle east* (e.g., Qatar, Morocco, Turkey), *Sub-Sahara Africa* (e.g., Namibia, Zambia, Zimbabwe), *Southern Asia* (e.g., India, Philippines, Malaysia) and *Confucian Asia* (e.g., Taiwan, Singapore, South Korea). Cross-cultural leadership researchers can use this clustering to identify countries that fall into different clusters. This would ensure adequate cultural distance in the study so that any conclusions have the maximum potential for generalization. Our point in this section is basically not all multinational samples are equivalent. A study that has sampled data from England, Australia, and New Zealand is not as good for testing the robustness of a theory as a study that has sampled data from England, Spain, and Sweden.

Another sampling issue with cross-cultural leadership research focuses on how many nations to include in a study. Researchers agree that it is impossible to determine whether an empirical finding is moderated by culture when data is sampled from only one culture. However, the ability to unambiguously identify findings that are meaningfully moderated by culture does not dramatically improve when data from two, three, or even four nations are studied. Significant differences between a small subset of nations may simply be due to unique characteristics of the nations included in the study rather than some hypothesized cultural dimension (e.g., collectivism).

Significant differences between countries cannot be interpreted as evidence for cultural moderation unless there are a sufficient number of countries included in a study. Experimental researchers have made this argument in their literature with the distinction between “fixed” effects and “random” effects. An independent variable (e.g., nations) is treated as a “fixed” effect when a researcher is interested in generalizing significant findings to only the specific levels of the independent variable included in the study (e.g., researcher only wants to generalize results to the specific countries included in his/her study). An independent variable is treated as a “random” effect when a researcher is interested in generalizing significant findings beyond the specific levels of the independent variable included in his/her study. In the present context, this would be the case of a researcher interested in generalizing his/her findings to all countries as opposed to only those specific countries included in his/her study. Clearly, when testing whether a theoretical construct such as culture moderates some empirical relationship, it is critical that a random effects approach is taken. Unfortu-

nately, statistical power requirements for a random effect independent variable is substantially greater than the 2, 3, or 4 countries included in some cross-cultural research.

Of course, adequate sampling for variables treated as random effects dramatically increases the difficulty of conducting cross-cultural studies. One way around this challenge is to develop a consortium of researchers who are interested in studying some phenomena and who are distributed throughout the world. Such a consortium can greatly increase the ease of conducting meaningful studies because these research teams can simultaneously collect data in multiple countries for multiple studies. While global research teams are not without their problems (c.f., Hanges/Lyon/Dorfman, in press), the payoffs of such teams are priceless.

### ***Cross-cultural research and measurement equivalence***

One critical issue for all cross-cultural research endeavors is to ensure that the construct of interest is meaningful in all sampled cultures. The cause of any discovered differences (or similarities) is never clear. Response differences might be due to cultural differences, lack of consistent operationalization and definition of constructs across countries, or even culturally-specific response biases (Dorfman 1996). Measurement equivalence is one way by which the lack of consistent operationalization and definition as an explanation for a study's results is ruled out.

The usual strategy for assessing measurement equivalence is to conduct multi-group confirmatory factor analyses. In this analysis, the same factor structure is imposed on the data sampled from each country. Progressively restrictive models are imposed on the data until the most restrictive (i.e., equivalence in factor structure, factor loadings, factor variance, and factor covariances) is imposed. If the fit indices for this last model meet pre-set standards, then there is evidence that the scale is operating similarly in the various countries. In other words, significant country differences are not due to measurement problems.

While the traditional measurement equivalence analysis is extremely useful, researchers need to realize that this analytic technique only works for scales measuring individual level phenomena. What happens when group level constructs are measured in a study? There really has not been a discussion of how to assess the measurement equivalence of group level scales. One possible solution is to use the MCFA procedure discussed earlier and to combine it with a multigroup approach. For example, if a cross-cultural researcher collects data from multiple countries and in each country, data from multiple organizations is collected, and in each organization, multiple followers are asked to describe their leader, then multigroup MCFA is needed to determine the measurement equivalence of the leadership scale. Specifically, the researcher needs to impose the same MCFA structure simultaneously across all the countries in the researcher's data. If the fit indices of the analysis exceed accepted standards, the measurement equivalence of this group-level construct is ascertained. Unfortunately, to date, no one has applied this approach to determine measurement equivalence of a group-level construct.

Even with multigroup MCFA, there are still unresolved issues surrounding assessing the measurement equivalence of group-level scales. Specifically, how is the measurement equivalence of societal-level constructs determined? With scales measur-



ing society-level constructs such as GLOBE societal culture scales, it is assumed that the individuals from each country are all providing their responses about the same target (i.e., societal culture). In other words, the true score for societal-level scales is a constant within a country and the true score for these scales only vary when data from multiple countries is collected. The typical measurement equivalence approach as well as the multigroup MCFA approach described above assume that there are true score differences within each country. Thus, these statistical techniques cannot be used for assessing the measurement equivalence of society-level scales. Unfortunately, a statistical procedure by which the equivalence of such scales can be determined has yet to be developed. Clearly, there is a real need in the cross-cultural literature for such a statistical procedure.

### **Alternative measurement strategies**

Frequently, researchers are interested in measuring a leader's mental processes or personality characteristics in order to test theories or predict behavior. Unfortunately, the most common strategy employed to measure such constructs is by asking the leader to self-disclose. Such measures are contaminated by self-presentation biases in which respondents' scores reflect not only their beliefs, attitudes, and other characteristics, but also the deliberate and conscious manipulation of responses to regulate their impression to others (Dunton/Fazio 1997; Plant/Devine 1998).

Social psychologists have recently proposed a new measurement protocol to minimize the influence of self-presentation bias. This new measurement protocol involves the inclusion of implicit attitude (IA) measures in their studies. Implicit attitude measures seek to indirectly assess the value of a construct without directly asking participants for a self-report (Fazio/Olson 2003). The IA measures attitudes and other characteristics of participants (e.g., personality, emotions, goals, needs) by recording the speed and accuracy with which they can categorize words. For example, when using an IA measure of racist attitudes, Ziegert and Hanges (in press) asked participants to sort words into either a Black-White category or into a pleasant-unpleasant category. The categorization task was accomplished by having participants hit one computer key if a word either belonged to the Black category or the pleasant category and another key if the word belonged to either the White category or the unpleasant category. The speed and accuracy of this categorization process were recorded as was the speed and accuracy of the categorization when race categories were switched with the pleasant-unpleasant category. Ziegert and Hanges found that participants' speed and accuracy on the IA measure subsequently predicted their degree of bias on an in-basket exercise whereas a self-report measure of racist attitudes did not. Thus, implicit measures are able to assess participant characteristics that may not be accurately assessed with self-report measures. Indeed, the discrepancy between implicit and explicit measures has led to questions of the trustworthiness of easily monitored explicit responses such as verbal reports and self-report ratings (Fazio et al. 1995). Indeed, self-report measures may be especially questionable in field research where followers are explicitly asked to rate management on a variety of dimensions. In this context, implicit measurement may be effective in preventing conscious monitoring and modification of responses due to follower fears of retaliation or hopes of ingratiation.

While there is an active debate in the social psychology literature concerning the meaningfulness of the IA measure (e.g., Devine 2001), there is a substantial amount of validity evidence for it. Greenwald/Nosek (2001) reviewed over 30 studies discussing the psychometric and validity evidence for the IA measure. This research has shown that IA measures detect stable differences (Greenwald et al. 1998) and that these measures show convergent validity with other latency and priming measures (Rudman/Kilianski 2000) as well as with physiological measures (Phelps et al. 2000).

Overall, there is growing evidence for the convergent, discriminant, and predictive validity of implicit measurement. Such measures can prove to be an extremely useful for leadership researchers. One clear application of IA measures is in research on implicit leadership theory. Robert Lord and his colleagues have been discussing and testing the implicit leadership perspective for nearly 25 years (Lord/Foti/DeVader 1984; Lord, Foti/Phillips 1982; Lord/Maher 1991; Lord/Brown 2004). Reaction time measures and other IA measures could be used to assess various aspects of a person's leadership schema. For example, IA can assess the extent to which leadership schemas are chronic and easily accessible as well as the extent to which socially undesirable aspects (i.e., stereotypes/prejudices against women) influence the leadership schema. These IA measures might even be helpful for measuring personal characteristics that participants might not be consciously aware of. In summary, implicit measures appear to have promise and should be explored further in the leadership research.

### ***A note on qualitative research***

With the opportunities inherent to implicit measures acknowledged, we would be remiss not to mention the benefits that could accrue if researchers actively attempted to use both qualitative and quantitative methods to study leadership phenomena. Bachiocchi and Weiner (2002) recommend incorporating qualitative methodologies into research strategies when the context is central to the research question. The role of context factors (e.g., leader actions, socio-cultural events) plays an increasing critical part in many leadership theories (e.g., Hanges, Lord/Dickson 2000). Qualitative methodology provide richer description of these contextual factors and can yield insights into the meaning underlying the numbers and relationships obtained using quantitative methods. More work integrating qualitative and quantitative methodology will help our field discover truths and reduce the possibility that our findings are limited by our methodology.

### **Concluding remarks**

In this paper, we have discussed a number of methodological challenges that confront leadership researchers. While not all of these challenges have been solved (e.g., measurement equivalence for societal-level constructs), strategies for many of them have been developed. We covered three major areas where these methodological issues can arise. First, we discussed how multilevel constructs and hierarchically nested data create statistical problems that can potentially bias results if researchers ignore the multilevel nature of their data. WABA and HLM are two statistical techniques specifically designed to handle multilevel data. Both of these techniques have their benefits and

their drawbacks. Comparison of these statistical procedures is just beginning. While we cannot definitively recommend one over the other, it is clear that these statistical tools are better than the traditional approaches (i.e., previously discussed aggregate or disaggregate approaches) used to test hypotheses with multilevel data. We also mentioned the need for exploring the construct validity and factor structure of group level scales. Unfortunately, this is an often neglected topic. The utility of MCFA for assessing and validating the structure of group level scales should be recognized.

Second, we also discussed methodology concerns that arise when conducting cross-cultural leadership research. We discussed issues concerning the sampling of countries. Specifically, which countries should be included in a study and how many countries to include? Researchers need to be more careful and deliberate when choosing which cultures they sample their data from. We suggested that the culture clusters developed by Ronen and Shenkar (1985) or Gupta and Hanges (2004) might be helpful in strategically identifying countries to include in a study. We also discussed the critical role that measurement equivalence plays in data interpretation. One cannot meaningfully interpret any empirical findings without knowing that the scale is operating consistently within all countries. While traditional measurement equivalence statistical procedures are useful for assessing the equivalence of individual level scales, researchers need to start using multigroup MCFA for group-level constructs.

Finally, we discussed the problem of using only self-report measures throughout the leadership literature. Self-report measures only provide information that a participant is consciously aware of and such measures are highly susceptible to self-presentation biases. We suggested that implicit attitude measurement protocols be incorporated into leadership research. This alternative measurement strategy can go a long way to resolve the conscious level processing and self-presentation bias that limit the utility of self-report measures.

There are still many methodological challenges that need to be addressed. However, as we discuss, there are solutions to some of the commonly occurring problems encountered in the leadership literature. It is our hope that leadership researchers will find the methods discussed in this paper beneficial and informative to their endeavors.

## References

- Bachiochi, P.D./Weiner, S.P. (2002): Qualitative data collection and analysis. In S.G. Rogelberg (Ed.): *Handbook of Research Methods in Industrial and Organizational Psychology*. (pp. 161-183). Malden, MA: Blackwell.
- Barrett, G.V./Bass, B.M.(1976): Cross-cultural issues in Industrial and Organizational Psychology. In: M. D. Dunnette (Ed.): *Handbook of Industrial and Organizational Psychology*. (pp. 1639-1686). Chicago, IL: Rand McNally.
- Bass, B.M. (1990): *Bass & Stogdill's handbook of leadership: Theory, research, and managerial applications* (3<sup>rd</sup> ed.). New York: Free Press.
- Bliese, P.D./Halverson, R. R./Schriesheim, C.A. (2002): Benchmarking multilevel methods in leadership: The articles, the data, and the data set. In: *Leadership Quarterly*, 13, 3-14.
- Bliese, P.D./Hanges, P.J. (In press): Being both too liberal and too conservative: The perils of treating grouped data as though it is independent. In: *Organizational Research Methods*.
- Bryk, A. S./Raudenbush, S. W. (1987): *Hierarchical linear models*. Newbury Park, CA: SAGE Publications.

- Carl, D./Gupta, V./Javidan, M. (2004): Power distance. (pp. 513-560). In: R. J. House, P. J. Hanges, M. Javidan, P.W. Dorfman/V. Gupta (Eds.): *Culture, leadership, and organizations: The GLOBE study of 62 societies*. Thousand Oaks, CA: Sage Publications, Inc.
- Castro, S.L. (2002): Data analytic methods for the analysis of multilevel questions: A comparison of intraclass correlation coefficients,  $r_{wg(i)}$ , hierarchical linear modeling, within- and between-analysis, and random group resampling. In: *Leadership Quarterly*, 13, 69-93.
- Chan, D. (1998): Functional relations among constructs in the same content domain in different levels of analysis: A typology of composition models. In: *Journal of Applied Psychology*, 83, 234-246.
- Chong, L. M. A./Thomas, D. C. (1997): Leadership perceptions in cross-cultural context: Pakeha and Pacific islanders in New Zealand. In: *Leadership Quarterly*, 8(3), 275-293.
- Cronbach, L. J. (1976): *Research on classrooms and schools: Formulation of questions, design, and analysis*. Unpublished manuscript, Stanford University, California.
- Dansereau, F./Alutto, J.A./Yammarino, F.J. (1984): *Theory testing in organizational behavior: The variant approach*. Englewood Cliffs, NJ: Prentice-Hall.
- Dansereau, F./Yammarino, F.J. (1998a): *Leadership: The Multiple-Level Approaches, Classical and New Wave*. Stamford, Connecticut: JAI Press, Inc.
- Dansereau, F./Yammarino, F.J. (1998b): *Leadership: The Multiple-Level Approaches, Contemporary and Alternative*. Stamford, Connecticut: JAI Press, Inc.
- Dansereau, F./Yammarino, F.J./Markham, S.E. (1995): Leadership: The multiple-level approaches. In: *Leadership Quarterly*, 6, 251-263.
- DeMeuse, K.P. (1986): A compendium of frequently used measures in Industrial/Organizational Psychology. In: *The Industrial and Organizational Psychologist*, 23, 53.
- Den Hartog, D./House, R. J./Hanges, P. J./Ruiz-Quintanilla, S. A./Dorfman, P. W./Associates. (1999): Culture specific and cross culturally generalizable implicit leadership theories: Are attributes of charismatic/transformational leadership universally endorsed? In: *Leadership Quarterly*, 10(2), 219-256.
- Devine, P. G. (2001): Implicit prejudice and stereotyping: How automatic are they? Introduction to the special section. In: *Journal of Personality and Social Psychology*, 81, 757-759.
- Dorfman, P.W./House, R.J. (2004): Cultural influences on organizational leadership. In: R. J. House/P.J. Hanges/M. Javidan/P.W. Dorfman/V. Gupta (Eds.): *Leadership, Culture, and Organizations: The GLOBE study of 62 societies*. (pp. 51-73). Thousand Oaks, CA: SAGE.
- Dorfman, P.W. (1996): International and cross-cultural leadership. In: J. Punnitt/O. Shanker (Eds.): *Handbook for International Management Research*. (pp. 267-349). Cambridge, MA: Blackwell.
- Dorfman, P./Hanges, P. J./Brodbeck, F. C. (2004): Leadership and cultural variation: The identification of culturally endorsed leadership profiles. In R. J. House/P. J. Hanges/M. Javidan/P. Dorfman/V. Gupta (Eds.), *Leadership, culture, and organizations: The GLOBE study of 62 societies* (pp. 667-718). Thousand Oaks, CA: Sage Publications, Inc.
- Dunton, B. C./Fazio, R. H. (1997): An individual difference measure of motivation to control prejudiced reactions. In: *Personality and Social Psychology Bulletin*, 23, 316-326.
- Dyer, N./Hanges, P.J./Hall, R.A. (2004): Applying multilevel confirmatory factor analysis techniques to the study of leadership. Working paper.
- Fazio, R. H./Olson, M. A. (2003): Implicit measures in social cognition research: Their meaning and use. In: *Annual Review of Psychology*, 54, 297 – 327.
- Fazio, R. H./Jackson, J. R./Dunton, B. C./Williams, C. J. (1995): Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? In: *Journal of Personality and Social Psychology*, 69, 1013-1027.
- Greenwald, A. G./Nosek, B. A. (2001): Health of the Implicit Association Test at age 3. In: *Zeitschrift für Experimentelle Psychologie*, 48, 85 – 93.
- Greenwald, A. G./McGhee, D. E./Schwartz, J. L. K. (1998): Measuring individual differences in implicit cognition: The Implicit Association Test. In: *Journal of Personality and Social Psychology*, 74, 1464-1480.
- Gupta, V./Hanges, P.J. (2004): Regional and climate clustering of social cultures. In: R. J. House/P.J. Hanges/M. Javidan/P.W. Dorfman/V. Gupta (Eds.): *Leadership, Culture, and Organizations: The GLOBE study of 62 societies*. (pp. 178-218). Thousand Oaks, CA: SAGE.

- Hall, R.J./Hanges, P.J./Dyer, N.G. (in press): Two-level structural equation modeling with latent variables: An illustration of confirmatory factor analysis with applications to organizational research. In: *Organizational Research Methods*.
- Hall, R. J./Makiney, J. D./Marchioro, C. A./Tan, J. A./Klein, K. J. (April 1999): Applying multilevel structural equation modeling techniques to the study of organizational behavior. Paper presented at the 14<sup>th</sup> Annual Conference of the Society for Industrial and Organizational Psychology, Atlanta, GA.
- Hanges, P.J./Dickson, M.W. (2004): The development and validation of the GLOBE culture and leadership scales. In: R. J. House, P.J. Hanges, M. Javidan, P.W. Dorfman/V. Gupta (Eds.): *Leadership, Culture, and Organizations: The GLOBE study of 62 societies*. (pp. 122-151). Thousand Oaks, CA: SAGE.
- Hanges, P. J./Dickson, M. W./Sipe, M. T. (2004): Rationale for GLOBE statistical analyses: Societal rankings and test of hypotheses. In R. J. House, P. J. Hanges, M. Javidan, P. Dorfman/V. Gupta (Eds.): *Leadership, culture, and organizations: The GLOBE study of 62 societies* (pp. 219-234). Thousand Oaks, CA: Sage.
- Hanges, P.J./Dorfman, P.W./Shteynberg, G/Bates, A. (in press): Implications of culture and leadership information processing: Leadership style doesn't work a bit? Maybe you're a cultural misfit! In: W. H. Mobley/E. Weldon (Eds): *Advances in Global Leadership* (Vol 4). NY: JAI Press.
- Hanges, P. J./Lord, R. G./Dickson, M. W. (2000): An information processing perspective on leadership and culture: A case for connectionist architecture. In: *Applied Psychology: An International Review*, 49, 133-161.
- Hanges, P.J./Lyon, J./Dorfman, P.W, (In press): Managing an international research team: Lessons from project GLOBE. In: D. Shapiro, M.A. Von Glinow, and J. Cheng (Eds.) *Managing Multinational Teams: Cultural, Organizational, and National Influences*. Elsevier Press.
- Harnqvist, K. (1978): Primary mental abilities of collective and individual levels. In: *Journal of Educational Psychology*, 70, 706-716.
- Hofmann, D. A./Griffin, M. A./Gavin, M. B. (2000): The application of hierarchical linear modeling to organizational research. In K. J. Klein/S. W. J. Koslowski (Eds.), *Multilevel theory, research, and methods in organizations* (pp. 467-511). San Francisco: Jossey-Bass.
- Hofmann, D. A. (1997): An overview of the logic and rationale of hierarchical linear models. In: *Journal of Management*, 23, 723-744.
- House, R. J./Hanges, P. J./Javidan, M./Dorfman, P. W./Gupta, V./GLOBE Associates. (2004): *Culture, leadership, and organizations: The GLOBE study of 62 societies*. Thousand Oaks, CA: Sage Publications, Inc.
- House, R. J./Hanges, P. J./Ruiz-Quintanilla, S. A./Dorfman, P. W./Javidan, M./Dickson, M./Gupta, V./170 co-authors. (1999): *Cultural influences on leadership and organizations: Project GLOBE*. In: W. F. Mobley/M. J. Gessner/V. Arnold (Eds.): *Advances in global leadership* (Vol. 1, pp. 171-233). Stamford, CT: JAI Press.
- James, L. R./Williams, L. J. (2000): The cross-level operator in regression, ANCOVA, and contextual analysis. In: K. J. Klein/S. W. J. Koslowski (Eds.): *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions*. San Francisco: Jossey-Bass.
- Javidan, M./House, R. J./Dorfman, P. W. (2004): A nontechnical summary of GLOBE findings. In: R. J. House/P. J. Hanges/M. Javidan/P. Dorfman/V. Gupta (Eds.), *Leadership, culture, and organizations: The GLOBE study of 62 societies* (pp. 29-48). Thousand Oaks, CA: Sage.
- Klein, K.J./Bliese, P.D./Kozlowski, S.W.J./Dansereau, F./Gavin, M.B./Griffin, M.A./Hofmann, D.A./James, L.R./Yammarino, F.J./Bligh, M.C. (2000): Multilevel analytic techniques: Commonalities, differences, and continuing questions. In: K. J. Klein/S. W. J. Koslowski (Eds.), *Multilevel theory, research, and methods in organizations* (pp. 512-556). San Francisco: Jossey-Bass.
- Klein, K.J./Kozlowski, S.W.J. (2000): From micro to meso: Critical steps in conceptualizing and conducting multilevel research. In: *Organizational Research Methods*, 3, 211-236.
- Kreft, I./De Leeuw, J. (1998): *Introducing multilevel modeling*. Thousand Oaks, CA: SAGE Publications.
- Leslie, J./Van Velsor, E. (1998): *A cross-national comparison of effective leadership and teamwork: Toward a global workforce*. Greensboro, NC: Center for Creative Leadership.

- Lord, R. G./Brown, D. J. (2004): Leadership processes and follower self-identity. Mahwah, NJ: Lawrence Erlbaum.
- Lord, R. G./Foti, R. J./DeVader, C. L. (1984): A test of leadership categorization theory: Internal structure, information processing, and leadership perceptions. *Organizational Behavior and Human Performance*, 34, 343-378.
- Lord, R. G./Foti, R. J./Phillips, J. S. (1982): A theory of leadership categorization. In J. G. Hunt/U. Sekaran/C. A. Schriesheim (Eds.), *Leadership: Beyond establishment views* (pp. 104-121). Carbondale: Southern Illinois University Press.
- Lord, R.G./Maher, K.J. (1991): Leadership and information processing: Linking perceptions and performance. Cambridge, MA: Unwin Hyman.
- Maczynski, J./Jago, A./Reber, G./Boehnisch, W. (1994): Culture and leadership styles: A comparison of Polish, Austrian, and U.S. managers. In: *Polish Psychological Bulletin*, 25(4), 303-315.
- Mitchell, T. R. (1979): Organizational behavior. In: *Annual Review of Psychology*, 30, 243-281.
- Muthen, B.O. (1990): Mean and covariance structure analysis of hierarchical data. Paper presented at the Psychometric Society, Princeton, New Jersey
- Muthen, B.O. (1994): Multilevel covariance structure analysis. In: *Sociological Methods/Research*, 22, 376-398.
- Nunnally, J.C./Bernstein, I.H. (1994): *Psychometric Theory*. New York: McGraw Hill.
- Phelps, E. A./O'Connor, K. J./Cunningham, W. A./Gatenby, J. C./Funayama, E. S./Gore, J. C./Banaji, M. R. (2000): Amygdala activation predicts performance on indirect measures of racial bias. In: *Journal of Cognitive Neuroscience*, 12, 729-738.
- Plant, E. A./Devine, P. G. (1998): Internal and external motivation to respond without prejudice. In: *Journal of Personality and Social Psychology*, 75, 811-832.
- Ronen, S./Shenkar, O. (1985): Clustering countries on attitudinal dimensions: A review and synthesis. In: *Academy of Management Review*, 10, 435-454.
- Rudman, L. A./Kilianski, S. E. (2000): Implicit and Explicit attitudes toward female authority. In: *Personality and Social Psychology Bulletin*, 26, 1315-1328.
- Stogdill, R.M. (1974): *Handbook of leadership* (1<sup>st</sup> ed.). New York: Free Press.
- Suutari, V. (1996): Leadership ideologies among European managers: A comparative survey in a multinational company. In: *Scandinavian Journal of Management*, 12(4), 389-409.
- Waldman, D.A./Yammarino, F.J. (1990): A multiple level investigation of personnel ratings. In: *Personnel Psychology*, 43, 811-835.
- Yammarino, F.J./Bass, B.M. (1991): Person and situation views of leadership: A multiple levels of analysis approach. In: *Leadership Quarterly*, 2, 121-139.
- Yammarino, F.J./Markham, S.E. (1992): On the application of within and between analysis: Are absence and affect really group-based phenomena? In: *Journal of Applied Psychology*, 77, 168-176.
- Zaccaro, S.J./Klimoski, R.J. (2001): The nature of organizational leadership: An introduction. In: S.J. Zaccaro/R.J. Klimoski (Eds.): *The nature of organizational leadership: Understanding the performance imperatives confronting today's leaders*. (pp. 3-41). San Francisco, CA: Jossey-Bass.
- Ziegert, J. C./Hanges, P.J. (In press): Employment Discrimination: The Role of Implicit Attitudes, Motivation, and a Climate for Racial Bias. In: *Journal of Applied Psychology*.