Abstract

DATTA, ANASUYA. Measurement Equivalence of English and Spanish Versions of the Perceived Leader Integrity Scale. (Under the direction of John Michael and Bart Craig.)

Research over the last three decades has addressed the importance of integrity in leadership (e.g., Burns, 1978; Fairholm, 1991; Posner & Schmidt, 1984; Vitell & Davis, 1990). Research and application are stunted without adequate measures that assess the extent to which leaders demonstrate ethical or unethical behaviors. As business activities between countries increase (Stephens & Greer, 1995), having tests available in multiple languages can have various benefits (Zumbo, 2003). The availability of a measure in different languages can allow researchers and practitioners to facilitate assessment without having to build a new test, develop understandings of new cultural differences, and conduct comparative research. This study used the differential functioning of items and tests (DFIT; Raju, van der Linden, & Fleer, 1995) framework, based on item response theory (IRT), to assess the measurement equivalence between two language versions of the Perceived Leader Integrity Scale (PLIS; Craig & Gustafson, 1998) using samples collected from the United States, New Zealand, and Mexico. The U.S. and New Zealand samples formed the English speaking or US-NZ group and the Mexico sample formed the Spanish speaking group. Two indices of DFIT were used to determine item level (NCDIF) and test level (DTF) inequivalence between the comparison groups. Results showed 17.9% (5 out of 28) of the items to be differentially functioning. No significant DTF was identified at the test level. Post hoc explanations of the items with significant NCDIF in terms of possible cultural and linguistic influences provide information about the possible reasons why the items are functioning differentially (e.g. translation errors, cultural differences, or both). Practical implications of the current study are discussed.

MEASUREMENT EQUIVALENCE OF ENGLISH AND SPANISH VERSIONS OF THE PERCEIVED LEADER INTEGRITY SCALE

by

ANASUYA DATTA

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Master of Science

August 2005

PSYCHOLOGY

APPROVED BY

Joan J. Michael, Ph.D.
Chair of Advisory Committee

S. Bartholomew Craig, Ph.D.
Co-Chair of Advisory Committee

Adam W. Meade, Ph.D.

Biography

Anasuya Datta was born in New Delhi, India, the daughter of Mr. Supriyo and Mrs. Supriya Datta. Anasuya attended Delhi Public School until the seventh grade. As a child she enjoyed reading mystery novels, collecting stamps and singing classical music. At the age of 14 Anasuya moved to San Diego, California with her parents and siblings. She completed 8th grade in Meadowbrook Middle School and was in the first four-year graduating class from Rancho Bernardo High School where she was a member of the choir and track team. Anasuya went on to attend the University of California at Irvine, where she majored in Psychology and competed as a varsity collegiate athlete in track and field. Her favorite events were the 200m, 400m, and 4x400m relay. She graduated with her B.A. in 1998. Following graduation, Anasuya decided to gain a few years of work experience, which helped her to make the decision to continue her academic career. In 2003, Anasuya decided to move to Raleigh, NC to begin her Masters in I/O Psychology at North Carolina State University. While in Raleigh, Anasuya met her future husband Chetan and also successfully completed the requirements for the M.S. in I/O Psychology during the Summer of 2005. Anasuya plans on pursuing a career in human resources and consulting hereafter.

Acknowledgements

There are several people who have helped me along that I would like to thank and convey my sincere gratitude to. First, I would like to thank my committee members: Drs. Joan Michael, Bart Craig and Adam Meade. Without their consultation, mentorship, and continued support, I would not have been able to produce this level of research work. I especially want to thank Bart Craig for suggesting the topic of research and sharing his resources with me, which ultimately made this study possible. I am obliged to him for his unimaginable patience in showing me how to conduct the analyses that were performed in this study as well as enduring revisions of the text throughout the writing process. I also thank Joan Michael and Adam Meade for their guidance and unfailing support in providing me their technical knowledge regarding the material and their feedback within a short period of time.

A special thanks to Florina Arredondo, Sigrid Gustafson, Ken Parry, and Sarah B. Proctor-Thomson for contributing data for this study. Finally, I would like to thank all my friends and family who supported me selflessly. I have the deepest appreciation for the invaluable emotional support, encouragement, and understanding I received from Chetan and from my dear family, Supriyo, Supriya, Tani, Jordan, and Raja.

TABLE OF CONTENTS

Page

Page

Page

Page

Page

LIST OF FIGURES

Introduction

The Importance of Leadership in Organizations

Much like a conductor of an orchestra, a leader plays a vital role in an organization. Organizations are structural units made up of leaders and followers who by their varied roles and position levels accomplish objectives essential to the continued improvement of the organization (Kanungo & Mendoza, 1996). Organizations need leaders to mobilize followers towards a shared sense of vision, accountability, ownership and values that are based on trust and integrity (Allen, Bordas, Hickman, Matusak, Sorenson, Whitmire, 1998; De Pree, 1987; Kanungo & Mendoza, 1996). In addition, organizations call for leaders to foster a work environment that facilitates learning (Allen et al., 1998), integrates and energizes organizational members (Fairholm, 1991), and develops future leaders (Bass & Avolio, 1993; Burns, 1978; DePree, 1987). As the workplace is changing to become more fast paced, technologically sophisticated, globally competitive, and diverse in nature, organizations have a greater need for managers and leaders who are able to function effectively in the resulting conditions while upholding the reputation and credibility of the organization (Kouzes & Posner, 2002). The multifaceted tasks, roles, and responsibilities of leaders have a significant impact on the life of an organization (Fairholm, 1991).

Leadership is a complex field. The research, discussion and practice of leadership over the last several years have helped develop the understanding of this field by great strides. However, much remains to be explored and investigated in terms of different aspects of leadership that contribute to the effectiveness of the leader within the organization.

Leadership Theories

A great deal of literature on leadership has examined various determinants of leader effectiveness. Researchers historically have been guided by a number of theoretical perspectives. Leader effectiveness has been measured by objective work group performance, attitudinal measures of subordinate job satisfaction and motivation, and perceptual measures that address leaders' qualities, behaviors, and abilities (Yukl, 1981, 1994). Although, discussing all the theories and related findings as they pertain to leader effectiveness is beyond the scope of the present study, an attempt will be made to highlight the theories that have played a large role in laying out the groundwork for further studies, thus contributing to the field's continued efforts to develop a finer understanding of the determinants of leader effectiveness within organizations.

Trait Theories. Trait, behavioral, and situational theories have dominated leadership research (Yukl, 1981). Several theorists (e.g. Mann, 1959; Hogan & Hogan, 2001; Stogdill, 1948, 1974) have attempted to explain leadership in terms of personality or character. Early trait theorists believed that there were distinct traits that separated an effective leader from an ineffective one. An extensive review of the trait studies conducted during the early period of trait research (1904-1947), revealed there were inconsistencies in the findings that had otherwise suggested that effective leaders have various traits in common (Stogdill, 1948). A second review of trait studies from a later period (1949-1970) reported positive findings for trait patterns when the situation of the leader was considered (Stogdill, 1974). Stogdill's (1974) findings led him to assert that the association of certain traits with effective leadership was a function of the leaders' situations, thus leading research into a new paradigm, which recognizes that "certain traits increase the likelihood that a leader will be effective, but they do not guarantee effectiveness, and the relative importance of different traits is dependent on the nature of the leadership

situation" (Yukl, 1981, p. 70). Trait research has become more productive in recent years (Yukl, 1994) lending to general acceptance of characteristics such as energy level, stress tolerance, self-confidence, internal control orientation, emotional maturity, and integrity as attributes of successful leaders (Yukl, 1981).

Behavioral Theories. The behavioral perspective in leadership research began with the studies at the Ohio State University and at the University of Michigan in the 1950s and lasted for the next 30 years. These studies emphasized initiating structure (task-orientation) and consideration (relationship-orientation) as two broad categories of observable behaviors that distinguish effective from ineffective leaders (Yukl, 1981, 1994). Empirical studies have associated high initiating structure with increased worker performance, but lower worker satisfaction, higher turnover, and increased grievance incidents. On the other hand, leaders who are rated high on consideration were associated with increased worker job satisfaction, lower turnover, and decreased number of grievances, but lower worker job performance (e.g., Fleishman & Harris, 1962). Although the results of this study seemed promising, researchers have not been able to consistently produce the same results for other criteria besides job satisfaction and performance, which are also important to organizations (e.g., Yukl, 1971).

Like the trait theories, the behavioral approach also lacked a consideration of the situational or cultural elements that can influence leadership effectiveness (Fairholm, 1991). This deficiency is addressed in the situational approach to leadership research, where the focus is on "aspects of the situation that enhance or nullify the effects of leader's traits or behavior" (Yukl, 1994, p. 285). Three distinct situational perspectives will be discussed next.

Contingency Theory. Fiedler's (1967) contingency model of leadership posited "situation" to act as a moderator of the relation between a leader's traits or behavior and a

leader's effectiveness (Riggio, 2003; Yukl, 1994). By the use of the Least Preferred Coworker (LPC) method, leaders rate how they feel about working with their "least preferred" member of an organization at any point in their work history. The scores then are used to assess whether the leader is more oriented or motivated by task objectives or relationship development. A high score on the LPC means that leaders are less critical of that co-worker and more motivated by supportive and considerate relationships. A low score means they are more critical of that coworker, and hence more interested in task objectives. According to the model, effectiveness of the leader's behavioral style is moderated by the leader's relationship with the coworker, the level of task structure and the leader's power, which describe the situational control of the leader. Research has shown that task-oriented leaders tend to be effective in situations that are either highly favorable or highly unfavorable for the leader (Yukl, 1994). A highly favorable situation would be one where the leader-member relationship was good, the task was structured, and the leader had strong position power. An unfavorable situation would be the opposite circumstance for all three variables. Further, research has shown that relationship-oriented leaders do better when the situation is moderately favorable or unfavorable such as in a situation where the leader's control and influence are neither high nor low. Although some studies have supported the theory (e.g., Chemers, 1969; Fiedler, 1967), others have failed to do so (e.g., Vecchio, 1977). Researchers have concluded that there are various strengths and weaknesses to this model (Ayman, Chemers, & Fiedler, 1995). However, this model is considered to be an important contribution to our understanding of leadership effectiveness because it was the first theory to address both the characteristics of the leader and the situation.

Path Goal Theory. Path goal theory (House, 1971) is concerned with the effects of managers on subordinates. This theory addresses the situations in which specific leader

behaviors are likely to be effective in improving subordinate satisfaction, motivation, and work unit performance (House, 1996; Yukl, 1994). Leader behavior is classified under the two broad categories of initiating structure (directive and achievement oriented) and consideration (supportive and participative). Optimal leader behavior is contingent on the type of task and the characteristics of workers. For example, if the task is stressful and boring, relationship oriented behavior by the leader would be expected to increase subordinate satisfaction because the leader increases the "valence" of doing the job and the "expectancy" that the job can be done effectively. The model has received both support (e.g., House & Dessler, 1974) and criticism due to the mixed impact of situational variables on the leader's behavior and the outcome and due to methodological limitations and conceptual deficiencies (e.g., Schriesheim & Von Glinow, 1977). However, the model is still held in high regard because it provides a guide for future researchers to identify potentially relevant situational variables (Yukl, 1994).

Leader Member Exchange Theory. Leader member exchange (LMX) theory assumes that leaders behave differently toward different workers and it is the quality of the relationships between the leader and each follower that determines effectiveness of the leader (e.g., Graen, Novak, & Sommerkamp, 1982b; Graen & Uhl-Bien, 1995). LMX has been found to predict subordinate satisfaction (e.g., Graen et al., 1982b; Scandura & Graen, 1984), performance (e.g., Dansereau, Alutto, Markham, & Dumas, 1982), intention to quit (e.g., Vecchio, 1982), and quality of leader-member relationships over a period of time (e.g., Liden, Wayne, & Stilwell, 1993). However, Schriesheim, Castro, and Cogliser (1999) reviewed 147 studies and concluded that the theory has various shortcomings and much further refinement of LMX theory is needed.

New Leadership Theory. More recently, a new genre of leadership theory has evolved that focuses on "exceptional" leaders alternatively referred to as "charismatic,"

"transformational," or "visionary" (Burns, 1978; Bass, 1985; House, 1977; Shamir, House & Arthur, 1993). The charismatic and transformational leadership literatures indicate that these types of leadership have an extraordinary impact on the follower, such that leaders change the needs, values, and perspectives of the followers and energize individuals in setting higher standards, embracing change, and becoming innovative (Shamir et al., 1993; Riggio, 2003). Further, leaders who are charismatic or transformational are able to draw others toward them and are able to influence and inspire others by appealing to their emotional needs and ideologies. They clearly communicate a vision that promises to change the present conditions for the better. They heighten employees' self-confidence and motivation by setting high expectations for everyone while treating every person as an individual. For example, research has found that leaders exhibiting transformational-type behaviors elicit superior results from subordinates in areas such as work performance (Bass, 1985; Elenkov, 2003; Howell & Avolio, 1993; Jung & Avolio, 2000; Kirkpatrick & Locke, 1996; Mackenzie, Podsakoff, & Rich, 2001; Shea & Howell, 1999); higher order motives (Sparks & Schenk, 2001), and attitudes and perceptions (Kirkpatrick & Locke, 1996; Waldman, Bass, & Yammarino, 1990; Yammarino & Bass, 1990).

Integrity in Leadership

Several researchers have addressed a topic that has been popularly referred to as the "dark side of leadership" (Conger, 1990; Hogan & Hogan, 2001; Howell & Avolio, 1992; Kets de Vries, 1986), which essentially brings attention to leaders' potential to produce negative outcomes as well as positive. Exceptional leaders that are described as charismatic and transformational are "celebrated as heroes of management" because they can revitalize followers to achieve high results during crises and change (Howell & Avolio, 1992). However, even leaders that have effective attributes and styles can use unethical tactics to expedite optimal

results that ultimately cause harm to the organization and its members in the long run (Howell & Avolio, 1992). The contention that the impact of the leader can vary by the ethical or unethical undertone of the leader has become a core concern for today's organizations (Allen et al., 1998).

Additionally, there has been a recent line of research that has focused on identifying the negative attributes associated with incompetent leaders. According to Hogan and Hogan (2001), negative traits of leaders or managers are seldom detected during the hiring process because many of these traits "coexist with strong social skills" (p. 50). The authors created the Hogan Development Survey (HDS), an 11 scale / 154 item instrument, to assess "dysfunctional dispositions" of managers that only begin to show after they have been in their positions for a period of time. The authors assert that the "dysfunctional dispositions" ultimately lead to ineffective leadership.

Research over the last three decades has addressed the importance of integrity in leadership (e.g., Burns, 1978; Fairholm, 1991; Posner & Shmidt, 1984; Vitell & Davis, 1990). Integrity has been conceptualized in various leadership theories as a central component of effective leadership behavior (Bass, 1990; Bass & Stedlmeier, 1999; Conger, 1990; Kirkpatrick & Locke, 1996; Kouzes & Posner, 1993). For example, integrity has been identified as a dimension of charismatic leadership behavior (Kramer, 1996). Additionally, leadership models have posited an association between a follower's perception of leader integrity and the follower's engagement in selfless acts (Mayer, Davis, & Schoorman, 1995).

Researchers and practicing managers have long been in search of the qualities and behaviors that augment leaders' effectiveness (Bass, 1985; Yukl, 1981, 1994). Studies have found that integrity is just as important for managers to have as abilities such as intelligence and competence (Stogdill, 1981; Hogan, Curphy, & Hogan 1994). Integrity has been conceptualized

as closely linked to the development of trust, such that the higher the perceived leader integrity, the more trust employees have in their managers (Mayer et al., 1995). The impact of higher trust has been documented in higher levels of work performance, job satisfaction, and organizational commitment and lower levels of intention to quit (Dirks & Ferrin, 2002). Some researchers have hypothesized that leaders' unethical behavior has various impacts on the organization such as decline in worker performance, organizational reputation, and subordinates' trust in their managers (Bass, 1990; Becker, 1998; Hogan & Hogan, 2001).

Keeping pace with the views of researchers on this topic, organizations are increasingly coming to realize that integrity is important to have in leadership in order to create a valued institution that can survive over time (Parry & Proctor-Thomson, 2002b; Goldsmith, Greenberg, Robertson, and Hu-Chan, 2003; Magill & Prybil, 2004). Corporate scandals and leader corruption within benchmark organizations, such as Enron, Arthur Anderson, and WorldCom, have led to a widespread uncertainty among the public about the moral solvency of today's organizations (Magbill & Prybil, 2004; Mitchell, 1993). There is a growing realization magnified by corporate debacles that unethical business practices can have real, costly, and sometimes irreversible consequences that affect not only the organization, but also other constituents such as employees, shareholders, business partners, customers, and even the general public (Premeaux, 2004).

In the midst of a period where the perception is that much of the business community is tainted (Magbill & Prybil, 2004), there is a need to address concerns about ethical leadership and revive individuals' confidence in their leaders' and organizations' ethical practices. There is some evidence that organizational members want to see integrity take priority over other demands of the organization. For example, one study discovered that peers, superiors, and

subordinates consider integrity to be the most essential quality to have in executives, middle managers, and supervisory managers (Posner & Schmidt, 1984). Additionally, surveys of subordinates have produced results that rank managers as having the most influence on the subordinates' own ethical behavior (Arlow & Ulrich, 1988; Brenner & Molander, 1977; Morgan, 1993; Posner & Schmidt, 1984).

Although research has highlighted the importance of integrity in leadership, a lack of standardized tools for the measurement of the leader integrity construct has impeded research examining the role of leader integrity on various organizational outcomes (Craig & Gustafson, 1998; Simons, 1999; Conger & Hunt, 1999; Bass & Steidlmeier, 1999). For example, Vitell and Davis (1990) were among the first to document an empirical relation between perceptions of managers' ethical behavior and subordinates' job satisfaction. They studied the complex role of leaders' ethical behavior by using a two-item global measure for ethics, which were: "MIS managers in my company often engage in behavior that I consider to be unethical" and "MIS managers in my industry often engage in behavior that I consider to be unethical." Results showed there was a significant positive correlation between perceptions of managers' ethical behavior and subordinates' job satisfaction.

Another study produced a global measure of perceived ethical behavior, but combined three factor analytically derived scales consisting of integrity, trust, and self-serving behavior into one scale. The ethical measure consisted of items such as "accepts responsibility for own actions," "considers the ethical implications of actions," and "will take advantage of others to accomplish own goals" (Morgan, 1993). The study found a positive relation between subordinates' perceptions of their managers' ethics and the regard in which they held their managers as leaders. Further, the relation was found to be more robust for subordinates than for

peers. Interestingly, the study also found an inverse relation between subordinates' perceptions of managers' ethical behavior and managers' salary.

More recently, Kramer (1996) found that leader integrity was significantly correlated with subordinates' satisfaction with supervision, organizational integrity, teamwork, sense of empowerment, and perceived quality of leadership. Although all of these studies attempted to empirically examine the impact of ethical or unethical behavior on the organization, no study used a scale that was designed to specifically measure the construct underlying leader integrity. In addition, none of the measures used in these studies had undergone stringent psychometric evaluation, which allows investigators to establish criteria for comparison across studies.

The Perceived Leader Integrity Scale

In an attempt to address the obvious need for an instrument for research and practical application, Craig and Gustafson (1998) designed and developed the *Perceived Leader Integrity Scale* (PLIS). The PLIS evaluates the perceptions of employees about their leaders' ethical and unethical behavior in the workplace. Several features of the original PLIS are worth noting. One feature of the original PLIS was that respondents rate their immediate supervisor. By having responses focused on one individual, the PLIS can be utilized as a feedback, coaching, or succession-planning tool. In addition, the PLIS was designed for broad applicability by including only items that would be relevant to nearly any type of organization.

Another feature of this instrument is that its development considered supererogatory as opposed to obligatory acts with respect to ethical and moral behavior (Craig & Gustafson, 1998). Because there are some acts that can be deemed as commendable but not required, individuals are not morally obligated to partake in those acts. For example, running into a house that is on fire to save individuals that might still be inside is not generally required of a civilian bystander,

but would be considered morally commendable if done. By having only items geared towards either morally required or clearly unethical behavior, confusion about supererogatory actions is avoided.

Simons (1999) proposed the notion of "behavioral integrity" as a match between "espoused and enacted" moral values (p. 89). Others have gone further and maintained that the focus in research should be on the ethical behavior of the manager as perceived by others and not on the self-perceived integrity of the leader (Parry & Proctor-Thomson, 2002b). There has also been consensus among researchers that organizations need to have leaders who can make obvious that their business practice and decision-making processes are ethical (Mortenson, Smith, & Cavanagh, 1989; Minkes, Small, & Chatterjee, 1999). Further, expansion of organizations into the global market has heightened the demand for leaders to "adhere to the highest levels of integrity and ethics in the operation of their organization" (Goldsmith et al., 2003, p. 220). Leaders must also be able to promote organization-wide ethical practice by modeling integrity in their own actions (Goldsmith et al., 2003). Enacting integrity, therefore, has become not just an important criterion for organizations, but has also become a focus of research (Parry & Proctor-Thompson, 2002b; Simons, 1999).

The PLIS instrument has opened up possibilities for research related to leader integrity that were not available previously. Tedious steps were taken to develop this scale. The PLIS development began with an initial pool of items based on seven categories (training and development, resource/workload allocation, truth-telling, unlawful discrimination, compliance with policies and procedures, maliciousness, and self-protection). These categories were used to restrict items to occurrences most likely to be common across organizations. Additionally, six items were included to measure global perceptions (e.g. "is evil").

Craig & Gustafson (1998) demonstrated that the PLIS was both a reliable and valid instrument. They found evidence for high internal consistency (Cronbach's alpha > .97) as well as convergent and discriminant validity.

The PLIS was also designed to serve as a global measure of perceptions. Craig and Gustafson (1998) found that there was a positive relation between the behavioral items in the scale and global impressions, accounting for 81% of the variance in those overall perceptions. Craig and Gustafson also tested for differences due to whether respondents were rating the leader's behavior towards themselves as opposed to their coworkers. Their results indicated that leaders' behavior toward coworkers did not contribute uniquely to overall impressions of integrity. An exploratory factor analysis (EFA) determined that the PLIS was unidimensional.

Since the development of the PLIS, the instrument has been used to assess relations between perceived leader integrity and follower job satisfaction as well as turnover intentions (Craig & Gustafson, 1998), transformational and transactional leadership behaviors, satisfaction with leadership, perceived leader effectiveness, follower extra effort, follower motivation, organizational effectiveness, and bottom-line achievement by the organization (Parry & Proctor-Thomson, 2002b). These studies found that perceived leader integrity correlated significantly with these leadership and effectiveness measures, thus providing empirical support for what was only theoretically suggested previously (Craig & Gustafson, 1998; Parry & Proctor-Thomson, 2002b).

The PLIS has been useful in measuring perceptions of leaders' ethical behavior and in establishing empirical support for the association between leader integrity and a broad range of organizational outcomes. One limitation of the PLIS has been its focus on subordinate perceptions of leader integrity. Because members at different levels of the organization can have

varied perspectives, which can be useful to understand the role and impact of leaders' integrity (Parry & Proctor-Thomson, 2002b; Posner & Schmidt, 1984), the items in the PLIS were later tailored by the authors to be appropriate for use as a multisource or "360 degree" assessment tool. The newest version of the PLIS is referred to as PLIS-360 for the purposes of this paper. For example, the item "would limit my training opportunities to prevent me from advancing" was altered to "would try to hurt someone's career because of a grudge." A second limitation of the PLIS is that the scale has only been available in English. As a result, Arredondo (2004) developed a Spanish version of the PLIS-360, the psychometric evaluation of which is the focus of the current study.

Scale Translation. The Spanish version of the PLIS was developed to facilitate leader integrity research in Spanish speaking countries as well as among Spanish speaking subpopulations in the U.S. A back-translation procedure (Brislin, 1970) was used to adapt the English version of the PLIS into Spanish. Seven bilingual subject matter experts, representing both Mexico and the U.S., served as either forward-translators or back-translators, with no individual serving as both. The forward-translators adapted the instrument from its original English (source) language into the Spanish (target) language. The back-translators then translated the target version back to the source language. The investigators involved in the adaptation process compared the two versions of the source language for discrepancies. Any differences between the two versions were discussed and addressed by modifying the target version and undergoing the back-translation process again. The target items were kept when there was agreement among the investigators that the two source versions were identical or comparable. Hambleton and Patsula (1999) suggested that, "format suitability, restrictive time limits, unclear directions, inappropriate content, and more can still be problematic even with a good literal

translation." For that reason, close consideration of various aspects of the content was given, such as accessibility of the vocabulary and emotional load. For example, in item 26 ("would engage in sabotage against the organization"), the word "sabotage" was replaced with the word "damage" and the item was translated as "se involucraría en acciones que dañen la organización." This approach retained the original meaning of the word "sabotage" but used an alternative word that is more likely to be associated with such behavior in the workplace in Hispanic cultures. Another consideration was that the emotional load of translated words should be equivalent to avoid a decrease or increase in the emotional strength of the words relative to the original version. For example, in item 7 ("is evil"), the Spanish word for "evil" was determined to be inappropriate because it is usually reserved for supernatural entities and not applied to humans. As a result, the item was initially translated into "es malo(a)" (is a bad person). Upon closer examination, "es malo(a)" was later reconsidered and changed to "es perverso." Although "perverse" is not a direct translation of "evil," the literally closer alternative "malo(a)" (bad) would have lowered the emotional load of the original "evil" in the English item.

Back-translation has been reported to be an effective technique for adapting test instruments, but several problematic factors can still affect the translation quality (Brislin, 1970; Hui & Trandis, 1985) and compromise the interpretability of the adapted test. In order to conclude that the two versions are measuring the construct (e.g. perceived leader integrity) equivalently, a common metric must be established (Cronbach & Furby, 1970).

Another factor that must be considered is whether the construct is relevant to the different cultures equivalently. In other words, it must be determined whether the construct of interest is universal or is culturally specific, which if not established by investigating psychometric

equivalence of the tests can lead to potential bias in the results and interpretations of later research findings (Hui & Trandis, 1985; Hulin, 1987).

Measurement Equivalence

Because one goal of the Spanish version of the PLIS is to facilitate cross-cultural comparisons, it is necessary to establish the measurement equivalence of the two language versions of the instrument. Measurement equivalence is obtained "when the relations between observed scores and latent constructs are identical across subpopulations" (Drasgow, 1984, p. 134). The idea is that subjects who have the same standing on a latent construct (e.g. leader integrity), but are sampled from different subpopulations (e.g. American versus Mexican) should have the same expected observed score on a measure of that construct. Therefore, to be able to compare two groups for similarity or differences, the scale on which they are measured has to be invariant across the two groups (Drasgow, 1984).

Establishing measurement equivalence is important to further the field of psychology (Reise, Widaman, & Pugh, 1993). For example, several studies have examined the stability of measurement scales across different cultural groups (Ellis & Mead, 2000; Guerrero, 2001; Weber, 1996); occupational groups (Drasgow & Kanfer, 1985; Idaszak, Bottom, & Drasgow, 1988) and rater groups (Mount, 1984; Maurer, Raju, & Collins, 1998; Swander, 1999; Facteau & Craig, 2001).

As discussed earlier, there can be difficulties with test instruments such that a construct may not be equivalent across two groups (Vandenberg & Lance, 2000) or that adaptation or translation of items may not necessarily yield comparable tests (Brislin, 1970). For this reason, researchers are increasingly applying advanced statistical techniques (e.g. item response theory or IRT) to establish the equivalence of translated instruments (Weber, 1996; Ellis & Mead, 2000;

Guerrero, 2001). However, more work by investigators is needed to establish measurement equivalence in the leadership domain in order to draw meaningful conclusions about different groups (e.g., Zagorsek, Jaklic, & Stough, 2004).

Methods of Assessing Measurement Equivalence. Measurement equivalence can be evaluated by using approaches such as differential item functioning (DIF) (Ellis & Mead, 2000). The term *differential functioning* (DF) refers to differences in an instrument's statistical properties, at the item or test level, in two different groups that are matched on a latent construct (θ) (Millsap & Everson, 1993). DIF in translated instruments reflects variation between groups in conceptualizing the items in a scale (Swander, 1999). These differences can be linked to translation errors or there may be cultural and linguistical differences that limit the use of the instrument for the target purpose.

Measurement equivalence can be assessed using statistical procedures such as the Mantel-Haenszel (MH; Mantel & Haenszal, 1959), Standardization (STDN; Dorans, 1989), logistic regression model (LR; Swaminathan & Rogers, 1990a), confirmatory factor analysis (CFA; Joreskog, 1971), and item response theory (IRT; Lord, 1980) (for detailed review of DIF models refer to Millsap & Everson, 1993; Potenza & Dorans, 1995). Although there are several methods of establishing measurement equivalence, some techniques have more advanced capabilities than others. Mantel-Haenszel, Standardization and Logistic Regression are powerful techniques for detecting item level equivalence and can be used for dichotomous (e.g. correct or incorrect) and polytomous (e.g. Likert type items) items. One problem common across all three of these techniques is their reliance on the total observed score as a substitute for the latent construct (θ), which can affect DIF detection by falsely indicating there is DIF when in actuality there is none (Millsap & Everson, 1993). In other words, these techniques assume perfect

measurement, that there is no error in the scale. This is a weak assumption, which can lead to DIF detection errors.

Reise et al. (2003) suggested CFA to be an easy and user-friendly application. One advantage of CFA is that it can assess the equivalence between two factor structures (Vandenberg & Lance, 2000). However, some limitations also exist with the CFA method. First, CFA assumes that each measured variable is a linear function of the latent variable (Reise et al., 2003); therefore, any nonlinearity in the relations between items and factors can result in imprecision in the detection of DIF. Second, at the item level, CFA can assess slope equivalence, which is similar to identifying the discrimination (*a*) parameters in IRT models, but does not assess inequivalence of difficulty (*b*) or guessing (*c*) parameters of IRT models (see Maurer, Raju & Collins, 1998; Reise et al., 1993; see Meade & Lautenschlager, 2004 for discussions of differences between CFA and IRT).

Item Response Theory

IRT models have become quite popular among psychometricians and measurement experts over the past two decades (Hambleton, Swaminathan, & Rogers, 1991). Researchers have used IRT to assess the measurement equivalence of translated tests (e.g. Drasgow & Hulin, 1990; Ellis, 1989; Ellis & Kimmel, 1992; Hulin et al., 1982) and have suggested that this technique should be applied to all translated tests to establish measurement equivalence (Ellis & Mead, 2000). IRT models have several desirable features (Hambleton, et al., 1991; Park & Lautenschlager, 1990). One distinct feature of IRT is that item and test characteristics are sample invariant, which means that item parameters are independent of the range of ability present in the sample from which the parameters are estimated, allowing DIF to be detected independently of any true population differences (Hambleton, et al., 1991).

Briefly, IRT describes the relations between items and the constructs they measure in terms of the probability of specific item responses, which can be graphically illustrated as an item response function (IRF) or item characteristic curve (ICC). The latent construct is referred to as theta (θ). The y-axis of an ICC plot represents the respondents' expected responses to a given item and the x-axis represents θ. After estimating item parameters, the detection of DIF essentially reduces to the process of comparing item parameters or their resultant ICCs (IRFs). ICCs can be compared using any of several approaches, including comparing the item parameters directly or calculating the area between the two ICCs that are being compared. If the area or the differences between the two-parameter sets are not statistically different from zero, then that is an indication that there is no DIF (Hambleton et al., 1991).

Most IRT models are governed by some common assumptions that must be verified before it can be determined that IRT is appropriate for evaluation of a particular test. One assumption is that the expected probability of an individual's response to an item is a function of that individual's ability and the characteristics of the item. Specifically, the higher respondents' standing on the latent construct, the higher the probability of positive responses. Most IRT models also assume unidimensionality, which means that the test is only measuring a single construct (Hambleton et al., 1991). Factor analysis is a widely used and powerful technique for assessing the dimentionality of an instrument (Swander, 1999). An established approach to assessing the factor structure of newly developed or adapted scales is to conduct an exploratory factor analysis (Hambleton, Swaminathan, & Rogers, 1991). Lastly, the local independence assumption postulates that the responses to any two items are statistically independent from one another if the traits or abilities that influence test performance are held constant (Hambleton, et al., 1991). The assumption of local interdependence can be checked by examining the variance-

covariance matrix. The assumption holds if the off-diagonals of the variance-covariance matrix have a value close to zero, after partialling for θ. Local independence can be assumed if the unidimensionality assumption is verified to be true; however, local independence can also hold if the test is found to be multidimensional (Hambleton, et al., 1991).

An IRT model is selected once its assumptions are checked. Three models appropriate for unidimensional tests consisting of dichotomous response items are the one (1PL), two (2PL) and three (3PL) parameter logistic models (Hambleton, et al., 1991). The 1PL estimates only the *b* parameter, which denotes item difficulty in ability or trait type tests. Item difficulty is the point on the trait continuum at which individuals have a 50% chance of endorsing the item, which is more commonly the point where the ICC steepens more noticeably. The 2PL model additionally incorporates the *a* parameter, which corresponds to item discrimination. Adding *a* to the ICC allows the examination of item discrimination, which refers to how well the item can distinguish among individuals with different latent trait levels. Item discrimination is the maximum slope of the curve at ranges of the latent trait continuum near the *b* parameter value. The steeper the slope, the better the differentiation among individuals at different levels of theta near the *b* value. A higher *a* parameter would mean two individuals with slightly different levels of theta are more likely to give different item responses (presuming these theta values are near the item's *b* parameter value). A lower *a* parameter would mean people with different thetas may give the same item response because the item is less sensitive to theta level differences. The 3PL is similar to the 2PL but adds the *c* parameter, which estimates the likelihood of respondents getting an item correct by guessing. In other words, a nonzero lower asymptote is allowed for respondents who are able to guess the answer right at a low ability level. The *c* parameter is generally of interest only for dichotomous items scored as "right" or "wrong."

The above models apply to dichotomous items. Other models have been developed to work similarly for polytomous items. The graded response model (GRM) developed by Samejima (1969) is a two-parameter model that is similar to 2PL. The GRM assumes that the scale is a monotonically increasing (Likert-type) scale with ordered response categories (e.g. a four point rating scale).

Once a model is chosen, procedures such as marginal maximum likelihood or Bayesian estimation can be used to estimate θ and item parameters. Several fit indices are available to assess whether there is a good fit between model and data, though their interpretation has been a topic of some debate (Hambleton et al., 1991). Next, the ICCs of the two groups are compared.

Differential item and test functioning

Various approaches are available to compare the ICCs for indications of differential functioning at the item level such as Lord's chi square test and the area between curves (Hambleton et al., 1991). But these two methods do not allow detection of differential functioning at the test level. Raju, van der Linden, and Fleer (1995) proposed a framework they called DFIT for detecting differential functioning both at the item and test level. DFIT includes two DIF measures (NCDIF and CDIF) and one DTF measure. Non-compensatory differential item functioning (NCDIF) and compensatory differential item functioning (CDIF) are item level indices and differential test functioning (DTF) is a test level index. The relation between CDIF and DTF is illustrated through an example provided by Ellis and Mead (2000) that if a translated item is more difficult for the focal group than it is for the reference group and another item is less difficult for the focal group than it is for the reference group, then the two items can cancel each other out and the impact of DIF on the overall test score remains insignificant. Because investigators are often more interested in overall test scores than scores on individual items,

ensuring that there is evidence of scale level equivalence between two groups can suffice for some purposes. Additionally, because DIF has been found to exist in anywhere from 1.5% to 64% of items in adapted instruments (Budgell, Raju, & Quartetti, 1995), investigators may find it more practical to use DTF over DIF (Ellis & Mead, 2000).

The application of DTF, CDIF, and NCDIF can be differentially useful depending on the purpose of the study and the information it will be used for (Raju el al., 1995). DTF is useful when total test scores are used for determining effectiveness. CDIF is useful if items that would favor a focal group at some instances and reference group in others are required to be on the test. NCDIF is useful when item level responses are of concern. Because this PLIS has been constructed as a feedback tool for managers in organizations, both item and scale level scores are important; therefore NCDIF and DTF indices will be examined closely.

Present Study

The purposes of the current study are (1) to examine the measurement equivalence of the English and Spanish versions of the *Perceived Leader Integrity Scale* instrument and (2) to examine differences in leader integrity perceptions between the English and Spanish speaking samples.

Two categories of research questions are of primary interest to this study and are as follows:

Category 1: Measurement Equivalence

Research Question 1. Do the English and the Spanish versions demonstrate equivalent factor structures?

Research Question 2. Within each factor, are items equivalently related to their respective constructs in both English and Spanish versions?

Category 2: Comparison across cultures with different languages

Research Question 3. Where measurement equivalence does not hold, what is the form of the inequivalence found?

Research Question 4. Can measurement inequivalence (if any) be linked to known differences between U.S. and Mexican cultures?

Research Question 5. Where measurement equivalence holds, is there a significant mean level difference between the English and Spanish speaking workers?

Research Question 6. Where equivalence holds and perceptual differences exist, can those differences be linked to known differences between U.S. and Mexican cultures?

Method

This study empirically tested for measurement equivalence between the English and Spanish versions of the PLIS instrument with data collected in the United States (U.S.), in New Zealand (N.Z.) and in Mexico. Measurement equivalence was examined with a combination of exploratory factor analysis (EFA) and item response theory (IRT). Because the number of respondents can be an issue where the IRT graded response model is applied (Embretson & Reise, 2000), data from the U.S. and N.Z. samples were combined into one group and referred to as the US-NZ or the English speaking sample in this paper. Equivalence between the English versions of the PLIS administered to the U.S. and N.Z. samples was established before combining the data from these two samples. The English version of the PLIS was then compared with the Spanish version of the PLIS. The Spanish version was administered to workers in Mexico and this group is referred to as the Mexican or Spanish speaking sample in this paper.

Samples

The study used archival PLIS data that were collected from the United States (U.S.), New Zealand (N.Z.), and Mexico. Data used for the U.S. group were collected as part of the original development of the PLIS instrument in 1995 (Craig & Gustafson, 1998). The N.Z. group data came from a study conducted in 2002 (Parry & Proctor-Thomson, 2002b) that used a revised version of the PLIS that was applicable in gathering ratings from multiple perspectives around focal leaders. Particularly, managers at different levels of their organizations provided ratings of the perceived integrity of a focal leader subordinate to them. The data for the Mexican group were collected as part of a separate study of transformational leadership, for which the Spanish translation of the PLIS was created (Arredondo, 2004).

U.S. Sample. The United States sample contained responses from 377 subordinates who provided ratings of focal managers using the original English version of the PLIS. The sample was 57% male, with an average age between 40 and 49 years. Respondents came from field and student settings consisting of 55% university faculty, 24% university staff, and 21% employed university students. The ethnicity of the field sample was primarily Caucasian (97.23%). Ethnicity information was not available for the student sample. The students in the sample were either currently employed for at least 20 hours per week or had been for at least three consecutive months during the last two years.

N.Z. Sample. Data were collected as part of a large-scale study on perceived leader integrity and transformational leadership in New Zealand (Parry & Proctor-Thomson, 2002b). This study used a revised version of the PLIS (PLIS-R) where three items were excluded from the original version and replaced by three additional items. Some items were modified to be applicable to raters with varied relationships to ratees (e.g., superiors). The data pool included

1,354 superiors who each gave ratings of a focal leader subordinate to them. The sample was composed of 77.6% male and 22.4% female at a mean age range of between 40 and 55 years. The ethnicity of the sample was largely European (95%). Respondents were managers from organizations in the public and private sectors.

Mexico Sample. Data were collected as part of a study on transformational leadership in Mexico. As discussed earlier, a back translation procedure (Brislin, 1970) was used for the adaptation of the PLIS-360 into Spanish (Arredondo, 2004). Total respondents equaled 439. All participants identified themselves as Mexican. Samples were collected from two different manufacturing firms in Mexico.

Procedure

The participants in the U.S. sample provided data using a paper and pencil form of the PLIS instrument (see Craig & Gustafson, 1998 for details). Student participants received extra credit for their participation. The field participants, composed of faculty and staff employees, were given surveys accompanied by a cover letter from the dean of the college promoting participation in the study. Completed surveys were returned by campus mail to the researchers. The N.Z. participants also completed a paper and pencil form of the PLIS. They received a cover letter endorsed by national institutions relevant to their professions encouraging them to take part in the study. Completed surveys were returned in pre-stamped envelopes. The Mexico sample was collected using both a web-based survey and paper-pencil form of the PLIS, where 60% of the sample came from the web-based form and 40% from the paper-based form. Although data collected across the studies were collected using multiple administration methods, data collected with different techniques (e.g. paper vs. electronic) have generally been found to be equivalent (Donovan, Drasgow, & Probst, 2000). In all samples, the PLIS was administered to respondents

along with other questionnaires, which are outside the scope of this study. As mentioned earlier, the U.S. and the N.Z. respondents completed the English version of the PLIS, and the Mexico sample completed the Spanish version of the PLIS.

Measure

Perceived Leader Integrity Scale. For the purposes of this study, only items in the English versions that matched those in the Spanish version of the PLIS were included in the analyses; however, several items were retained for analysis even though subsequent revisions had resulted in substantive differences between the English and Spanish versions. This occurred because the Spanish version was translated from the newest English version for which data were not yet available. For example, the English item "lacks high morals" was compared to the Spanish item "has high moral standards" because data were not yet available on the newer English version "has high moral standards."

The following discussion will facilitate an understanding of how items were matched across the different versions of the PLIS. The U.S. sample used the English PLIS as originally developed by Craig and Gustafson (1998). The U.S. study was conducted in two phases where 77 items were administered to the working students and 43 items were administered to the faculty and staff employees (for a detailed review of the study refer to Craig & Gustafson, 1998).

The N.Z. study used a slightly revised version of the original PLIS (PLIS-R) where nine items were modified to get ratings from a range of observers, rather than just subordinates as had been the case with the original version. For example, the item in the original PLIS "Would deliberately exaggerate my mistakes to make me look bad when describing my performance to his/her superiors" was changed to "Would deliberately exaggerate my mistakes to make me look bad when describing my performance to others." Respondents were superiors who provided

ratings on their subordinate focal leaders (refer to Parry & Proctor-Thomson, 2002b for a detailed review of the study).

The Mexico study, as described earlier, used the Spanish translation of the most recent version of the PLIS, version PLIS-360. The PLIS-360 consists of 32 items designed to assess observer perceptions of leader integrity. Respondents were subordinates who rated the perceived integrity of their immediate superiors. Four items from the Spanish version were omitted from the current study (21, 79-81; a list of these items is provided in Table A3 of Appendix A). Items 79, 80, and 81 were excluded because these are the most recent items in the PLIS and no data have been collected yet on these items using the English version of the instrument. Item 21 was excluded because after conducting a subsequent DIF analysis, it appeared that item 21 was anomalous, and an inspection of the frequency distribution in the Mexican sample suggested that raters might have misread the item. Specifically, item 21 showed considerable negative skew in the Mexican data set but a positive skew in all the other samples. It is possible that the word "can" was misread as "can't" by many of the Mexican raters due to the small number of positively worded items on the scale. The remaining 28 items in the Spanish version were used in the analyses.

All respondents in the current study received similar instructions to indicate the extent to which each item describes the person they are rating. Items in the PLIS are short sentence fragments that might describe a leader (e.g. "has it in for me"; "would deliberately distort what I say"). Raters respond using a four point Likert type scale where 1 = not at all, 2 = somewhat, 3 = very much and 4 = exactly. The data sets used have scoring structured in such a way that higher scores on the PLIS indicate lower perceived leader integrity and lower scores indicate

higher perceived leader integrity. A list of all the items corresponding to the PLIS as used in the present study is shown in Tables A1, A2, and A3 of Appendix A.

Analyses

Measurement Equivalence

Factor Structure Assessment. The measurement equivalence of the English and Spanish versions of the PLIS was assessed using a combination of exploratory factor analysis (EFA) and item response theory (IRT). Preliminary to combining the U.S. and N.Z. samples to form a single "English" sample, EFA was used to establish that the factor structure of the English version used in New Zealand (Parry & Proctor-Thomson, 2002a) mirrored the unidimensional structure previously found in the U.S. data (Craig & Gustafson, 1998). Thereafter, EFA was again used to evaluate the factor structure of the Spanish version used in Mexico (Arredondo, 2004) to see if I could replicate the one factor solution established as most appropriate for the English version.

EFA was conducted using principle axis factor (PAF) analysis and the promax oblique rotation. Because the primary analyses were conducted with IRT, the question of interest for the EFA portion of the analysis was to find out the dimensionality of the PLIS. PAF seeks to identify a set of latent factors that represent a parsimonious solution while still accounting for a reasonable amount of common item variance. Several criteria were examined to identify the best factor solution. The first criterion was that the proportion of variance accounted for by the first factor is greater than 20%, which has been suggested as adequate to meet IRT's assumption of unidimensionality (Reckase, 1979). A second criterion was to check the scree plot of the eigenvalues for a dominant factor. A third criterion was to examine the number of eigenvalues greater than 1.0. Finally, factor loadings and loading pattern were examined for interpretability.

Item Response Theory. The application of item response theory (IRT) allows us to gain information about a latent construct (e.g., perceived leader integrity) from responses to items because the responses to the items have a certain relationship to the latent construct. Differential functioning is said to occur if the relationship between responses to an item and the latent construct is not the same for different target groups. If the relationship is the same, then there is no differential functioning on the item level (DIF) between the comparison groups. There are many different ways to look at DIF. The DFIT framework proposed by Raju, van der Linden, and Fleer (1995) is one way that has been gaining in popularity in the recent years (Swander, 1999; Facteau & Craig, 2001; Guerrero, 2001). One advantage of DFIT is that it focuses on the expected differences in observed scores due to the DIF. In other words, DFIT operationalizes DIF as whether individuals at the same level of the latent construct would respond with a different item response as a consequence of group membership. In the current study, DFIT was used to assess whether each item equivalently relates to its factor for the English and Spanish speaking samples.

DFIT includes two DIF measures (NCDIF and CDIF) and one DTF measure. NCDIF is an item level index that considers each item separately in determining whether the item is differentially functioning. DTF estimates the scale-level difference between two groups that have the same standing on the latent construct, but come from different subpopulations. CDIF is an item-level index that estimates each item's net contribution to DTF, explicitly accounting for the "directionality" of differential functioning. Specifically, CDIF is the change in DTF associated with the removal of a given item from the scale.

As discussed earlier, most IRT models require that the assumption of unidimensionality is checked (Hambleton et al., 1991). By establishing that both the English and Spanish versions

of the PLIS are measuring a single construct, the instrument met the unidimensionality assumption of the graded response IRT model developed by Samejima (1969). Because the PLIS uses a Likert type scale, Samejima's two-parameter graded response model is the only IRT model appropriate for the current study.

The MULTILOG computer program (Thissen, 1995) was used to estimate person parameters (theta) for each rater and item parameters for each of the 28 items. This program implements the marginal maximum likelihood estimation procedure. EQUATE v2.1 (Baker, 1995) was used to estimate the linear transformation coefficients (slope and intercept) that are necessary for placing item parameters from different groups on the same metric so they can be compared. EQUATE implements the iterative linking procedure developed by Stocking and Lord (1983) to estimate these coefficients. Using the coefficients from EQUATE, the DFITPS6 computer program (Raju, 2003) was used to transform the item parameters to a common scale in order for the item parameters to be comparable.

The DFITPS6 program was also used to calculate several indices. The primary criterion applied to assess measurement equivalence was the NCDIF index, with the DTF index being a secondary consideration. Two conditions were applied for determining if there is significant NCDIF: (1) NCDIF values are greater than the .054 critical value and (2) NCDIF is accompanied by a χ^2 test with a significance level of $p < .01$. Both conditions must be met. If items were found to have significant DIF, then new linking constants were computed with those items removed. The new constants were then used to re-estimate the DFIT indices to determine whether more items demonstrated NCDIF.

Additionally, DTF was examined. The DFIT program applied the cut off for DTF, which was .054 multiplied by the number of items on the scale. For example, if the scale consists of 28

items, then the DTF cut off would be .054 x 28 = 1.51. In this case, DTF greater than 1.51 would be significant. Because the PLIS is intended to be useful as a managerial feedback instrument (i.e., both item and scale scores are important) and because a purpose of the current study is to evaluate the translation of the instrument into Spanish, any item that displayed significant NCDIF was examined for retranslation, even though DTF might have been nonsignificant.

Comparison across cultures with different languages

Because scores on the English and Spanish versions of the PLIS can only be compared on those items that are found to be invariant across the two groups, it was necessary to exclude any differentially functioning items from this portion of the analysis. Where measurement equivalence held, SPSS for Windows (version 12.0) was used to conduct a *t*-test for independent samples at the scale level. First, a mean score for each respondent was computed for the English and the Spanish speaking samples. Means of the items instead of sums were used to reduce any influence of missing data on the test scores. Next, a *t*-test for independent samples was used to compare the two groups on mean level of perceived leader integrity. Differential item functioning can occur due to various reasons (Gierl, 2005). Any measurement inequivalence detected in this study was interpreted in terms of either translation errors or cultural differences between the groups. Where translation errors could be ruled out, a speculative *post hoc* interpretation was offered based on the literature regarding cultural differences between the English and the Spanish speaking groups. Because there is no way to know in advance which of the 28 items might demonstrate inequivalence, it was impossible to predict which cultural factors might be active, and thus, what form such an interpretation might take.

Results

Preliminary Analyses to construct the English speaking sample

The English PLIS data consist of samples gathered from the U.S. (Craig & Gustafson, 1998) and from N.Z. (Parry & Proctor-Thomson, 2002b). In order to combine the samples into one group, the data from the N.Z. sample was first analyzed through an exploratory factor analysis (EFA) and the IRT-based DFIT procedure to establish measurement equivalence between the U.S. and N.Z. samples. Although, Parry & Proctor-Thompson (2002) had found the 28 item PLIS instrument to be an unidimensional construct, in constructing the English speaking sample, another factor analysis using the N.Z. sample was conducted to see if a two factor solution was called for. EFA revealed that the N.Z. sample did indeed show a single factor solution as had been previously established for the U.S. data. Specifically, the 25 items common to the two English versions were submitted to principal axis factor analysis. The first factor accounted for 34% of the variance, which is greater than the 20% suggested by Reckase (1979) as adequate for IRT's assumption of unidimensionality. An examination of the scree plot indicated that the first factor was dominant compared to the other factors (see Figure 2). The first unrotated factor had an eigenvalue of 8.51, a magnitude five times greater than the eigenvalue of the second factor (1.74). All except one item loaded .4 or greater on the first factor. In addition to performing a one factor extraction, a two factor extraction was also conducted. The two factor extraction resulted in high loadings of all the items on the first factor except for five items (item 14, 59, 67, 72, and 25) that loaded on the second factor. The small cluster of items that loaded on the second factor may be describing unethical behavior that is illegal or extreme. Furthermore, six items appeared to cross load on both factors and item 46 did not seem to load strongly on either of the two factors. A correlation of .71 between the first and the second factors indicated

that the two factors are not orthogonal, but are oblique. A comparison of the EFA results for a one and two factor extraction, between the N.Z. and the U.S. samples showed a similar pattern of factor loadings. Overall, these results led to the conclusion that a one factor solution appears to be the most parsimonious for the English version of the PLIS administered to the N.Z. sample, as is the case with the English version administered to the U.S. sample (Craig & Gustafson, 1998). For a summary of results, refer to Tables B1a-c and B2a-c of Appendix B. Scree plots are shown in Figures 1 and 2 to illustrate the variance that is associated with each factor.

Next, the DFIT procedure (Raju et al., 1995) was conducted to compare the groups that used the English version of the PLIS. The N.Z. group (N = 1,354) and the U.S. group (N = 377) were compared on the 25 items. Six items (10, 15, 19, 36, 52, 75) showed DIF, indicating that they functioned differently for the two English speaking groups. As result, responses to these six items from the N.Z. sample were not combined with the U.S. sample (i.e., data on those six items in the combined English sample consist only of U.S. responses). DFIT found no DIF for item 46 despite the parameters for the U.S. and N.Z. samples being very different. The reason may have been that DIF may be occurring at extremely high ranges of theta, which are not well represented in this sample. The standard errors for the parameter estimates for this item are high as a result of the small number of observed responses for the low response options. Data from the remaining 19 items that functioned similarly in the U.S. and N.Z. samples were combined (see Table E of Appendix E for NCDIF values for all the items). From this point forward in the paper, the combined sample of English data is referred to as the US-NZ or the English speaking sample. See Table D1 of Appendix D for the item parameters and theta estimates for a comparison of the two English versions of the PLIS.

Factor Structure Assessment using the English and Spanish speaking Samples

Spanish speaking Sample. Principal axis factor (PAF) analysis was used to assess the factor structure of the 28 items from the Spanish PLIS using the 2004 sample provided by Arredondo. A summary of results of the factor analysis is provided in Tables B3a-c of Appendix B. The first factor accounted for 58.52% of the variance, which is greater than the 20% suggested by Reckase (1979) as adequate for IRT's assumption of unidimensionality. An examination of the scree plot of eigenvalues indicated that the first factor was clearly dominant compared to the other factors (see Figure 3). Examining the magnitude of the difference between the eigenvalues of the first factor and the rest of the factors with eigenvalues greater than 1 revealed that the eigenvalue of the first unrotated factor (16.4) was ten times greater than the eigenvalue of the second factor (1.5) and 16 items greater than the eigenvalue of the third factor (1.0). In addition, the loading pattern indicated that all items loaded high on the first factor (factor loadings >.6), with one item loading .3. Because the results clearly met the previously set criteria, a single factor solution appears to be the most parsimonious for the Spanish version of the PLIS, as is the case with the English version.

English speaking sample. Principal axis factoring was next used to assess the factor structure of the 28 items from the PLIS using the English speaking (US-NZ) sample. Because factor analysis requires complete data (e.g., no missing values) and every observation in the New Zealand sample had missing values for the three items that were not administered (6, 53, and 61), EFA was conducted without including these three items. A summary of results is provided in Tables B4a-c of Appendix B. The first factor accounted for 38.87% of the variance, which meets IRT's unidimensionality criterion of greater than 20% of the variance (Reckase, 1979). The second criterion, a scree plot of eigenvalues indicated that the first factor was clearly dominant

relative to the other factors (see Figure 4). Furthermore, the first eigenvalue (9.33) was six times the magnitude of the second eigenvalue (1.69). In addition, both one factor and two factor extraction were conducted. When one factor was extracted, all the items loaded high on the first factor (> .4). When two factors were extracted, the correlation between the two factors was .71. Five items cross loaded on both the factors and six items (67, 59, 25, 14, 72, and 77) seemed to load higher on the second factor than on the first factor. An examination of the six items revealed that the behaviors assessed were more extreme in severity. A further study of the highest loading items in a one factor extraction and the lowest loading items in a two factor extraction indicated that those items that loaded high on the second factor in a two factor extraction differed only in the degree of severity of the unethical behavior and not in the kind of unethical behavior. In conclusion, a single factor solution appears to be the most parsimonious for the English version of the PLIS, based on the combined English speaking sample.

For the sake of completeness, another factor analysis was conducted using the three items previously left out of the analysis, but this procedure required that the missing values be replaced by the variable mean. Results similar to the ones above were found. Refer to Tables B4d-f of Appendix B for a summary of the findings.

In sum then, results indicated that the Spanish and the English versions demonstrate equivalent factor structures. Both the Spanish and the English speaking groups yielded a one factor solution.

Item Response Theory

Parameter estimation. The item parameters, which consisted of one *a* parameter (*a* = discrimination) and three *b* parameters (*b* = difficulty) for each item, were estimated using the MULTILOG program for Windows 7.0 (Thissen, 1995). The item parameters were estimated

separately for the English and Spanish speaking groups. Under the DFIT framework, theta estimates are required to be estimated for one of the groups (the "focal" group in DFIT terminology) so they may be rescored using the parameters from the other group in the assessment of DIF. Here, estimates of theta were calculated for the version of the PLIS with the larger sample, which in this case was the English speaking group. See Table D2 of Appendix D for the item parameters and theta estimates for a comparison of the two versions of the PLIS using the English and Spanish speaking groups.

Equating parameter metrics. The EQUATE program used an iterative procedure (Stocking & Lord, 1983) to link the estimated item parameters between the English (N = 1731) and Spanish (N = 439) speaking groups. Using the linear transformation coefficients (slope and intercept) from EQUATE, the DFITPS6 computer program (Raju, 2003) was used to transform the item parameters for the Spanish speaking group to be on the same scale as the item parameters of the English speaking group. Having the item parameters on a common scale allowed the item parameters to be compared. Because the transformation coefficients were not considered accurate if based on items with DIF, the linking process using EQUATE was conducted twice. Any items that showed DIF after the first DFIT analysis were excluded from the estimation process conducted the second time using EQUATE. DFIT indices were then re-estimated using the corrected transformation coefficients.

Differential functioning analysis

The DFITPS6 program (Raju, 2003) was used to estimate the NCDIF and DTF indices. NCDIF is the square of the mean difference on expected item score between individuals from two different groups who are at the same level of theta. Similarly, the square root of DTF is the unsigned mean difference on expected test score between individuals in two different groups

who are at the same level of theta (Facteau & Craig, 2001). The expected item score given theta is also referred to as true score in the DFIT framework. The metric of the true score is derived from the response format used for the items (e.g., 1 - 4 scale on a 4 point Likert type scale). As mentioned earlier, the significance of NCDIF was determined by (a) a significant χ^2 statistic $p < .01$ and (b) an index greater than the .054 critical value associated with a given item. Significant DTF would require (a) the index to exceed the critical value composed of .054 multiplied by the number of items in the test and (b) a significant χ^2 statistic $p < .01$.

Comparisons Across Cultures with Different Languages

Across all comparisons conducted for the 28 items common between the English and Spanish versions of the PLIS, a total of five items (17.9%) were identified as functioning differentially using the NCDIF index. DTF was found to be nonsignificant. The items with DIF were removed from the sample of items and new constants were estimated. Using these new equating estimates, the second run of the DFIT analysis again produced five items that demonstrated significant NCDIF; item 6: NCDIF = .114; $\chi^2 = 6072$ ($p < .01$), item 15: NCDIF = .113; $\chi^2 = 5088$ ($p < .01$), item 27: NCDIF = .476; $\chi^2 = 17091$ ($p < .01$), item 46: NCDIF = .106; $\chi^2 = 2101$ ($p < .01$), and item 59: NCDIF = .056; $\chi^2 = 2213$ ($p < .01$). See Table E of Appendix E for NCDIF values for all the items. Again, the DTF index was nonsignificant, DTF = .73848; $\chi^2 = 8492$, $p < .01$.

Research question three asked, “Where measurement equivalence does not hold, what is the form of the inequivalence found?” Using the estimated item parameters and constants, graphical representations of DIF results were used to illustrate the differences in the response patterns of the two groups at similar theta levels. These “true score functions” depict how difficult an item was and how discriminating the item was for one group compared to the other.

Figures 5, 6, 7, 8, and 9 are the true score functions for items with DIF. Figures 5, 8, and 9 display different levels of uniform DIF while figures 6 and 7 display nonuniform DIF. Note that nonuniform DIF describes a situation where the true score functions in the respective figures are not parallel (Zumbo, 2003).

Items 6 and 59, as represented in figures 5 and 9 respectively, discriminate similarly in the two groups as evident from the similarity of the *a* parameters. For item 6, $a = 2.31$ for the English speaking workers and equated $a = 2.30$ for the Spanish speaking workers; for item 59, $a = 3.33$ for English speaking workers and equated $a = 3.43$ for the Spanish speaking workers.

Items 15 and 27 are represented in figures 6 and 7, respectively. These items' true score functions cross, indicating that they discriminate differently in the two groups. For item 15, at moderately high levels of theta (0 to 2.0), it is more likely for the Spanish than for the English speaking workers to give higher (less ethical) ratings, whereas, at higher levels of theta (2.5 to 3.5), the Spanish speaking workers tended to give lower (more ethical) ratings than the English speaking workers. For example, at theta = 0.6, the Mexican workers had a higher probability of giving a less favorable rating than the US-NZ workers (e.g. where the item true score for the Mexican workers = 1.5 and the US-NZ workers = 1). However, at theta = 3.0, the US-NZ workers had a higher probability of giving favorable ratings than the Mexican workers (e.g. item true score for Mexican workers = 3.5 and the US-NZ workers = 4.0). This item is also more discriminating for the US-NZ sample ($a = 3.58$) than it is for the Mexican sample (equated $a = 1.71$).

For item 27, Mexican workers tended to give higher (less ethical) ratings at moderately low to high levels of theta (-2.0 to 2.0) than the US-NZ workers, but lower (more ethical) ratings

than the US-NZ workers at higher levels of theta (> 2.4). This item is also more discriminating for the US-NZ workers ($a = 2.31$) than it is for the Mexican workers (equated $a = 0.71$).

In conclusion, if these items were functioning equivalently, both groups would have been expected to give the same responses at the same levels of theta. Differences in the responses indicate that items either did not translate accurately and/or that there are subtle cultural differences between the groups.

Research question five asked, "Where measurement equivalence holds, how similar are perceptions of the English and Spanish speaking workers?" Because no significant DTF was found, averages were computed using all items on the scale to compare the mean ratings on the overall dimension for the two groups (see Figure 10 for an illustration of the mean difference between the English speaking [US-NZ] and Spanish speaking [Mexico] groups on the scale). An independent samples *t*-test analysis of the comparison between the English and the Spanish speaking samples was conducted. Levene's test for Equality of Variances indicated variances for the Spanish and the English speaking workers differed significantly from each other ($p < .05$), which called for the use of an unequal-variance *t* test. Results indicated that there is a significant mean difference in perceived leader integrity between the Spanish speaking group ($M = 1.53$, $SD = 0.66$) and the English speaking group ($M = 1.29$, $SD = 0.49$), $t_{528} = 7.06$, $p < .05$ (two-tailed), $d = 0.41$ and that the average is higher for the Spanish speaking group than it is for the English speaking group. Because lower ratings indicate higher perceived leader integrity, the higher average is an indicator of lower perceived leader integrity. Furthermore, according to guidelines proposed by Cohen (1988), the effect size of 0.4 would be considered small in magnitude. In the interest of completeness, the *t*-test was also repeated with items showing DIF excluded, which produced essentially the same result ($M_{\text{Spanish}} = 1.48$, $SD_{\text{Spanish}} = 0.68$;

$M_{English}$ = 1.30, $SD_{English}$ = 0.52), t_{533} = 5.06, p < .05 (two-tailed), d = 0.30.

Discussion

The discussion section is organized to make comparisons across cultures with different languages (e.g. US-NZ versus Mexico). First, *post hoc* explanations for item inequivalence and mean difference are considered, based on existing literature. Second, recommendations for re-translation are suggested. Third, implications of the current study are offered. Finally, limitations and future research are addressed.

Comparison across cultures with different languages

Comparison of the English and Spanish versions of the PLIS instrument found some NCDIF, but no significant DTF. Five items (6, 15, 27, 46, 59) were found to have significant NCDIF.

Post-hoc explanations. Research question four asked, "Can measurement inequivalence be linked to known differences between the U.S. and Mexican cultures? For item 6, the description of the leader in the English version was that of a person who "gives special favors to certain [pet] employees, but not to me." This item was translated into Spanish for which the back-translation was "shows favoritism." Differences in the difficulty parameters are exhibited in the US-NZ and Mexico comparison, with US-NZ workers (b_1 = 0.65, b_2 = 1.52, b_3 = 2.05) being less likely to perceive leaders as unethical than workers in Mexico (equated b_1 = 0.25, b_2 = 0.96, b_3 = 1.61). In other words, US-NZ workers have to be, on the average, 0.66 standard deviation above the theta mean, to give a rating of 1 or higher, 1.52 standard deviation above the theta mean to give a rating of 2 or higher, and 2.05 above the theta mean to give rating of 3 or higher, whereas workers in Mexico have to be 0.25, 0.96, and 1.61 standard deviations above the theta mean to give the same ratings. Where Mexican workers are predicted to give a response of 2 on

the rating scale for this item at theta score of 1.8, US-NZ workers to give the same response would have to score 2.2 on theta.

There may be a translation difference that is causing the item to be functioning differentially. The English version of the item implied favoritism to be directed personally towards the individual rating the leader, whereas the Spanish version did not make such a personal reference. This difference in translation of the phrases may have resulted in differing emotional loadings for the two groups.

Another explanation for the differential functioning of this item between the two groups may be cultural differences in how favoritism is perceived. In the Anglo-American culture, rules and regulations clearly prohibit employers from treating workers unequally based on race, gender, age, creed, or social status. These regulations carry weight in various aspects of decisions that managers have to make during the course of the worker's employment cycle. Additionally, workers in the U.S. are legally protected against retaliation from managers and coworkers for presenting a case against the manager for unfair treatment. It is the manager or the employer who has to provide evidence that appropriate measures have been taken to resolve the concerns of the employee with the complaint. Additionally, because in the United States the power distance between managers and workers is lower than in Mexico, workers can question the manager's behavior and ask for justifications based on reason. Therefore, favoritism may not be as adverse to U.S. employees as it is for employees in the Mexican culture. In the Mexican culture, where the power distance is substantially larger between managers and workers (Schuler, Jackson, Jackofsy, & Slocum, 1996), it is not considered to be respectful or appropriate for workers to question the decisions of managers nor can they speak in public against their managers (Stephens & Greer, 1995). Favoritism may be perceived more adversely by Mexican workers than by

workers in the United States because in Mexico, personal relationships carry into the work environment more than they do in the United States and favors are given based on whom you know and their status (Stephens & Greer, 1995).

Item 15 describes the leader as one who "Would use my performance appraisal to criticize me as a person" in the English version. This item was adapted to "would use feedback as an excuse to criticize others". An examination of the difficulty parameters indicates that US-NZ workers ($b_1 = 1.28$, $b_2 = 1.90$) are less likely than Mexican workers (equated $b_1 = 0.71$, $b_2 = 1.54$) to perceive leaders as unethical at moderate levels of theta, but that US-NZ workers ($b_3 = 2.22$) are more likely than Mexican workers (equated $b_3 = 2.46$) to perceive leaders as unethical at higher levels of theta. In other words, US-NZ workers have to be 1.28 standard deviation above the theta mean, to give a rating of 1 or higher, 1.90 standard deviation above the theta mean to give a rating of 2 or higher, and 2.22 above the theta mean to give rating of 3 or higher, whereas Mexican workers have to be 0.71, 1.54, and 2.46 standard deviations above the theta mean to give the same ratings.

One possible reason for the differential functioning of this item in the two groups may be a translation difference between the items. The English version refers to a formal process of providing feedback, whereas the Spanish version does not specifically make a reference to a "performance appraisal." This difference in wording could cause the phrases to have different emotional loadings for one group versus the other.

The differential functioning of item 15 may also be a result of cultural differences. One possible explanation could be that in the Mexican culture, leaders are expected to be more supportive and relationship-oriented (Drost, Ayman, & Chemars, 1983). Thus, feedback used as an excuse to criticize would be seen very negatively by those of the Mexican culture. As

Stephens and Greer (1995) pointed out, "Mexicans are far less tolerant of abrasiveness and insensitivity in managerial styles than are Americans" (p. 42). They also reported that Mexico is a "softer culture" where workers are more congenial towards one another, more sensitive to criticisms, and require more assurance than do workers in the United States (p. 43). Thus, it is apparent that Mexican norms for politeness are stricter than U.S. norms. As a result, saying anything negative in a direct and open manner to someone would be evaluated more harshly by Mexicans (Riding, 1985).

The Spanish version was created from the most recent English version for which there were no data available. The English version of item 27 was phrased "lacks high morals" and in the Spanish version, the same item was phrased "has high moral standards." This item was reverse coded before analysis. It seems likely that this item is functioning differentially because it was reverse phrased or positively phrased in the Spanish version. This item was one of only two items in the Spanish version of the scale that were worded positively which means more than 90% of the items were phrased to describe unethical behaviors. Reverse wording this item may have contributed to differences in how this item was read and rated by some or all of the observers.

Item 46 was phrased in the English version as "would treat me better if I belonged to a different ethnic group" and in the Spanish version back translated as "would treat people depending on gender, ethnic group or social status." Differential functioning might be explained by an obvious translation difference between the two items. The reason this item was translated differently is because the most recent English version of the PLIS makes a reference to gender as a reason for discrimination. During the translation process, Arredondo (2004) decided that adding "social status" would more effectively tap into the range of discrimination observed in the

Mexican workplace. As a result, the Spanish version addresses three types of discrimination as opposed to a single type addressed in the original English version.

Item 59 was phrased in the English version as "would blackmail an employee if (s)he thought (s)he could get away with it" and translated into Spanish as "would blackmail an employee if believed would not be caught." As shown in Table E of Appendix E, DIF was significant for the comparison of the US-NZ and Mexican samples on this item. An examination of the difficulty parameters indicates that US-NZ workers ($b_1 = 2.06$, $b_2 = 2.29$, $b_3 = 2.40$) are less likely to perceive leaders as unethical than are Mexican workers (equated $b_1 = 1.37$, $b_2 = 1.89$, $b_3 = 2.44$). In other words, US-NZ workers have to be 2.06 standard deviations above the theta mean, to obtain a rating of 1 or higher, 2.29 standard deviations above the theta mean to obtain a rating of 2 or higher, and 2.40 above the theta mean to obtain a rating of 3 or higher, whereas Mexican workers have to be 1.37, 1.89, and 2.44 standard deviations above the theta mean to obtain the same ratings.

Differential functioning of item 59 may be due to cultural differences in how adversely blackmailing is perceived. The following explanation is formed using Schuler et al.'s (1996) comparison of cultural values between workers in the U.S. and in Mexico. Based on their research, it appears that a manager blackmailing an employee would be considered more severely by Mexican workers than by workers in the United States. One reason may be the fact that the Mexican work culture is more family-oriented or paternalistic, in which the expectations are that managers would not only care for subordinates as for their own families, but also watch out for them and show concern for them. In return, subordinates are expected to show loyalty towards their superiors and follow directions from their leaders. In contrast, the U.S. is based on individualism, where workers value independence and have greater tolerance for confrontations

and adversarial relations with their respective managers. Evidence such as this may explain why it can be expected that a worker of the Mexican culture would give a less favorable rating to a leader who is known to blackmail employees than would a worker of the U.S. culture.

Mean differences in perceived leader integrity. The test score average for the group that used the English version of the PLIS was found to be slightly higher than the test score average for the group that used the Spanish version of the PLIS. Although computing the mean test score on the PLIS using the data from the U.S. sample revealed a similar mean test score as the US-NZ sample, it must be noted that the samples compared here may not be representative of their respective cultures.

Research question six asked if the mean difference between U.S. and Mexico could be linked to any known differences between these cultures. There is evidence that leadership differs across cultures (Bass, 1990). Factors that can be linked to cultural differences are management practices and values that organizations place on leadership and employee motivation (Drost & Von Glinow, 1998). Recently, as U.S. organizations have been setting up operations in Mexico, they are finding management and business practices in Mexico are different from their own (Stephens & Greer, 1995).

Drost and Van Glinow's (1998) description of the various attributes of leaders in Latin America such as Mexico will be used as a model to provide possible reasons why workers in the Mexican culture tend to perceive their leaders to be slightly more unethical than do workers of the U.S. culture. These authors suggested that there is less potential for "trust, openness, and the rational expression of feelings" (p. 6) in the Mexican culture compared to that found in cultures such as the United States. The authors further described corruption as an "ingredient of doing business" where the working individual must make a decision at each step whether to take the

"long frustrating honest route or the short efficient corrupt route (p. 9)." The authors also indicated that the behaviors of the Mexican leaders tend to make the employees susceptible to exploitation. Although it is reasonable to assume that dishonesty and unethical behaviors occur in any culture in varying degrees, there seems to be the suggestion that in Mexico, "typical business is conducted around the law, in personal loyalties, in bribes and fees, in tax evasion, in fraud, and in codes of honor" (p. 9). The reasons provided so far are simply speculative, but may provide some rationale for why perceived leader integrity is lower in Mexico than in the United States.

Retranslation recommendations for the Spanish Version of the PLIS

Items with DIF linked to possible translation errors should be considered for retranslation prior to using item level scores to provide targeted feedback to managers or to make cross-cultural comparisons. Some suggestions are made with reference to retranslation of items 27 and 46. Item 27 "has high moral standards" was reversed in the Spanish version, but is more congruent to the newest English version. Because no data were yet available on the newest English version, future comparisons between the Spanish version and the newest English version may not find the same DIF shown here. Item 46 "would treat people depending on gender, ethnic group or social status" is addressing three types of discrimination as opposed to one. This item in the Spanish version is triple barreled. Thus, phrasing the item with three types of discrimination requires respondents to reply with one answer to a question that has three components about which they may have differential feelings (Dillman, 2000). Once again, the Spanish version of this item is more congruent to the newest English version, but DIF could not be tested between these two versions because no data have been collected using the new English version. An obvious suggestion would be to revise this item by splitting the item into three individual items

in which the first item would be "would treat people depending on their gender," the second item would be "would treat people depending on the ethnic group they belong to, and the third item would be "would treat people depending on the social status they belong to." Items that did not have obvious translation errors should be used with caution; that is DIF items should only be used if the goal is to conduct a scale level analysis. However, DIF items should be removed if item-level scores are used to make personnel recommendations or to provide feedback to managers on specific indicators of behavior on the PLIS.

Implications

Establishing that different language versions of the same measure are on the same scale is central to cross-cultural research. Equivalent measures may provide interpretable test results relevant to target groups and meaningful comparisons between groups across different cultures (Hambleton & Kanjee, 1995). Consequently, the results of this study have a number of practical implications.

1. There apparently had been no studies prior to this one that had established measurement equivalence in the context of cross-cultural perceptions of leader integrity. Additionally, this study sought to consider both language and cultural influences as possible sources of measurement inequivalence.

2. This study replicated the one factor solution of the U.S. developed version of the PLIS and found that the construct remained unidimensional even after translations and revisions made to the measure.

3. This study adds to the growing body of research that has assessed item and test level equivalence by using IRT analysis. It is apparent that even the most careful forward and back translations cannot be assumed sufficient to obtain psychometrically equivalent measures.

4. This study shows the advantages of using DFIT to measure equivalence of adapted tests. Comparison of the two versions of the measure revealed that 17.9% of the items were differentially functioning at the item level. An appropriate practice when DIF is discovered is to either remove the items or to rewrite DIF items and retest them for DIF. However, constraints of time, cost and available participants might not make it feasible for repeated studies to be conducted by the same organization or research team every time DIF is found. If DIF is found and a new research study is impractical, then composite results of the measure can be used when test-level analysis (DTF) is found to be nonsignificant. In other words, the findings of this study suggest that, to the extent that these results are generalizable, scale-level scores on the PLIS can be meaningful indicators of perceived leader integrity for workers in Mexico that are directly comparable to scores on the English version of the instrument.

5. Although the focus of the post-hoc explanations was limited to the cultural differences between U.S. and Mexico, the results of this study indicate that the perceptions of Mexican workers can be compared to the perceptions of workers in the United States and New Zealand to derive meaningful correlates for comparative research as well as to investigate the value of integrity in leaders' behavior.

6. Implications of DIF must be considered when combining data from two samples (e.g., U.S. and New Zealand). A comparison of the U.S. and N.Z. groups revealed that DIF existed for item 10 "would risk me to protect themselves in work matters", item 15 "would use feedback to me to criticize me as a person, item 19 "has it in for me", item 36 "is vindictive", item 52 "is a hypocrite", and item 75 "would try to get people fired just because they don't like them." All these items appear to be referencing professional harm to the rater. It is possible that the use of superior ratings in the N.Z. sample and the use of subordinate ratings in the U.S. sample might

explain the DIF that was found. For example, in the N.Z. sample, it could be that there was no opportunity for the subordinate ratees to perform the behaviors listed in some of the items that showed DIF. An alternate explanation could be that DIF resulted from the minor wording differences between the two English versions. For example, item 15 used the word "feedback" in the N.Z. version instead of the term "performance appraisal" used in the U.S. version. DIF could also be a result of cultural differences between the two English speaking groups and in that case future researchers should be cautioned against routinely combining together these two groups or assuming that these two cultures are the same until further studies can be done to establish their measurement equivalence. The present study dealt with the DIF between the U.S. and N.Z. samples by only combining data for those items where equivalence held. For the items that showed DIF, I used only data from the U.S. sample when combining the two groups to form the English speaking sample. Even though removing the DIF items from the N.Z. sample did not drastically alter the inter-item correlations or the overall factor structure of the construct, it cannot be known for certain whether the construct domain might shift slightly when some items are absent for one sample but not the other. For a summary of results of the factor analyses of the English PLIS used by the N.Z. sample after the DIF items were removed, see Tables B2d-f of Appendix B.

7. Finally, as organizations increasingly move towards establishing operations in other countries, it will become increasingly important for businesses to have a keen understanding of cultural differences and how to train managers to perform effectively in cross-cultural and cross-national environments. Not becoming culturally knowledgeable of the differences in managerial practices can lead to unfavorable outcomes such as "extended periods of poor performance, disharmony, missed opportunity,…" (Stephens & Greer, 1995, p. 40). The use of translated

measures of perceived leader integrity that have established measurement equivalence can improve the understanding of these kinds of differences in leadership practices and provide a framework for leadership training and development around the needs and values of various constituents of the organization that managers will work with.

Limitations and Future Research

There are a number of limitations in the present study that should be addressed in future studies.

1. The current study used archival data; therefore, the researcher conducting this study had no control over collecting the data. However, it is inevitable that archival data is often the most feasible alternative in academic research for conducting a study utilizing the DFIT methodology due to the large sample sizes needed. As a result, the researcher was limited to information that was available from the archival data set and from published articles. For example, respondents across samples were assured of confidentiality; however, participants' actual confidence in that assurance is unknown. Future research is needed that would replicate the current study to extend its generalizability.

2. Only the items for which wording matched across the three samples were used in this study and nonmatching items were excluded from the analysis. For example, three items (79, 80, and 81) on the Spanish version were not included in the analysis because no data had yet been collected with their English counterparts. Although a comparison suggested that the inter-item correlations of the included items and the excluded items were similar for all the samples used in this study (see Table C1-3 of Appendix C), it is suggested that future research on perceived leader integrity use the most recent version of the PLIS measure to increase future opportunities for comparative research. Also, as discussed previously, items not analyzed could have affected

responses to items that were analyzed. For example, in comparing the English versions of the PLIS using the U.S. and N.Z. samples, six items (10, 15, 19, 36, 52, and 75) were found to have DIF. These six items were removed from the N.Z. sample before combining the U.S. and N.Z. data sets. Using the N.Z. sample, a comparison of the factor structures obtained when including and excluding DIF items revealed similar patterns of results (see Table B2a-f of Appendix B). The factor structures appeared similar whether the six items were included or excluded from the N.Z sample. A similar analysis was done using the English speaking sample (US-NZ) and results once again were similar when DIF items were included or excluded from this sample (see Tables B4a-c and B4g-i of Appendix B). Item 46 did not appear to contribute to the factor structure with or without the DIF items in either the N.Z. or the US-NZ samples. Previously, Parry and Proctor-Thomson (2002a) had come to a similar conclusion that this particular item had a very low correlation with the other items. As a result, they had left this item out of the factor analysis that they conducted.

3. Each group responded to items presented in different sets and different orders. Thus, if the position of an item on a scale influences responses to the item, then that is a potential alternative explanation for these findings. Future research is needed where different groups are administered the same test without systematic variation in the items or the order in which they are presented. Doing so would reduce competing explanations and increase the interpretability of the results.

4. The subpopulation that used the Spanish version consisted only of raters with a Mexican background. Given the diversity within the Spanish speaking population in the U.S. and elsewhere in the world, future studies should include Spanish speakers from varied backgrounds such as Spain, Columbia, Nicaragua, Cuba, and Puerto Rico.

5. Mexican respondents completed an electronic version of the PLIS, while English speakers completed a paper form. Although previous research has generally supported the equivalence of these two modes of administration (Donovan, Drasgow, & Probst, 2000), the design of the current study does not permit a direct examination of this issue.

6. Another issue to consider is that differences in the rater-ratee relationships across the three samples may have affected the ratings that were provided. The U.S. and Mexico samples consisted of subordinate ratings, whereas superior ratings were collected from the New Zealand sample. Previous research has consistently found that reporting relationship is not a source of measurement inequivalence in multirater leadership assessments (e.g., Facteau & Craig, 2001), but mean level differences among rating sources are common (Evans, 1995). Because managers in the New Zealand sample had discretion regarding whom they chose to evaluate, ratings may have been biased towards being lenient and positive. Bias in choice of rating target could have occurred if the managers selected individuals to rate whom they regard more highly. Differences in mean ratings could have also occurred if managers rating subordinates rated more leniently than did managers rating superiors as a result of differential opportunities to observe unethical behavior. Superiors may not have as much opportunity to watch the day to day unethical activities of the subordinates as do subordinates to observe superiors. On the other hand, subordinates have a greater opportunity to observe and be aware of the unethical activities of their managers. Although this limitation should be noted, it seems unlikely to account for the mean difference between the English and Spanish samples because no significant mean difference was found between the U.S. and the New Zealand groups even though the same difference in rater-ratee relationship existed in that comparison.

7. Finally, the post hoc explanations geared towards cultural differences were speculative and by no means definitive. Swander (1999) recommended that the only way to address this limitation is to give importance to the DIF itself so that the findings would allow researchers in the future to consider various cultures when constructing new tests or modifying existing tests.

Conclusion

Item response theory is a popular method for detecting measurement inequivalence (Hambleton, et al., 1991). Raju et al's (1995) DFIT framework is the only IRT method that can detect differential functioning at both the item and test level, which makes DFIT convenient for practical as well as for research use. The current study found 17.9% of the items to have DIF; however, with limited resources, retranslation of the items with DIF is not always an option. Because no DTF was found in the comparison of the English and Spanish versions of the PLIS, using the overall mean score for each individual as opposed to item level scores would be appropriate. One advantage of the PLIS is that it can be used to provide targeted feedback to managers. However, targeted feedback requires an item level interpretation, which in light of the non-compensatory differential item functioning (NCDIF) found in the present study should wait until the Spanish version of the PLIS is free from item level inequivalence. A post-hoc explanation for the items that showed inequivalence suggests that when using a translated test instrument, specific items should be evaluated for cultural relevance or translation error before forming decisions based on those items. Similarly, researchers who use translated measures, should evaluate the individual items before making comparisons based on those items.

References

Allen, K. E., Bordas, J., Hickman, G. R., Matusak, L. R., Sorenson, G. J., & Whitmire, K. J. (1998). Leadership in the 21st Century. In G. R. Hickman (Ed.), *Leading Organizations: Perspectives for a New Era* (pp. 572-580). Thousand Oaks: Sage Publications.

Arlow, P., & Ulrich, T. A. (1988). A longitudinal survey of business school graduates' assessments of business ethics. *Journal of Business Ethics, 7*, 295-302.

Arrendondo, F. (2004). *La integridad del líder de empresa y su relación con el liderazgo.* Unpublished doctoral dissertation, Tecnológico de Monterrey.

Ayman, R., Chemers, M. M., & Fiedler, F. (1995). The contingency model of leadership effectiveness: Its levels of analysis. *The Leadership Quarterly, 6*, 147-167.

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

Bass, B. M. (Ed.). (1990). *Bass and Stogdill's handbook of leadership*. New York: Free Press.

Bass, B. M., & Avolio, B. J. (1993). *Improving organizational effectiveness through transformational leadership*. Thousand Oaks: Sage Publications.

Bass, B. M., & Avolio, B. J. (2000). Multifactor Leadership Questionnaire.

Bass, B. M., & Steidlmeier, P. (1999). Ethics, character, and authentic transformational leadership behavior. *Leadership Quarterly, 10*(2), 181-217.

Becker, T. E. (1998). Integrity in organizations: Beyond honesty and conscientiousness. *The Academy of Management Journal, 23*(1), 154-161.

Brenner, S., & Molander, E. (1977). Is the ethics of business changing? *Harvard Business review, 55*(1), 57-71.

Brislin, R. W. (1970). Back-translations for cross-cultural Research. *Journal of Cross-Cultural Psychology, 1*(3), 185-216.

Budgell, G. R., Raju, N. S., & Quartetti, D. A. (1995). Analysis of differential item functioning in translated assessment instruments. *Applied Psychological Measurement, 19*, 309-321.

Burns, J. M. (1978). *Leadership*. New York: Harper & Row, Publishers, Inc.

Chemers, M. M. (1969). Cross-cultural training as a means for improving situational favorableness. *Human Relations, 22*, 531-546.

Conger, J. A. (1990). The dark side of leadership. *Organizational Dynamics, 19*(2), 44-55.

Conger, J. A., & Hunt, J. G. (1999). Overview charismatic and transformational leadership: Taking a stock of the present and future (Part 1). *Leadership Quarterly, 10*(2), 121-127.

Craig, S. B., & Gustafson, S. B. (1998). Perceived leader integrity scale: An instrument for assessing employee perceptions of leader integrity. *Leadership Quarterly, 9*(2), 127-145.

Cronbach, L. J., & Furby, L. (1970). How we should measure "change"-or should we? *Psychological Bulletin, 74*(1), 68-80.

Dansereau, F., Alutto, J. A., Markham, S. E., & Dumas, M. (1982). Multiplexed supervision and leadership: An application of within and between analysis. In J. G. Hunt, U. Sekaran & C. A. Schriesheim (Eds.), *Leadership: Beyond establishment views* (pp. 81-103). Carbonale: Southern Illinois University Press.

De Pree, M. (1987). *Leadership is an art*. East Lansing: Michigan State University Press.

Dirks, K. T., & Ferrin, D. L. (2002). Trust in leadership: Meta-analysis findings and implications for research and practice. *Journal of Applied Psychology, 87*(4), 611-628.

Donovan, M. A., Drasgow, F., & Probst, T. M. (2000). Does computerizing paper-and-pencil job attitude scales make a difference? New IRT analyses offer insight. *Journal of Applied Psychology, 85*(2), 305-315.

Dorans, N. J. (1989). Two new approaches to assessing differential item functioning: Standardization and the Mantel-Haenszel method. *Applied Measurement in Education, 3*, 217-233.

Drasgow, F. (1984). Scrutinizing psychological tests: Measurement equivalence and equivalent relations with external variables are the central issues. *Psychological Bulletin, 95*(1), 134-135.

Drasgow, F., & Hulin, C. L. (1990). Item response theory. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology: Vol. 1.* (2nd ed., pp 577-636). Palo Alto: Consulting Psychologists Press.

Drasgow, F., & Kanfer, R. (1985). Equivalence of psychological measurement in heterogeneous populations. *Journal of Applied Psychology, 70*(4), 662-680.

Drost, E. A., & Von Glinow, M. A. (1998). Leadership behavior in Mexico: Etic philosophies-emic practices. *Research in International Business and International Relations, 7*, 3-28.

Elenkov, D. S. (2002). Effects of leadership on organizational performance in Russian companies. *Journal of Business Research, 55*(6), 467-480.

Ellis, B. E. & Kimmel, H. D. (1992). Identification of unique cultural response patterns by means of item response theory. *Journal of Applied Psychology, 77*, 177-184.

Ellis, B. E., & Mead, A. D. (2000). Assessment of the measurement equivalence of a Spanish translation of the 16PF questionnaire. *Educational and Psychological Measurement, 60*(5), 787-807.

Embretson, S. E. & Reise, S. P. (2000). Item response theory for psychologists. Mahwah: Lawrence Erlbaum Associates.

Evans, C. R. (1995). *Rating source differences and performance appraisal policies: Performance is in the "i" of the beholder.* Unpublished doctoral dissertation, The University of Guelph, Canada.

Facteau, J. D., & Craig, S. B. (2001). Are performance appraisal ratings obtained from different rating sources comparable? *Journal of Applied Psychology, 86*(2), 215-227.

Fairholm, G. W. (1991). *Values leadership: Towards a new philosophy of leadership*. New York: Praeger.

Fisher, R. A. (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika, 10*, 507-521.

Fiedler, F. E. (1967). *A theory of leadership effectiveness*. New York: McGraw-Hill.

Fleishman, E. A., & Harris, E. F. (1962). Patterns of leadership behavior related to employee grievances and turnover. *Personnel Psychology, 15*, 43-56.

Gierl, M. J. (2005). Using dimensionality-based DIF analyses to identify and interpret constructs that elicit group differences. *Educational Measurement: Issues and Practice, 24*(1), 3-14.

Goldsmith, M., Greenberg, C. L., Robertson, A., & Hu-Chan, M. (2003). *Global leadership: The next generation*. Upper Saddle River: Financial Times Prentice Hall.

Graen, G. B., Novak, M. A., & Sommerkamp, P. (1982b). The effects of leader-member exchange and job design on productivity and job satisfaction: Testing a dual attachment model. *Organizational Behavior and Human Performance, 30*, 109-131.

Graen, G. B., & Uhl-Bien, M. (1995). Relationship-based approach to leadership: Development of leader-member exchange (LMX) theory of leadership over 25 years: Applying a multi-level multi-domain perspective. *Leadership Quarterly, 6*, 219-247.

Guerrero, G. (2001). *Measurement equivalence of English and Spanish versions of the Campbell interest and skill survey.* Unpublished doctoral dissertation, The University of Texas at El Paso.

Hambleton, R. K., & Patsula, L. (1999). *Increasing the validity of adapted tests: Myths to be avoided and guidelines for improving test adaptation practices*. Retrieved August 18, 2004, from http://www.testpublishers.org/journal01.htm

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of Item Response theory*. Newbury Park: Sage Publications.

Hambleton, R.K., & Kanjee, A. (1994). Increasing the validity of cross-cultural assessments: Use of improved methods for test adaptations. *European Journal of Psychological Assessments, 11*(3), 147-157.

Hogan, R., & Hogan, J. (2001). Assessing leadership: A view from the dark side. International. *Journal of Selection and Assessment, 9*, 40-21.

House, R. J. (1971). A path-goal theory of leadership effectiveness. *Administrative Science Quarterly, 16*, 321-332.

House, R. J. (1977). A 1976 theory of charismatic leadership. In J. G. Hunt & L. L. Larson (Eds.), *Leadership: The cutting edge* (pp. 189-207). Carbondale: Southern Illinois University Press.

House, R. J. (1996). Path-goal theory of leadership: Lessons, legacy, and a reformulated theory. *The Leadership Quarterly, 7*(323-352).

House, R. J., & Dessler, G. (1974). *The path goal theory of leadership: Some post hoc and a priori tests*. Carbondale: Southern Illinois University Press.

Howell, J. M., & Avolio, B. J. (1992). The ethics of charismatic leadership: Submission or liberation? *Academy of Management Executive, 6*(2), 43-54.

Howell, J. M., & Avolio, B. J. (1993). Transformational leadership, transactional leadership, locus of control, and support for innovation: Key predictors of consolidated-business-unit performance. *Journal of Applied Psychology, 78*(6), 891-902.

Hui, C. H., & Triandis, H. C. (1985). Measurement in cross-cultural psychology: A review and comparison of strategies. *Journal of Cross-Cultural Psychology, 16*(2), 131-152.

Hulin, C. L. (1987). A psychometric theory of evaluations of item and scale translations: Fidelity across languages. *Journal of Cross-Cultural Psychology, 18*(2), 115-142.

Hulin, C. L., Drasgow, F., & Komocar, J. (1982). Applications of item response theory to analysis of attitude scale translations. *Journal of Applied Psychology, 67*, 818-825.

Idaszak, J. R., Bottom, W. P., & Drasgow, F. (1988). A test of the measurement equivalence of the revised job diagnostic survey: Past problems and current solutions. *Journal of Applied Psychology, 73*(4), 647-656.

Joreskog, K. J. (1971). Simultaneous factor analysis in several populations. *Psychometrika, 36*, 409-426.

Jung, D. I., & Avolio, B. J. (2000). Opening the black box: An experimental investigation of the mediating effects of trust and value congruence on transformational and transactional leadership. *Journal of Organizational Behavior, 21*, 949-964.

Kanungo, R. N., & Mendonca, M. (1996). *Ethical dimensions of leadership*. Thousand Oaks: Sage Publications.

Kets de Vries, M. F. (1986). Personality, culture, and organization. *Academy of Management Review, 11*(2), 266-279.

Kirkpatrick, S. A., & Locke, E. A. (1996). Direct and indirect effects of three core charismatic leadership components on performance and attitudes. *Journal of Applied Psychology, 81*(1), 36-51.

Kouzes, J. M., & Posner, B. Z. (1993). *Credibility: How leaders gain and lose it and why people demand it*. San Francisco: Jossey-Bass.

Kouzes, J. M., & Posner, B. Z. (2002). *The leadership challenge* (3rd ed.). San Francisco: Jossey-Bass.

Kramer, T. R. (1996). *An empirical assessment of the behavioral dimensions of charismatic leadership.* Unpublished Dissertation, The Graduate School of the University of Alabama, Tuscaloosa.

Liden, R. C., Wayne, S. J., & Stilwell, D. (1993). A longitudinal study on the early development of leader-member exchanges. *Journal of Applied Psychology, 78*, 662-674.

Mackenzie, S. B., Podsakoff, P. M., & Rich, G. A. (2001). Transformational and transactional leadership and salesperson performance. *Journal of the Academy of Marketing Science, 29*(2), 115-134.

Magill, G., & Prybil, L. (2004). Stewardship and integrity in healthcare: A role for organizational ethics. *Journal of business ethics, 50*, 225-238.

Mann, R. D. (1959). A review of the relationships between personality and performance in small groups. *Psychological Bulletin, 56*, 241-270.

Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute, 22*, 719-748.

Maurer, T. J., Raju, N. S., & Collins, W. C. (1998). Peer and subordinate performance appraisal measurement equivalence. *Journal of Applied Psychology, 83*, 693-702.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review, 20*, 709-734.

Meade, A. W., & Lautenschlager, G. L. (2004). A comparison of item response theory and confirmatory factor analytic methodologies for establishing measurement equivalence/invariance. *Organizational Research Methods, 7*(4), 361-388.

Millsap, R. E., & Everson, H. T. (1993). Methodology review: Statistical approaches for assessing measurement bias. *Applied Psychological Measurement, 17*(4), 297-334.

Minkes, A. L., Small, M. W., & Chatterjee, S. R. (1999). Leadership and business ethics: Does it matter? Implications for management. *Journal of Business Ethics, 20*, 327-335.

Mitchell, T. R. (1993). Leadership, values, and accountability. In M. M. Chemers & R. Ayman (Eds.), *Leadership theory and research: Perspectives and directions* (pp. 109-136). San Diego: Academic Press.

Morgan, R. B. (1993). Self- and co-worker perceptions of ethics and their relationships to leadership and salary. *The Academy of Management Journal, 36*(1), 200-214.

Mortensen, R. A., Smith, J. E., & Cavanagh, G. F. (1989). The importance of ethics to job performance: An empirical investigation of manager's perceptions. *Journal of Business Ethics, 8*, 253-260.

Mount, M. K. (1984). Psychometric properties of subordinate ratings of managerial performance. *Personnel Psychology, 37*, 687-702.

Park, D., & Lautenschlager, G. L. (1990). Improving IRT item bias detection with iterative linking and ability scale purification. *Applied Psychological Measurement, 14*(2), 163-173.

Parry, K. W., & Proctor-Thomson, S. B. (2002a). *Subordinate Integrity Rating Scale (SIRS): An instrument for assessing leader ratings of subordinate integrity.* Unpublished manuscript.

Parry, K. W., & Proctor-Thomson, S. B. (2002b). Perceived integrity of transformational leaders in organizational settings. *Journal of Business Ethics, 35*, 75-96.

Posner, B. Z., & Schmidt, W. H. (1984). Values and the American manager: An update. *California Management Review, 26*(3), 202-216.

Potenza, M. T., Dorans, N. J., 19(1), & 23-37. (1995). DIF assessment for polytomously scored items: A framework for classification and evaluation. *Applied Psychological Measurement, 19*(1), 23-37.

Premeaux, S. R. (2004). The current link between management behavior and ethical philosophy. *Journal of Business Ethics, 51*, 269-278.

Raju, N.S., Laffitte, L. J., & Byrne, B. M. (2002). Measurement Equivalence: A comparison of methods based on confirmatory factor analysis and item response theory. *Journal of Applied Psychology*, *87*(3), 517-529.

Raju, N. S., van der Linden, W. J., & Fleer, P. F. (1995). IRT-based internal measures of differential functioning of items and tests. *Psychological Measurement, 19*(4), 353-368.

Reckase, M. D. (1979). Unifactor latent trait models applied to multi-factor tests: Results and implications. *Journal of Education Statistics, 4*, 207-230.

Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: Two approaches for exploring measurement equivalence. *Psychological Bulletin, 114*(3), 552-566.

Riggio, R. E. (2003). *Introduction to Industrial/Organizational psychology* (4th ed.). Upper Saddle River: Prentice Hall.

Riding, A. (1985). *Distant Neighbors: A portrait of the Mexicans.* New York: Alfred A. Knopf.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika, 39*, 111-121.

Scandura, T. A., & Graen, G. B. (1984). Moderating effects of initial leader-member exchange status on the effects of a leadership intervention. *Journal of Applied Psychology, 69*, 428-436.

Schriesheim, C. A., Castro, S. L., & Cogliser, C. C. (1999). Leader-member exchange (LMX) research: A comprehensive review of theory, measurement, and data-analytic practices. *The Leadership Quarterly, 10*(63-113).

Schriesheim, C. A., & Von Glinow, M. A. (1977). The path-goal theory of leadership: a theoretical and empirical analysis. *Academy of Management Journal*, *20,* 398-405.

Schuler, R. S., Jackson, S. E. Jackofsky, E., & Slocum, J. W., Jr. (1996). Managing human resources in Mexico: A cultural understanding. *Business Horizon,* 55-62.

Shamir, B., House, R. J., & Arthur, M. B. (1993). The motivational effects of charismatic leader: A self-concept based theory. *Organizational Science, 4*(4), 577-594.

Shea, C. M., & Howell, J. M. (1999). Charismatic leadership and task feedback: A laboratory study of their effects on self-efficacy and performance. *Leadership Quarterly, 10*(3), 375-396.

Simons, T. L. (1999). Behavioral integrity as a critical ingredient for transformational leadership. *Journal of Organizational Change Management, 12*(2), 89-104.

Sparks, J. R., & Schenk, J. A. (2001). Explaining the effects of transformational leadership: An investigation of the effects of higher-order motives in multilevel marketing organizations. *Journal of Organizational Behavior, 22*, 849-869.

Stephens, G. K., & Greer, C. R. (1995). Doing business in Mexico: Understanding cultural differences. *Organizational Dynamics, 24*(1), 39-55.

Stogdill, R. M. (1948). Personal factors associated with leadership: A survey of the literature. *Journal of Psychology, 25*, 35-71.

Stogdill, R. M. (1974). *Handbook of leadership: A survey of theory and research*. New York: Free Press.

Stogdill, R. M. (1981). Traits of leadership: A followup to 1970. In B. M. Bass (Rev. ed.), *Stogdill's handbook of leadership: A survey of theory and research.* New York: The Free Press.

Swaminathan, H., & Rogers, H. J. (1990a). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement, 27*, 361-370.

Swander, C. (1999). *Assessing the differential functioning of items and tests of a polytomous employee attitude survey.* Virginia Polytechnic Institute and State University, Blacksburg.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*(1), 4-70.

Veccchio, R. P. (1977). An empirical examination of the validity of Fiedler's model of leadership effectiveness. *Organizational Behavior and Human Performance, 19*, 180-206.

Veccchio, R. P. (1982). A further test of leadership effects due to between-group variation and within-group variation. *Journal of Applied Psychology, 67*, 200-208.

Vitell, S. J., & Davis, D. L. (1990). The relationship between ethics and job satisfaction: An empirical investigation. *Journal of Business Ethics, 9*, 489-494.

Waldman, D. A., Bass, B. M., & Yammarino, F. J. (1990). Adding to contingent-reward behavior: The augmenting effect of charismatic leadership. *Group and Organizational Studies, 15*, 381-394.

Weber, J. E. (1996). *Equivalence of leadership scales.* Unpublished Dissertation, New Mexico State University, Las Cruces.

Yammarino, F. J., & Bass, B. M. (1990). Long-term forecasting of transformational leadership and its effects among naval officers: Some preliminary findings. In K. E. Clark & M. B.

Clarks (Eds.), *Measures of Leadership* (pp. 151-171). West Orange: Leadership Library of America.

Yukl, G. (1981). *Leadership in organizations*. Englewood Cliffs: Prentice-Hall, Inc.

Yukl, G. (1994). *Leadership in organizations* (3rd ed.). Englewood Cliffs: Prentice Hall.

Yukl, G. A. (1971). Toward a behavioral theory of leadership. *Organizational Behavior and Human Performance, 6*, 414-440.

Zumbo, B. D. (2003). Does item-level DIF manifest itself in scale-level analyses? Implications for translating language tests. *Language Testing, 20*(2), 136-147.

Appendix A

List of all items used in the current study

Table A1

List of all items used in the current study from the U.S. Sample

Item	Item description
1	Would use my mistakes to attack me personally
*3	Would assign me tasks which (s)he knows I can't possibly complete in the time available
5	Always gets even
6	Gives special favors to certain "pet" employees, but not to me
8	Lies to me
10	Would risk me to protect himself/herself in work matters
*12	Would "pad" his/her expense account if given the opportunity
13	Deliberately fuels conflict among employees
14	Is evil
15	Would use my performance appraisal to criticize me as a person
*17	Would deliberately give me tasks without allowing me access to the resources necessary to complete them
19	Has it in for me
*21	Could be trusted with information that I want kept confidential
23	Would allow me to be blamed for his/her mistake
25	Would falsify records if it would help his/her work situation
*26	Helps smooth relations among employees
27	Lacks high morals
*28	Makes fun of my mistakes instead of coaching me as to how to do my job better
*30	Doesn't recommend me for pay raises, even when policy says (s)he should
*32	Discriminates against me because of my gender
34	Would deliberately exaggerate my mistakes to make me look bad when describing my performance to his/her superiors
36	Is vindictive

Item	Item description
*37	Would blame me for his/her own mistake
*39	Would make personal use of company property, even if it violated policy
42	Avoids coaching me because he/she wants me to fail
*44	Denies me time off without good reason
46	Would treat me better if I belonged to a different ethnic group
48	Would deliberately distort what I say
*50	Would embezzle money from the organization if the opportunity arose
*51	Deliberately makes employees angry at each other
52	Is a hypocrite
53	Would limit my training opportunities to prevent me from advancing
*55	Deliberately makes it difficult for me to schedule time off, even when I am due it
*57	Discriminates against me because of my age
59	Would blackmail an employee if (s)he thought (s)he could get away with it
61	Enjoys turning down my requests
63	Would make trouble for me if I got on his/her bad side
65	Would take credit for my ideas
67	Would steal from the organization
68	Would risk me to get back at someone else
72	Would engage in sabotage against the organization
75	Would fire people just because (s)he doesn't like them if (s)he could get away with it
77	Would do things which violate organizational policy and then expect his/her subordinates to cover for him/her

 Items with asterick marks were nonmatching items; therefore, not used in the analysis.

Table A2

List of all items used in the current study from the N.Z. Sample

Item	Item description
1(1)	Would use my mistakes to attack me personally
2(5)	Always tries to get even
3(8)	Would lie to me
4(10)	Would risk me to protect themselves in work matters
5(13)	Deliberately fuels conflict among employees
6(14)	Is evil
7(15)	Would use feedback to me to criticise me as a person
8(19)	Has it in for me
9(23)	Would allow me to be blamed for their mistake
10(25)	Would falsify records if it would help their work situation
11(27)	Lacks high morals
12(34)	Would deliberately exaggerate my mistakes to make me look bad when describing my performance to others
13(36)	Is vindictive
*14(37)	Would blame me for their own mistake
15(42)	Avoids giving me constructive feedback because they want me to fail
16(46)	Would respect me more if I belonged to a different ethnic group
17(48)	Would deliberately distort what I say
*18(51)	Deliberately makes employees angry at each other
19(52)	Is a hypocrite
20(59)	Would blackmail an employee if they thought they could get away with it
*21(78)	Deliberately puts off completing assignments in order to disrupt the flow of work
22(63)	Would make trouble for me if I got on their bad side
23(65)	Would try to take credit for my ideas
24(67)	Would steal from the organisation
25(68)	Would risk me to get back at someone else

Item	Item description
26(72)	Would engage in sabotage against the organisation
27(75)	Would try to get people fired just because they don't like them
28(77)	Would do things which violate organisational policy and then expect others to cover for them

Note. Items numbers in parentheses correspond to the item number sequence in the original version of the PLIS and were used to refer to items in this paper.

Table A3

List of the back-translated items used in the current study from the Mexico Sample

Item	Item description
1(1)	Laughs at others' mistakes
2(5)	Seeks revenge
3(6)	Shows favoritism
4(8)	Would lie to me
5(10)	Would put others at risk to protect his/her job
6(13)	Deliberately creates conflict among people
7(14)	Is a bad person
8(15)	Would use feedback as an excuse to criticize others
9(19)	Is against me
10(23)	Would allow others to be accused of her/his own mistakes
11(25)	Would change information to be seen in a more favorable light at the workplace
12(27)	Has high moral standards
13(34)	Would dare to exaggerate others' mistakes to make them look bad to others
14(36)	Is vengeful
15(42)	Would keep constructive information or feedback so others fail
16(46)	Would treat people depending on gender, ethnic group or social status
17(48)	Would dare to distort what I say
18(52)	Is hypocrite
19(53)	Would try to affect someone's professional growth negatively, for revenge
20(59)	Would blackmail an employee if believed would not be caught
21(61)	Likes to say no when asked for a favor
22(63)	Would cause problems to anyone who were against her/him
23(65)	Would try to take others' credit for her/himself
24(67)	Would steal to the organization
25(68)	Would put someone at risk to have his/her revenge against a third party

Item	Item description
26(72)	Would take active part against the company
27(75)	Would try to fire someone because he/she dislikes that person
28(77)	Would violate organization policies, sure that someone will cover up for her/him
*29(79)	Would put the company at risk to obtain personal benefits
*30(21)	Can be trusted with confidential information
*31(80)	Tells the truth
*32(81)	Would deliberately work slower to cause someone problems

Note. Items numbers in parentheses correspond to the item number sequence in the original version of the PLIS and were used to refer to items in this paper. Items with asterisk marks were non-matching item; therefore, not used in the analysis.

Appendix B

Exploratory Factor Analysis Results of the PLIS

Table B1a

United States sample using the English version of the Perceived Leader Integrity Scale

Item	Factor 1 Loading
52	.885
63	.839
23	.831
48	.821
68	.816
36	.804
34	.788
05	.787
65	.786
08	.780
10	.774
01	.771
27	.766
75	.757
14	.737
25	.735
61	.724
15	.710
13	.708
19	.707
59	.651
77	.635
53	.631
06	.622
42	.603
46	.596
67	.593
72	.519

Note. A one factor extraction. Factor analysis of the U.S. sample using items that match the Spanish speaking sample.

Table B1b

United States sample using the English version of the Perceived Leader Integrity Scale

Item	Factor 1 Loadings	Factor 2 Loadings
19	1.005	-.301
01	.934	-.152
34	.858	-.050
15	.812	-.088
63	.806	.064
61	.767	-.022
05	.730	.089
06	.724	-.091
42	.669	-053
08	.664	.154
36	.641	.206
10	.625	.190
13	.613	.128
48	.598	.273
23	.578	.309
53	.573	.085
46	.547	.074
65	.522	.318
52	.502	.454
68	.481	.398
75	.434	.384
67	-.285	1.013
77	-.182	.943
25	.000	.852
59	.012	.739
72	.009	.588
27	.326	.516
14	.378	.423

Note. Two factors extracted. Factor analysis of the U.S. sample using items that match the Spanish speaking sample. Factor correlation = .73.

Table B1c

United States sample using the English version of the Perceived Leader Integrity Scale

Factor	Eigenvalues	% of Variance	Cumulative %
1	15.479	55.283	55.283
2	1.958	6.994	62.276
3	1.306	4.665	66.942
4	1.019	3.640	70.581
5	0.887	3.170	73.751
6	.730	2.607	76.358
7	.583	2.083	78.441
8	.569	2.031	80.472
9	.502	1.795	82.267
10	.493	1.760	84.027
11	.462	1.651	85.678
12	.390	1.394	87.072
13	.383	1.367	88.439
14	.361	1.290	89.729
15	.323	1.155	90.884
16	.303	1.083	91.967
17	.294	1.051	93.017
18	.255	.912	93.929
19	.232	.828	94.757
20	.221	.791	95.548
21	.210	.749	96.297
22	.188	.672	96.969
23	.168	.601	97.570
24	.154	.550	98.121
25	.147	.524	98.644
26	.136	.484	99.128
27	.132	.471	99.599
28	.112	.401	100.000

Note. Factor analysis of the U.S. sample using items that matched the Spanish speaking sample.

Table B2a

New Zealand sample using the English version of the Perceived Leader Integrity Scale

Item	Factor 1 Loading
34	.703
68	.696
52	.664
48	.660
23	.636
77	.613
36	.610
01	.608
19	.606
63	.593
05	.578
08	.576
27	.567
42	.561
14	.549
65	.528
15	.524
75	.519
25	.504
59	.500
10	.433
67	.419
72	.347
46	.178

Note. A one factor extraction. Factor analysis using items that matched the Spanish speaking sample and prior to conducting a DFIT analysis.

Table B2b

New Zealand sample using the English version of the Perceived Leader Integrity Scale

Item	Factor 1 Loadings	Factor 2 Loadings
01	.812	-.191
34	.717	.017
23	.706	-.047
15	.680	-.145
05	.627	-.028
63	.627	-.010
10	.619	-.183
13	.571	.018
65	.526	.025
48	.495	.209
52	.487	.222
08	.445	.168
36	.437	.216
75	.380	.175
42	.346	.228
14	-.069	.715
59	-.113	.707
67	-.157	.661
72	-.146	.564
25	.026	.554
77	.231	.448
68	.320	.445
19	.236	.436
27	.249	.374
46	-.019	.225

Note. Two factors extracted. Factor analysis using items that matched the Spanish speaking sample and prior to conducting a DFIT analysis. Factor correlation = .71.

Table B2c

New Zealand sample using the English version of the Perceived Leader Integrity Scale

Factors	Eigenvalues	% of Variance	Cumulative %
1	8.512	34.047	34.047
2	1.735	6.940	40.986
3	1.208	4.832	45.818
4	1.086	4.345	50.163
5	.974	3.897	54.061
6	.876	3.504	57.565
7	.829	3.316	60.880
8	.789	3.157	64.038
9	.764	3.057	67.095
10	.699	2.795	69.889
11	.658	2.631	72.520
12	.643	2.572	75.092
13	.630	2.520	77.612
14	.599	2.397	80.009
15	.572	2.288	82.297
16	.539	2.155	84.452
17	.526	2.103	86.555
18	.478	1.913	88.468
19	.471	1.884	90.352
20	.458	1.832	92.183
21	.455	1.821	94.004
22	.415	1.658	95.662
23	.400	1.600	97.262
24	.362	1.449	98.711
25	.322	1.289	100.000

Note. Factor analysis using items that matched the Spanish speaking sample and prior to conducting a DFIT analysis.

Table B2d

New Zealand sample using the English version of the Perceived Leader Integrity Scale after removing 6 items with DIF

Item	Factor 1 Loading
34	.706
68	.681
48	.670
23	.640
77	.625
01	.594
63	.586
42	.585
27	.581
08	.578
13	.560
05	.552
14	.548
25	.545
65	.526
59	.503
67	.438
72	.363
46	.176

Note. A one factor extraction after excluding items 10, 15, 19, 36, 52, and 75 found to have DIF.

Table B2e

New Zealand sample using the English version of the Perceived Leader Integrity Scale after removing 6 items with DIF

Item	Factor 1 Loadings	Factor 2 Loadings
01	.785	-.176
34	.753	-.016
23	.709	-.044
05	.657	-.088
63	.634	-.024
13	.601	-.018
48	.565	.145
65	.520	.029
42	.459	.164
08	.430	.188
68	.407	.333
67	-.149	.687
59	-.044	.643
25	.028	.609
14	.074	.559
72	-.105	.544
77	.257	.439
27	.276	.365
46	.010	.193

Note. Two factors extracted after excluding items 10, 15, 19, 36, 52, and 75 found to have DIF.

Factor correlation = .69.

Table B2f

New Zealand sample using the English version of the Perceived Leader Integrity Scale after removing 6 items with DIF

Factors	Eigenvalues	% of Variance	Cumulative %
1	6.668	35.094	35.094
2	1.533	8.067	43.161
3	1.065	5.606	48.766
4	.978	5.147	53.913
5	.914	4.813	58.726
6	.798	4.198	62.923
7	.728	3.831	66.755
8	.699	3.681	70.436
9	.659	3.469	73.905
10	.601	3.161	77.066
11	.588	3.092	80.159
12	.560	2.948	83.106
13	.549	2.887	85.994
14	.508	2.676	88.669
15	.506	2.661	91.330
16	.474	2.494	93.824
17	.421	2.217	96.041
18	.395	2.078	98.119
19	.357	1.881	100.000

Note. Factor analysis after excluding items 10, 15, 19, 36, 52, and 75 found to have DIF.

Table B3a

Mexico sample using the Spanish version of the PLIS

Item	Factor 1 Loading
52	.859
53	.857
36	.845
14	.836
48	.832
63	.827
59	.824
68	.821
34	.820
42	.812
65	.809
10	.799
75	.798
23	.790
05	.773
61	.768
25	.762
08	.753
13	.743
77	.743
19	.717
01	.680
06	.658
46	.655
72	.627
67	.615
15	.611
27	.266

Note. A one factor extraction. Factor analysis does not include items 79, 21, 80, and 81 because no data have been collected for these items using the English speaking samples.

Table B3b

Mexico sample using the Spanish version of the PLIS

Item	Factor 1 Loadings	Factor 2 Loadings
34	.901	-.070
01	.898	-.223
36	.887	-.025
10	.829	-.014
05	.814	-.027
06	.788	-.129
52	.755	.135
63	.713	.145
23	.700	.118
15	.699	-.084
08	.695	.071
13	.695	.071
48	.683	.185
75	.651	.182
61	.623	.178
19	.604	.142
65	.594	.258
14	.562	.324
46	.520	.165
53	.506	.411
27	.271	.001
67	-.331	1.083
72	-.187	.931
77	.121	.715
68	.258	.649
25	.307	.526
59	.377	.519
42	.425	.450

Note. Two factors extracted. Factor analysis does not include items 79, 21, 80, and 81 because no data have been collected for these items using the English speaking samples.

Table B3c

Mexico sample using the Spanish version of the PLIS

Factors	Eigenvalues	% of Variance	Cumulative %
1	16.385	58.518	58.518
2	1.542	5.506	64.025
3	1.015	3.627	67.651
4	.826	2.948	70.599
5	.733	2.617	73.217
6	.705	2.518	75.735
7	.594	2.120	77.855
8	.578	2.066	79.921
9	.531	1.895	81.816
10	.486	1.734	83.550
11	.471	1.684	85.234
12	.403	1.438	86.672
13	.391	1.395	88.067
14	.341	1.218	89.284
15	.316	1.128	90.412
16	.312	1.115	91.527
17	.285	1.018	92.545
18	.259	.925	93.470
19	.246	.878	94.349
20	.231	.827	95.979
21	.225	.804	95.979
22	.217	.775	96.754
23	.189	.677	97.431
24	.178	.637	98.067
25	.161	.576	98.643
26	.150	.535	99.178
27	.131	.470	99.648
28	.099	.352	100.000

Note. Factor analysis does not include items 79, 21, 80, and 81 because no data have been collected for these items using the English speaking samples.

Table B4a

English speaking (US-NZ) samples using English versions of the PLIS

Item	Factor 1 Loading
52	.720
68	.712
34	.701
48	.686
23	.666
36	.656
63	.635
01	.627
19	.620
05	.620
77	.618
27	.613
08	.604
14	.601
75	.579
13	.576
25	.562
65	.550
59	.548
42	.538
15	.530
10	.486
67	.470
72	.379
46	.252

Note. A one factor extraction. Factor analysis of the US-NZ sample without items 6, 53, and 61 because no data were collected on these items using the N.Z. sample. Analysis was conducted prior to a DFIT analysis.

Table B4b

English speaking (US-NZ) samples using English versions of the PLIS

Item	Factor 1 Loadings	Factor 2 Loadings
1	.851	-.208
34	.742	-.011
15	.703	-.161
23	.680	.015
63	.658	.005
05	.655	-.009
10	.622	-.125
13	.600	.000
48	.524	.206
65	.520	.057
36	.478	.222
08	.470	.172
52	.448	.327
42	.431	.139
68	.390	.382
19	.381	.288
75	.379	.243
67	-.207	.772
59	-.101	.744
14	.019	.670
25	-.015	.663
72	-.105	.549
77	.175	.514
27	.234	.441
46	.052	.230

Note. Two factors extracted. Factor analysis of the US-NZ sample without items 6, 53, and 61 because no data were collected on these items using the N.Z. sample. Analysis was conducted prior to a DFIT analysis. Factor correlation = .712.

Table B4c

English speaking (US-NZ) samples using English versions of the PLIS without the three items not used by the N.Z. sample

Factor	Eigenvalues	% of Variance	Cumulative %
1	9.350	37.398	37.398
2	1.718	6.872	44.271
3	1.140	4.449	48.830
4	1.052	4.206	53.036
5	.922	3.688	56.724
6	.867	3.470	60.193
7	.800	3.201	63.395
8	.730	2.919	66.313
9	.702	2.806	69.120
10	.681	2.722	71.842
11	.606	2.426	74.268
12	.603	2.414	76.682
13	.579	2.318	78.999
14	.556	2.224	81.224
15	.540	2.164	83.386
16	.518	2.073	85.458
17	.497	1.987	87.445
18	.454	1.817	89.262
19	.437	1.748	91.010
20	.425	1.700	92.710
21	.406	1.622	94.332
22	.391	1.563	95.895
23	.365	1.460	97.355
24	.347	1.387	98.742
25	.314	1.258	100.000

Note. Factor analysis of the US-NZ sample without items 6, 53, and 61 because no data were collected on these items using the N.Z. sample. Analysis was conducted prior to a DFIT analysis.

Table B4d

English speaking (US-NZ) samples using English versions of the PLIS with the three items not used by the N.Z. Sample

Item	Factor 1 Loading
52	.738
34	.724
48	.698
36	.689
68	.685
23	.675
63	.667
05	.642
01	.638
19	.631
27	.629
08	.623
14	.607
77	.606
13	.597
75	.589
25	.579
42	.562
65	.561
15	.557
10	.518
59	.496
67	.453
61	.432
53	.396
06	.388
72	.371
46	.268

Note. A one factor extraction. Factor analysis of the US-NZ sample with items 6, 53, and 61 that were not used in the N.Z. sample. Analysis was conducted prior to a DFIT analysis.

Table B4e

English speaking (US-NZ) samples using English versions of the PLIS with the three items not used by the N.Z. Sample

Item	Factor 1 Loadings	Factor 2 Loadings
34	.753	-.039
48	.730	-.045
68	.725	-.059
23	.696	-.028
77	.673	-.103
01	.657	-.026
52	.655	.144
63	.637	.056
36	.636	.093
05	.625	.033
13	.616	-.026
08	.612	.023
27	.610	.036
42	.600	-.056
65	.583	-.032
15	.582	-.035
25	.536	.075
75	.534	.097
19	.529	.176
59	.518	-.032
14	.510	.166
10	.489	.053
67	.467	-.020
72	.421	-.079
46	.181	.146
61	-.086	.904
53	-.032	.737
6	-.001	.667

Note. Two factors extracted. Factor analysis of the US-NZ sample with items 6, 53, and 61 that were not used in the N.Z. sample. Analysis was conducted prior to a DFIT analysis. Factor correlation = .51.

Table B4f

English speaking (US-NZ) samples using English versions of the PLIS with the three items not used by the N.Z. Sample

Factor	Eigenvalues	% of Variance	Cumulative %
1	9.350	37.398	37.398
2	1.718	6.872	44.271
3	1.140	4.449	48.830
4	1.052	4.206	53.036
5	.922	3.688	56.724
6	.867	3.470	60.193
7	.800	3.201	63.395
8	.730	2.919	66.313
9	.702	2.806	69.120
10	.681	2.722	71.842
11	.606	2.426	74.268
12	.603	2.414	76.682
13	.579	2.318	78.999
14	.556	2.224	81.224
15	.540	2.164	83.386
16	.518	2.073	85.458
17	.497	1.987	87.445
18	.454	1.817	89.262
19	.437	1.748	91.010
20	.425	1.700	92.710
21	.406	1.622	94.332
22	.391	1.563	95.895
23	.365	1.460	97.355
24	.347	1.387	98.742
25	.314	1.258	100.000

Note. Factor analysis of the US-NZ sample with items 6, 53, and 61 that were not used in the N.Z. sample. Analysis was conducted prior to a DFIT analysis.

Table B4g

English speaking (US-NZ) samples using English versions of the PLIS without DIF items found between United States and New Zealand.

Item	Factor 1 Loading
68	.705
34	.704
48	.694
23	.669
77	.634
63	.625
27	.617
01	.613
08	.603
25	.597
14	.597
05	.594
13	.575
42	.563
65	.554
59	.552
67	.485
72	.395
46	.246

Note. A one factor extraction.

Table B4h

English speaking (US-NZ) samples using English versions of the PLIS without DIF items found between United States and New Zealand.

Item	Factor 1 Loadings	Factor 2 Loadings
01	.810	-.177
34	.766	-.030
23	.692	.010
05	.666	-.047
63	.664	-.010
13	.628	-.028
48	.589	.148
65	.524	.059
42	.490	.106
08	.466	.178
68	.456	.305
67	-.177	.768
59	-.033	.681
25	.025	.669
14	.129	.549
72	-.063	.528
77	.210	.500
27	.273	.408
46	.072	.203

Note. Two factors extracted. Factor correlation = .69.

Table B4i

English speaking (US-NZ) samples using English versions of the PLIS without DIF items found between United States and New Zealand.

Factor	Eigenvalues	% of Variance	Cumulative %
1	7.231	38.060	38.060
2	1.538	8.096	46.156
3	.980	5.160	51.316
4	.925	4.871	56.187
5	.874	4.598	60.784
6	.764	4.018	64.803
7	.712	3.747	68.550
8	.662	3.483	72.033
9	.628	3.307	75.340
10	.572	3.009	78.348
11	.557	2.931	81.279
12	.525	2.761	84.040
13	.518	2.726	86.767
14	.481	2.534	89.300
15	.475	2.499	91.800
16	.435	2.289	94.089
17	.394	2.075	96.163
18	.381	2.004	98.167
19	.348	1.833	100.000

Appendix C

Correlation Matrix of all items analyzed including items that were not analyzed

Table C1a

Correlation Matrix of all items analyzed for the United States sample

	plis01	plis05	plis06	plis08	plis10	plis13	plis14	plis15	plis19	plis23	plis27	plis34	plis36	plis42	plis46	plis48	plis52	plis53	plis59	plis61	plis63	plis65	plis67	plis68	plis72	plis75	plis77
plis01	1.000	.690	.568	.614	.691	.578	.511	.686	.642	.654	.552	.667	.603	.462	.432	.660	.662	.467	.407	.601	.697	.595	.265	.617	.306	.547	.371
plis05	.690	1.000	.521	.578	.660	.649	.550	.556	.574	.639	.554	.610	.802	.427	.409	.623	.695	.400	.504	.600	.737	.552	.388	.667	.321	.661	.420
plis06	.568	.521	1.000	.529	.587	.428	.355	.441	.473	.560	.441	.509	.469	.369	.420	.506	.556	.396	.242	.533	.592	.532	.247	.425	.201	.562	.349
plis08	.614	.578	.529	1.000	.593	.623	.537	.552	.572	.562	.611	.684	.640	.478	.476	.705	.678	.534	.441	.578	.639	.589	.376	.639	.463	.548	.482
plis10	.691	.660	.587	.593	1.000	.563	.469	.537	.476	.761	.595	.623	.615	.383	.387	.578	.691	.451	.439	.489	.694	.658	.402	.651	.342	.642	.492
plis13	.578	.649	.428	.623	.563	1.000	.483	.495	.556	.578	.533	.559	.616	.389	.430	.617	.593	.341	.455	.480	.621	.490	.316	.644	.509	.497	.383
plis14	.511	.550	.355	.537	.469	.483	1.000	.592	.547	.572	.643	.518	.611	.581	.443	.555	.698	.513	.695	.502	.587	.512	.543	.541	.379	.558	.495
plis15	.686	.556	.441	.552	.537	.495	.592	1.000	.630	.545	.490	.725	.574	.482	.438	.547	.583	.509	.466	.585	.578	.540	.334	.548	.215	.486	.340
plis19	.642	.574	.473	.572	.476	.556	.547	.630	1.000	.544	.501	.649	.575	.685	.561	.597	.563	.549	.340	.619	.620	.514	.243	.471	.297	.499	.242
plis23	.654	.639	.560	.562	.761	.578	.572	.545	.544	1.000	.607	.634	.642	.487	.540	.668	.751	.504	.519	.570	.721	.759	.510	.695	.452	.623	.548
plis27	.552	.554	.441	.611	.595	.533	.643	.490	.501	.607	1.000	.561	.613	.423	.399	.663	.745	.455	.551	.521	.575	.616	.581	.610	.395	.565	.576
plis34	.667	.610	.509	.684	.623	.559	.518	.725	.649	.634	.561	1.000	.676	.520	.514	.669	.638	.595	.401	.613	.608	.631	.374	.628	.326	.541	.391
plis36	.603	.802	.469	.640	.615	.616	.611	.574	.575	.642	.613	.676	1.000	.412	.384	.624	.704	.449	.535	.612	.727	.576	.438	.687	.324	.657	.482
plis42	.462	.427	.369	.478	.383	.389	.581	.482	.685	.487	.423	.520	.412	1.000	.615	.479	.492	.553	.293	.449	.492	.441	.370	.397	.286	.406	.273
plis46	.432	.409	.420	.476	.387	.430	.443	.438	.561	.540	.399	.514	.384	.615	1.000	.511	.495	.432	.285	.407	.435	.491	.360	.442	.475	.408	.343
plis48	.660	.623	.506	.705	.578	.617	.555	.547	.597	.668	.663	.669	.624	.479	.511	1.000	.740	.480	.539	.589	.690	.656	.502	.694	.419	.603	.495
plis52	.662	.695	.556	.678	.691	.593	.698	.583	.563	.751	.745	.638	.704	.492	.495	.740	1.000	.521	.598	.596	.766	.675	.587	.728	.454	.722	.608
plis53	.467	.400	.396	.534	.451	.341	.513	.509	.549	.504	.455	.595	.449	.553	.432	.480	.521	1.000	.435	.561	.486	.542	.337	.414	.246	.378	.361
plis59	.407	.504	.242	.441	.439	.455	.695	.466	.340	.519	.551	.401	.535	.293	.285	.539	.598	.435	1.000	.426	.491	.461	.595	.577	.442	.530	.567
plis61	.601	.600	.533	.578	.489	.480	.502	.585	.619	.570	.521	.613	.612	.449	.407	.589	.596	.561	.426	1.000	.678	.564	.328	.614	.346	.500	.361
plis63	.697	.737	.592	.639	.694	.621	.587	.578	.620	.721	.575	.608	.727	.492	.435	.690	.766	.486	.491	.678	1.000	.691	.358	.741	.374	.700	.463
plis65	.595	.552	.532	.589	.658	.490	.512	.540	.514	.759	.616	.631	.576	.441	.491	.656	.675	.542	.461	.564	.691	1.000	.513	.680	.457	.520	.536
plis67	.265	.388	.247	.376	.402	.316	.543	.334	.243	.510	.581	.374	.438	.370	.360	.502	.587	.337	.595	.328	.358	.513	1.000	.501	.482	.486	.668
plis68	.617	.667	.425	.639	.651	.644	.541	.548	.471	.695	.610	.628	.687	.397	.442	.694	.728	.414	.577	.614	.741	.680	.501	1.000	.544	.630	.528
plis72	.306	.321	.201	.463	.342	.509	.379	.215	.297	.452	.395	.326	.324	.286	.475	.419	.454	.246	.442	.346	.374	.457	.482	.544	1.000	.326	.549
plis75	.547	.661	.562	.548	.642	.497	.558	.486	.499	.623	.565	.541	.657	.406	.408	.603	.722	.378	.530	.500	.700	.520	.486	.630	.326	1.000	.593
plis77	.371	.420	.349	.482	.492	.383	.495	.340	.242	.548	.576	.391	.482	.273	.343	.495	.608	.361	.567	.361	.463	.536	.668	.528	.549	.593	1.000

Table C1b

Correlation Matrix of all items analyzed including those not analyzed for the United States sample

	01	03	05	06	08	10	12	13	14	15	17	19	21	23	25	26	27	28	30	32	34	36	37	39	42	44	46	48	50	51	52	53	55	57	59	61	63	65	67	68	72	75	77
plis01	1.00	.458	.678	.554	.591	.677	.375	.557	.478	.675	.415	.639	.333	.634	.411	.398	.526	.478	.410	.375	.649	.589	.637	.351	.464	.323	.437	.651	.270	.544	.645	.424	.388	.296	.358	.584	.689	.578	.255	.601	.293	.547	.334
plis03	.458	1.00	.342	.371	.290	.365	.276	.229	.305	.337	.564	.390	.124	.445	.276	.208	.275	.377	.323	.220	.357	.296	.394	.280	.415	.435	.338	.348	.207	.273	.371	.451	.389	.211	.130	.450	.462	.437	.237	.326	.063	.357	.168
plis05	.678	.342	1.00	.493	.559	.656	.448	.635	.526	.534	.368	.566	.311	.624	.486	.394	.528	.477	.405	.300	.588	.794	.650	.463	.426	.293	.411	.611	.394	.620	.684	.360	.401	.303	.472	.575	.726	.531	.384	.648	.312	.652	.389
plis06	.554	.371	.493	1.00	.498	.574	.293	.384	.291	.402	.347	.463	.349	.528	.313	.373	.402	.367	.374	.370	.477	.420	.494	.310	.348	.270	.403	.485	.239	.337	.535	.343	.281	.325	.166	.493	.566	.501	.226	.373	.156	.543	.302
plis08	.591	.290	.559	.498	1.00	.570	.390	.587	.480	.512	.308	.567	.358	.521	.518	.402	.577	.335	.417	.343	.660	.619	.599	.376	.460	.251	.454	.693	.317	.581	.660	.481	.375	.303	.374	.544	.630	.562	.361	.612	.440	.553	.435
plis10	.677	.365	.656	.574	.570	1.00	.459	.545	.435	.517	.386	.470	.308	.746	.521	.382	.584	.386	.482	.375	.612	.604	.713	.470	.369	.295	.372	.570	.365	.551	.683	.405	.348	.311	.402	.462	.690	.645	.396	.641	.322	.652	.468
plis12	.375	.276	.448	.293	.390	.459	1.00	.287	.485	.360	.293	.254	.192	.482	.686	.247	.545	.321	.324	.363	.368	.398	.555	.733	.253	.339	.285	.442	.575	.281	.551	.332	.402	.307	.504	.394	.377	.440	.678	.438	.377	.522	.575
plis13	.557	.229	.635	.384	.587	.545	.287	1.00	.408	.455	.226	.550	.303	.543	.388	.399	.482	.343	.279	.270	.520	.585	.554	.331	.339	.135	.373	.607	.284	.793	.559	.239	.314	.235	.388	.426	.607	.455	.283	.610	.468	.502	.314
plis14	.478	.305	.526	.291	.480	.435	.485	.408	1.00	.543	.347	.548	.265	.528	.501	.303	.598	.493	.350	.408	.447	.574	.528	.479	.568	.466	.390	.542	.553	.411	.678	.398	.518	.325	.632	.443	.576	.466	.547	.487	.318	.566	.422
plis15	.675	.337	.534	.402	.512	.517	.360	.455	.543	1.00	.385	.628	.206	.509	.383	.325	.445	.676	.399	.455	.698	.548	.528	.358	.480	.385	.431	.521	.341	.446	.557	.448	.511	.396	.395	.559	.568	.505	.325	.517	.177	.476	.281
plis17	.415	.564	.368	.347	.308	.386	.293	.226	.347	.385	1.00	.398	.154	.459	.300	.264	.305	.397	.300	.247	.398	.342	.454	.300	.488	.312	.463	.409	.265	.307	.448	.497	.381	.345	.189	.405	.413	.395	.300	.370	.103	.385	.241
plis19	.639	.390	.566	.463	.567	.470	.254	.550	.548	.628	.398	1.00	.221	.539	.356	.307	.490	.569	.358	.417	.647	.572	.569	.227	.699	.367	.579	.595	.217	.570	.554	.558	.477	.412	.318	.620	.624	.512	.230	.462	.286	.499	.212
plis21	.333	.124	.311	.349	.358	.308	.192	.303	.265	.206	.154	.221	1.00	.241	.247	.426	.317	.123	.139	.140	.241	.314	.271	.218	.194	.102	.147	.329	.174	.303	.363	.169	.149	.163	.214	.240	.321	.278	.189	.242	.191	.283	.224
plis23	.634	.445	.624	.528	.521	.746	.482	.543	.528	.509	.459	.539	.241	1.00	.546	.397	.579	.435	.482	.447	.608	.615	.869	.504	.473	.308	.533	.656	.510	.601	.739	.444	.379	.355	.471	.533	.709	.743	.508	.671	.431	.623	.514
plis25	.411	.276	.486	.313	.518	.521	.686	.388	.501	.383	.300	.356	.247	.546	1.00	.295	.615	.368	.375	.337	.500	.548	.629	.708	.327	.334	.336	.593	.694	.389	.646	.468	.410	.345	.607	.433	.479	.551	.748	.558	.427	.632	.681
plis26	.398	.208	.394	.373	.402	.382	.247	.399	.303	.325	.264	.307	.426	.397	.295	1.00	.403	.256	.260	.217	.361	.425	.375	.242	.224	.167	.242	.398	.230	.375	.452	.247	.245	.220	.239	.288	.447	.384	.233	.350	.222	.407	.316
plis27	.526	.275	.528	.402	.577	.584	.545	.482	.598	.445	.305	.490	.317	.579	.615	.403	1.00	.389	.339	.373	.517	.584	.641	.560	.395	.287	.361	.653	.508	.501	.723	.380	.437	.343	.495	.476	.557	.595	.581	.576	.359	.570	.535
plis28	.478	.377	.477	.367	.335	.386	.321	.343	.493	.676	.397	.569	.123	.435	.368	.256	.389	1.00	.336	.311	.531	.453	.414	.279	.499	.466	.311	.401	.261	.296	.460	.437	.495	.418	.295	.547	.514	.384	.276	.371	.152	.451	.213
plis30	.410	.323	.405	.374	.417	.482	.324	.279	.350	.399	.300	.358	.139	.482	.375	.260	.339	.336	1.00	.364	.459	.454	.489	.280	.338	.349	.260	.365	.227	.326	.435	.349	.411	.243	.302	.473	.514	.484	.297	.451	.268	.385	.318
plis32	.375	.220	.300	.370	.343	.375	.363	.270	.408	.455	.247	.417	.140	.447	.337	.217	.373	.311	.364	1.00	.336	.298	.436	.376	.439	.387	.483	.333	.298	.244	.410	.297	.400	.404	.244	.253	.399	.440	.394	.238	.257	.411	.386
plis34	.649	.357	.588	.477	.660	.612	.368	.520	.447	.698	.398	.647	.241	.608	.500	.361	.517	.531	.459	.336	1.00	.655	.659	.378	.520	.288	.514	.660	.382	.521	.612	.540	.473	.406	.309	.584	.595	.608	.366	.600	.298	.532	.330
plis36	.589	.296	.794	.420	.619	.604	.398	.585	.574	.548	.342	.572	.314	.615	.548	.425	.584	.453	.454	.298	.655	1.00	.688	.460	.390	.275	.353	.617	.406	.576	.693	.388	.463	.328	.490	.570	.709	.546	.430	.653	.286	.643	.440
plis37	.637	.394	.650	.494	.599	.713	.555	.554	.528	.528	.454	.569	.271	.869	.629	.375	.641	.414	.489	.436	.659	.688	1.00	.562	.522	.282	.534	.693	.549	.651	.765	.461	.371	.393	.525	.604	.737	.721	.563	.715	.473	.653	.543
plis39	.351	.280	.463	.310	.376	.470	.733	.331	.479	.358	.300	.227	.218	.504	.708	.242	.560	.279	.280	.376	.378	.460	.562	1.00	.245	.285	.292	.482	.676	.320	.585	.373	.353	.273	.533	.386	.388	.477	.750	.491	.377	.478	.615
plis42	.464	.415	.426	.348	.460	.369	.253	.339	.568	.480	.488	.699	.194	.473	.327	.224	.395	.499	.338	.439	.520	.390	.522	.245	1.00	.407	.557	.494	.323	.423	.478	.540	.414	.282	.258	.430	.504	.443	.326	.371	.186	.431	.209
plis44	.323	.435	.293	.270	.251	.295	.339	.135	.466	.385	.312	.367	.102	.308	.334	.167	.287	.466	.349	.387	.288	.275	.282	.285	.407	1.00	.268	.333	.318	.153	.402	.527	.597	.317	.330	.438	.424	.307	.303	.255	.052	.386	.233
plis46	.437	.338	.411	.403	.454	.372	.285	.373	.390	.431	.463	.579	.147	.533	.336	.242	.361	.311	.260	.483	.514	.353	.534	.292	.557	.268	1.00	.542	.328	.438	.482	.377	.310	.526	.235	.378	.448	.503	.307	.418	.379	.446	.273
plis48	.651	.348	.611	.485	.693	.570	.442	.607	.542	.521	.409	.595	.329	.656	.593	.398	.653	.401	.365	.333	.660	.617	.693	.482	.494	.333	.542	1.00	.563	.649	.738	.462	.389	.403	.518	.574	.686	.642	.513	.688	.433	.609	.480
plis50	.270	.207	.394	.239	.317	.365	.575	.284	.553	.341	.265	.217	.174	.510	.694	.230	.508	.261	.227	.298	.382	.406	.549	.676	.323	.318	.328	.563	1.00	.349	.581	.333	.335	.162	.698	.317	.366	.502	.891	.520	.416	.492	.571
plis51	.544	.273	.620	.337	.581	.551	.281	.793	.411	.446	.307	.570	.303	.601	.389	.375	.501	.296	.326	.244	.521	.576	.651	.320	.423	.153	.438	.649	.349	1.00	.596	.339	.285	.218	.422	.447	.634	.537	.325	.702	.520	.521	.331
plis52	.645	.371	.684	.535	.660	.683	.551	.559	.678	.557	.448	.554	.363	.739	.646	.452	.723	.460	.435	.410	.612	.693	.765	.585	.478	.402	.482	.738	.581	.596	1.00	.470	.456	.426	.565	.571	.764	.661	.586	.716	.435	.739	.580
plis53	.424	.451	.360	.343	.481	.405	.332	.239	.398	.448	.497	.558	.169	.444	.468	.247	.380	.437	.349	.297	.540	.388	.461	.373	.540	.527	.377	.462	.333	.339	.470	1.00	.508	.363	.312	.516	.457	.496	.316	.338	.157	.365	.265
plis55	.388	.389	.401	.281	.375	.348	.402	.314	.518	.511	.381	.477	.149	.379	.410	.245	.437	.495	.411	.400	.473	.463	.371	.353	.414	.597	.310	.389	.335	.285	.456	.508	1.00	.245	.363	.502	.461	.362	.395	.324	.120	.389	.334
plis57	.296	.211	.303	.325	.303	.311	.307	.235	.325	.396	.345	.412	.163	.355	.345	.220	.343	.418	.243	.404	.406	.328	.393	.273	.282	.317	.526	.403	.162	.218	.426	.363	.245	1.00	.225	.298	.359	.308	.175	.268	.218	.436	.209
plis59	.358	.130	.472	.166	.374	.402	.504	.388	.632	.395	.189	.318	.214	.471	.607	.239	.495	.295	.302	.244	.309	.490	.525	.533	.258	.330	.235	.518	.698	.422	.565	.312	.363	.225	1.00	.359	.462	.404	.620	.534	.425	.525	.518
plis61	.584	.450	.575	.493	.544	.462	.394	.426	.443	.559	.405	.620	.240	.533	.433	.288	.476	.547	.473	.253	.584	.570	.604	.386	.430	.438	.378	.574	.317	.447	.571	.516	.502	.298	.359	1.00	.653	.530	.312	.568	.309	.481	.304
plis63	.689	.462	.726	.566	.630	.690	.377	.607	.576	.568	.413	.624	.321	.709	.479	.447	.557	.514	.514	.399	.595	.709	.737	.388	.504	.424	.448	.686	.366	.634	.764	.457	.461	.359	.462	.653	1.00	.675	.359	.723	.376	.694	.443
plis65	.578	.437	.531	.501	.562	.645	.440	.455	.466	.505	.395	.512	.278	.743	.551	.384	.595	.384	.484	.440	.608	.546	.721	.477	.443	.307	.503	.642	.502	.537	.661	.496	.362	.308	.404	.530	.675	1.00	.525	.657	.460	.508	.512
plis67	.255	.237	.384	.226	.361	.396	.678	.283	.547	.325	.300	.230	.189	.508	.748	.233	.581	.276	.297	.394	.366	.430	.563	.750	.326	.303	.307	.513	.891	.325	.586	.316	.395	.175	.620	.312	.359	.525	1.00	.499	.445	.506	.661
plis68	.601	.326	.648	.373	.612	.641	.438	.610	.487	.517	.370	.462	.242	.671	.558	.350	.576	.371	.451	.238	.600	.653	.715	.491	.371	.255	.418	.688	.520	.702	.716	.338	.324	.268	.534	.568	.723	.657	.499	1.00	.535	.623	.489
plis72	.293	.063	.312	.156	.440	.322	.377	.468	.318	.177	.103	.286	.191	.431	.427	.222	.359	.152	.268	.257	.298	.286	.473	.377	.186	.052	.379	.433	.416	.520	.435	.157	.120	.218	.425	.309	.376	.460	.445	.535	1.00	.348	.512
plis75	.547	.357	.652	.543	.553	.652	.522	.502	.566	.476	.385	.499	.283	.623	.632	.407	.570	.451	.385	.411	.532	.643	.653	.478	.431	.386	.446	.609	.492	.521	.739	.365	.389	.436	.525	.481	.694	.508	.506	.623	.348	1.00	.602
plis77	.334	.168	.389	.302	.435	.468	.575	.314	.422	.281	.241	.212	.224	.514	.681	.316	.535	.213	.318	.386	.330	.440	.543	.615	.209	.233	.273	.480	.571	.331	.580	.265	.334	.209	.518	.304	.443	.512	.661	.489	.512	.602	1.00

Table C2a

Correlation Matrix of all the items analyzed for the New Zealand sample (excluding the DIF items found between the comparison of U.S. and N.Z. samples)

	plis01	plis05	plis08	plis13	plis14	plis23	plis25	plis27	plis34	plis42	plis46	plis48	plis59	plis63	plis65	plis67	plis68	plis72	plis77
plis01	1.000	.462	.370	.436	.274	.415	.215	.321	.495	.331	.057	.419	.219	.441	.347	.187	.342	.154	.286
plis05	.462	1.000	.280	.314	.277	.381	.203	.268	.417	.315	.090	.412	.251	.434	.302	.166	.400	.132	.283
plis08	.370	.280	1.000	.370	.287	.403	.386	.375	.387	.322	.102	.400	.232	.309	.324	.298	.365	.140	.343
plis13	.436	.314	.370	1.000	.291	.377	.258	.322	.418	.407	.070	.353	.198	.403	.252	.210	.376	.131	.284
plis14	.274	.277	.287	.291	1.000	.262	.390	.398	.292	.408	.154	.340	.402	.293	.189	.332	.393	.271	.316
plis23	.415	.381	.403	.377	.262	1.000	.325	.354	.541	.392	.120	.407	.230	.411	.466	.214	.390	.146	.390
plis25	.215	.203	.386	.258	.390	.325	1.000	.444	.309	.305	.114	.305	.346	.279	.282	.420	.302	.223	.406
plis27	.321	.268	.375	.322	.398	.354	.444	1.000	.364	.350	.104	.325	.228	.319	.263	.298	.415	.175	.422
plis34	.495	.417	.387	.418	.292	.541	.309	.364	1.000	.394	.067	.566	.318	.424	.407	.227	.489	.216	.446
plis42	.331	.315	.322	.407	.408	.392	.305	.350	.394	1.000	.121	.422	.229	.294	.287	.184	.439	.169	.369
plis46	.057	.090	.102	.070	.154	.120	.114	.104	.067	.121	1.000	.114	.103	.051	.099	.136	.178	.043	.107
plis48	.419	.412	.400	.353	.340	.407	.305	.325	.566	.422	.114	1.000	.338	.389	.333	.263	.491	.248	.396
plis59	.219	.251	.232	.198	.402	.230	.346	.228	.318	.229	.103	.338	1.000	.259	.240	.347	.428	.404	.366
plis63	.441	.434	.309	.403	.293	.411	.279	.319	.424	.294	.051	.389	.259	1.000	.344	.143	.355	.241	.348
plis65	.347	.302	.324	.252	.189	.466	.282	.263	.407	.287	.099	.333	.240	.344	1.000	.202	.342	.135	.354
plis67	.187	.166	.298	.210	.332	.214	.420	.298	.227	.184	.136	.263	.347	.143	.202	1.000	.288	.309	.319
plis68	.342	.400	.365	.376	.393	.390	.302	.415	.489	.439	.178	.491	.428	.355	.342	.288	1.000	.266	.447
plis72	.154	.132	.140	.131	.271	.146	.223	.175	.216	.169	.043	.248	.404	.241	.135	.309	.266	1.000	.314
plis77	.286	.283	.343	.284	.316	.390	.406	.422	.446	.369	.107	.396	.366	.348	.354	.319	.447	.314	1.000

Table C2b

Correlation Matrix of all items analyzed including those not analyzed for the New Zealand sample

	plis01	plis05	plis08	plis10	plis13	plis14	plis15	plis19	plis23	plis25	plis27	plis34	plis36	plis37	plis42	plis46	plis48	plis51	plis52	plis78	plis59	plis63	plis65	plis67	plis68	plis72	plis75	plis77
plis01	1.000	.460	.371	.352	.422	.267	.496	.329	.407	.199	.310	.482	.380	.446	.321	.061	.403	.338	.407	.246	.212	.430	.348	.181	.340	.144	.288	.275
plis05	.460	1.000	.273	.266	.315	.270	.370	.368	.374	.190	.261	.408	.391	.431	.315	.094	.407	.351	.409	.208	.249	.430	.301	.173	.403	.140	.352	.279
plis08	.371	.273	1.000	.324	.370	.286	.327	.316	.388	.367	.369	.382	.338	.419	.316	.109	.398	.354	.394	.303	.247	.306	.329	.296	.382	.129	.216	.335
plis10	.352	.266	.324	1.000	.223	.148	.324	.186	.431	.201	.203	.339	.255	.270	.200	.077	.230	.235	.262	.166	.134	.341	.332	.107	.223	.105	.229	.234
plis13	.422	.315	.370	.223	1.000	.281	.366	.313	.366	.252	.319	.407	.401	.395	.398	.076	.346	.527	.369	.313	.188	.390	.253	.203	.378	.117	.366	.281
plis14	.267	.270	.286	.148	.281	1.000	.201	.529	.248	.381	.389	.273	.384	.311	.385	.164	.341	.320	.342	.301	.429	.279	.187	.352	.395	.291	.259	.320
plis15	.496	.370	.327	.324	.366	.201	1.000	.276	.341	.200	.266	.405	.300	.364	.320	.039	.338	.312	.287	.266	.225	.357	.290	.149	.334	.105	.288	.276
plis19	.329	.368	.316	.186	.313	.529	.276	1.000	.308	.217	.306	.404	.441	.361	.414	.164	.397	.381	.449	.431	.304	.335	.244	.280	.499	.263	.323	.346
plis23	.407	.374	.388	.431	.366	.248	.341	.308	1.000	.307	.341	.536	.360	.546	.382	.128	.402	.390	.436	.285	.245	.399	.466	.209	.404	.136	.293	.382
plis25	.199	.190	.367	.201	.252	.381	.200	.217	.307	1.000	.427	.285	.267	.285	.284	.116	.282	.333	.318	.234	.370	.270	.288	.420	.309	.209	.219	.393
plis27	.310	.261	.369	.203	.319	.389	.266	.306	.341	.427	1.000	.344	.323	.343	.332	.111	.304	.302	.437	.301	.243	.313	.268	.292	.426	.159	.296	.409
plis34	.482	.408	.382	.339	.407	.273	.405	.404	.536	.285	.344	1.000	.448	.503	.376	.072	.548	.424	.485	.390	.312	.412	.414	.220	.490	.205	.351	.434
plis36	.380	.391	.338	.255	.401	.384	.300	.441	.360	.267	.323	.448	1.000	.349	.287	.053	.378	.419	.405	.288	.315	.399	.240	.189	.455	.223	.384	.353
plis37	.446	.431	.419	.270	.395	.311	.364	.361	.546	.285	.343	.503	.349	1.000	.462	.090	.495	.377	.464	.266	.229	.382	.429	.260	.481	.106	.307	.375
plis42	.321	.315	.316	.200	.398	.385	.320	.414	.382	.284	.332	.376	.287	.462	1.000	.128	.416	.383	.323	.348	.243	.277	.288	.168	.447	.148	.255	.360
plis46	.061	.094	.109	.077	.076	.164	.039	.164	.128	.116	.111	.072	.053	.090	.128	1.000	.122	.038	.123	.074	.108	.055	.104	.142	.187	.046	.049	.113
plis48	.403	.407	.398	.230	.346	.341	.338	.397	.402	.282	.304	.548	.378	.495	.416	.122	1.000	.405	.492	.339	.330	.380	.347	.254	.488	.235	.334	.387
plis51	.338	.351	.354	.235	.527	.320	.312	.381	.390	.333	.302	.424	.419	.377	.383	.038	.405	1.000	.405	.375	.395	.367	.281	.253	.433	.264	.387	.345
plis52	.407	.409	.394	.262	.369	.342	.287	.449	.436	.318	.437	.485	.405	.464	.323	.123	.492	.405	1.000	.332	.273	.376	.382	.308	.428	.185	.328	.402
plis78	.246	.208	.303	.166	.313	.301	.266	.431	.285	.234	.301	.390	.288	.266	.348	.074	.339	.375	.332	1.000	.337	.288	.223	.261	.372	.366	.323	.317
plis59	.212	.249	.247	.134	.188	.429	.225	.304	.245	.370	.243	.312	.315	.229	.243	.108	.330	.395	.273	.337	1.000	.253	.252	.363	.421	.427	.266	.384
plis63	.430	.430	.306	.341	.390	.279	.357	.335	.399	.270	.313	.412	.399	.382	.277	.055	.380	.367	.376	.288	.253	1.000	.341	.134	.352	.234	.392	.346
plis65	.348	.301	.329	.332	.253	.187	.290	.244	.466	.288	.268	.414	.240	.429	.288	.104	.347	.281	.382	.223	.252	.341	1.000	.205	.361	.137	.224	.355
plis67	.181	.173	.296	.107	.203	.352	.149	.280	.209	.420	.292	.220	.189	.260	.168	.142	.254	.253	.308	.261	.363	.134	.205	1.000	.302	.279	.211	.308
plis68	.340	.403	.382	.223	.378	.395	.334	.499	.404	.309	.426	.490	.455	.481	.447	.187	.488	.433	.428	.372	.421	.352	.361	.302	1.000	.283	.380	.471
plis72	.144	.140	.129	.105	.117	.291	.105	.263	.136	.209	.159	.205	.223	.106	.148	.046	.235	.264	.185	.366	.427	.234	.137	.279	.283	1.000	.218	.300
plis75	.288	.352	.216	.229	.366	.259	.288	.323	.293	.219	.296	.351	.384	.307	.255	.049	.334	.387	.328	.323	.266	.392	.224	.211	.380	.218	1.000	.341
plis77	.275	.279	.335	.234	.281	.320	.276	.346	.382	.393	.409	.434	.353	.375	.360	.113	.387	.345	.402	.317	.384	.346	.355	.308	.471	.300	.341	1.000

Table C3a

Correlation Matrix of all items analyzed for the Spanish speaking (Mexico) sample

	plis01	plis05	plis06	plis08	plis10	plis13	plis14	plis15	plis19	plis23	plis25	plis27	plis34	plis36	plis42	plis46	plis48	plis52	plis53	plis59	plis61	plis63	plis65	plis67	plis68	plis72	plis75	plis77
plis01	1.000	.627	.541	.488	.589	.595	.558	.528	.476	.545	.449	.190	.682	.604	.520	.466	.591	.591	.533	.518	.514	.556	.536	.303	.468	.340	.560	.379
plis05	.627	1.000	.582	.644	.666	.621	.649	.461	.524	.631	.593	.241	.655	.801	.549	.461	.593	.620	.636	.596	.576	.642	.568	.382	.621	.457	.653	.526
plis06	.541	.582	1.000	.561	.558	.499	.526	.481	.442	.500	.446	.245	.579	.574	.479	.454	.557	.580	.522	.523	.539	.554	.484	.312	.479	.295	.570	.472
plis08	.488	.644	.561	1.000	.677	.601	.593	.449	.487	.612	.556	.276	.623	.679	.599	.481	.648	.675	.620	.622	.539	.578	.608	.434	.599	.394	.592	.563
plis10	.589	.666	.558	.677	1.000	.641	.642	.520	.489	.762	.627	.213	.667	.679	.651	.468	.665	.684	.639	.624	.582	.672	.759	.385	.579	.422	.613	.553
plis13	.595	.621	.499	.601	.641	1.000	.662	.502	.484	.601	.487	.201	.630	.610	.615	.455	.613	.605	.624	.545	.563	.649	.624	.429	.597	.418	.550	.562
plis14	.558	.649	.526	.593	.642	.662	1.000	.517	.658	.681	.649	.220	.678	.692	.669	.581	.662	.686	.711	.689	.684	.694	.648	.560	.708	.534	.611	.642
plis15	.528	.461	.481	.449	.520	.502	.517	1.000	.456	.453	.398	.178	.595	.531	.488	.368	.562	.495	.461	.458	.495	.532	.511	.307	.426	.304	.461	.463
plis19	.476	.524	.442	.487	.489	.484	.658	.456	1.000	.558	.528	.132	.617	.682	.620	.554	.627	.647	.628	.614	.623	.585	.544	.364	.566	.450	.591	.454
plis23	.545	.631	.500	.612	.762	.601	.681	.453	.558	1.000	.668	.218	.666	.674	.647	.490	.658	.678	.646	.652	.602	.613	.707	.429	.603	.464	.589	.517
plis25	.449	.593	.446	.556	.627	.487	.649	.398	.528	.668	1.000	.219	.570	.620	.640	.480	.601	.615	.668	.679	.542	.585	.659	.538	.666	.608	.614	.657
plis27	.190	.241	.245	.276	.213	.201	.220	.178	.132	.218	.219	1.000	.201	.251	.167	.178	.190	.255	.192	.205	.176	.176	.195	.154	.232	.111	.258	.226
plis34	.682	.655	.579	.623	.667	.630	.678	.595	.617	.666	.570	.201	1.000	.713	.633	.545	.725	.708	.671	.638	.633	.704	.656	.416	.591	.436	.685	.543
plis36	.604	.801	.574	.679	.679	.610	.692	.531	.682	.674	.620	.251	.713	1.000	.637	.545	.664	.764	.728	.643	.676	.734	.610	.395	.720	.447	.714	.555
plis42	.520	.549	.479	.599	.651	.615	.669	.488	.620	.647	.640	.167	.633	.637	1.000	.549	.717	.687	.767	.684	.619	.636	.676	.586	.678	.568	.590	.635
plis46	.466	.461	.454	.481	.468	.455	.581	.368	.554	.490	.480	.178	.545	.545	.549	1.000	.607	.622	.658	.581	.528	.519	.506	.396	.482	.422	.476	.373
plis48	.591	.593	.557	.648	.665	.613	.662	.562	.627	.658	.601	.190	.725	.664	.717	.607	1.000	.760	.734	.694	.621	.660	.702	.508	.617	.525	.644	.563
plis52	.591	.620	.580	.675	.684	.605	.686	.495	.647	.678	.615	.255	.708	.764	.687	.622	.760	1.000	.788	.729	.646	.708	.693	.453	.690	.473	.731	.616
plis53	.533	.636	.522	.620	.639	.624	.711	.461	.628	.646	.668	.192	.671	.728	.767	.658	.734	.788	1.000	.710	.623	.719	.654	.579	.748	.551	.681	.658
plis59	.518	.596	.523	.622	.624	.545	.689	.458	.614	.652	.679	.205	.638	.643	.684	.581	.694	.729	.710	1.000	.627	.623	.691	.583	.723	.631	.612	.675
plis61	.514	.576	.539	.539	.582	.563	.684	.495	.623	.602	.542	.176	.633	.676	.619	.528	.621	.646	.623	.627	1.000	.721	.613	.475	.613	.476	.612	.528
plis63	.556	.642	.554	.578	.672	.649	.694	.532	.585	.613	.585	.176	.704	.734	.636	.519	.660	.708	.719	.623	.721	1.000	.708	.477	.694	.479	.734	.594
plis65	.536	.568	.484	.608	.759	.624	.648	.511	.544	.707	.659	.195	.656	.610	.676	.506	.702	.693	.654	.691	.613	.708	1.000	.494	.665	.498	.625	.601
plis67	.303	.382	.312	.434	.385	.429	.560	.307	.364	.429	.538	.154	.416	.395	.586	.396	.508	.453	.579	.583	.475	.477	.494	1.000	.651	.742	.460	.692
plis68	.468	.621	.479	.599	.579	.597	.708	.426	.566	.603	.666	.232	.591	.720	.678	.482	.617	.690	.748	.723	.613	.694	.665	.651	1.000	.628	.708	.751
plis72	.340	.457	.295	.394	.422	.418	.534	.304	.450	.464	.608	.111	.436	.447	.568	.422	.525	.473	.551	.631	.476	.479	.498	.742	.628	1.000	.469	.565
plis75	.560	.653	.570	.592	.613	.550	.611	.461	.591	.589	.614	.258	.685	.714	.590	.476	.644	.731	.681	.612	.612	.734	.625	.460	.708	.469	1.000	.658
plis77	.379	.526	.472	.563	.553	.562	.642	.463	.454	.517	.657	.226	.543	.555	.635	.373	.563	.616	.658	.675	.528	.594	.601	.692	.751	.565	.658	1.000

Table C3b

Correlation Matrix of all items analyzed including those not analyzed for the Spanish speaking (Mexico) sample

	plis01	plis05	plis06	plis08	plis10	plis13	plis14	plis15	plis19	plis23	plis25	plis27	plis34	plis36	plis42	plis46	plis48	plis52	plis53	plis59	plis61	plis63	plis65	plis67	plis68	plis72	plis75	plis77	plis79	plis21	plis80	plis81
plis01	1.000	.621	.535	.480	.581	.588	.553	.545	.484	.537	.434	.185	.678	.596	.514	.460	.588	.584	.524	.508	.509	.549	.528	.285	.459	.324	.553	.366	.318	-.212	.198	.450
plis05	.621	1.000	.578	.639	.663	.612	.641	.481	.533	.623	.584	.237	.648	.796	.539	.450	.586	.612	.627	.586	.568	.635	.559	.359	.612	.439	.646	.515	.454	-.201	.230	.519
plis06	.535	.578	1.000	.555	.549	.493	.525	.494	.449	.494	.434	.240	.575	.570	.477	.452	.557	.576	.515	.517	.539	.550	.478	.304	.478	.286	.566	.464	.327	-.190	.241	.453
plis08	.480	.639	.555	1.000	.671	.594	.589	.465	.495	.605	.545	.272	.618	.673	.595	.476	.645	.671	.612	.614	.535	.571	.601	.426	.596	.382	.586	.554	.444	-.242	.342	.507
plis10	.581	.663	.549	.671	1.000	.636	.643	.540	.501	.760	.614	.206	.664	.675	.652	.466	.668	.681	.633	.617	.583	.669	.758	.373	.577	.413	.608	.543	.480	-.240	.302	.580
plis13	.588	.612	.493	.594	.636	1.000	.654	.523	.493	.592	.472	.196	.623	.599	.606	.443	.606	.597	.614	.533	.554	.641	.616	.408	.585	.395	.541	.552	.440	-.227	.184	.594
plis14	.553	.641	.525	.589	.643	.654	1.000	.545	.672	.674	.645	.218	.673	.684	.658	.569	.653	.680	.704	.680	.675	.688	.641	.539	.696	.512	.603	.636	.579	-.199	.211	.643
plis15	.545	.481	.494	.465	.540	.523	.545	1.000	.457	.473	.425	.183	.615	.556	.515	.388	.584	.513	.485	.484	.518	.552	.532	.364	.462	.355	.479	.484	.366	-.106	.180	.481
plis19	.484	.533	.449	.495	.501	.493	.672	.457	1.000	.568	.547	.134	.625	.696	.634	.564	.637	.656	.642	.629	.634	.594	.554	.396	.587	.483	.600	.465	.431	-.204	.151	.567
plis23	.537	.623	.494	.605	.760	.592	.674	.473	.568	1.000	.660	.214	.660	.666	.639	.479	.652	.672	.637	.643	.595	.605	.701	.409	.593	.446	.580	.506	.504	-.177	.212	.533
plis25	.434	.584	.434	.545	.614	.472	.645	.425	.547	.660	1.000	.212	.563	.609	.634	.472	.598	.608	.658	.668	.536	.577	.651	.520	.659	.595	.606	.646	.625	-.195	.231	.612
plis27	.185	.237	.240	.272	.206	.196	.218	.183	.134	.214	.212	1.000	.197	.247	.163	.175	.188	.252	.187	.200	.173	.171	.191	.150	.230	.104	.254	.221	.160	-.370	.430	.196
plis34	.678	.648	.575	.618	.664	.623	.673	.615	.625	.660	.563	.197	1.000	.707	.626	.537	.721	.703	.665	.630	.627	.699	.650	.400	.583	.421	.680	.535	.421	-.245	.213	.583
plis36	.596	.796	.570	.673	.675	.599	.684	.556	.696	.666	.609	.247	.707	1.000	.628	.535	.658	.759	.720	.632	.669	.728	.601	.366	.712	.423	.708	.544	.469	-.245	.279	.599
plis42	.514	.539	.477	.595	.652	.606	.658	.515	.634	.639	.634	.163	.626	.628	1.000	.536	.709	.682	.761	.676	.609	.628	.669	.568	.664	.547	.582	.628	.519	-.211	.180	.636
plis46	.460	.450	.452	.476	.466	.443	.569	.388	.564	.479	.472	.175	.537	.535	.536	1.000	.598	.616	.651	.571	.517	.510	.496	.370	.465	.398	.466	.362	.341	-.278	.209	.479
plis48	.588	.586	.557	.645	.668	.606	.653	.584	.637	.652	.598	.188	.721	.658	.709	.598	1.000	.756	.729	.689	.612	.654	.697	.493	.606	.511	.638	.556	.476	-.279	.257	.625
plis52	.584	.612	.576	.671	.681	.597	.680	.513	.656	.672	.608	.252	.703	.759	.682	.616	.756	1.000	.784	.724	.640	.703	.687	.440	.685	.460	.726	.608	.533	-.275	.305	.608
plis53	.524	.627	.515	.612	.633	.614	.704	.485	.642	.637	.658	.187	.665	.720	.761	.651	.729	.784	1.000	.701	.615	.712	.646	.565	.741	.534	.674	.649	.560	-.253	.236	.704
plis59	.508	.586	.517	.614	.617	.533	.680	.484	.629	.643	.668	.200	.630	.632	.676	.571	.689	.724	.701	1.000	.618	.614	.683	.567	.715	.618	.602	.666	.637	-.263	.266	.615
plis61	.509	.566	.539	.535	.583	.554	.675	.518	.634	.595	.536	.173	.627	.669	.609	.517	.612	.640	.615	.618	1.000	.716	.606	.456	.600	.456	.605	.520	.439	-.162	.193	.592
plis63	.549	.635	.550	.571	.669	.641	.688	.552	.594	.605	.577	.171	.699	.728	.628	.510	.654	.703	.712	.614	.716	1.000	.702	.463	.688	.465	.729	.586	.482	-.175	.209	.651
plis65	.528	.559	.478	.601	.758	.616	.641	.532	.554	.701	.651	.191	.650	.601	.669	.496	.697	.687	.646	.683	.606	.702	1.000	.480	.657	.483	.617	.592	.549	-.202	.218	.629
plis67	.285	.359	.304	.426	.373	.408	.539	.364	.396	.409	.520	.150	.400	.366	.568	.370	.493	.440	.565	.567	.456	.463	.480	1.000	.624	.708	.445	.694	.740	-.123	.154	.521
plis68	.459	.612	.478	.596	.577	.585	.696	.462	.587	.593	.659	.230	.583	.712	.664	.465	.606	.685	.741	.715	.600	.688	.657	.624	1.000	.600	.703	.749	.651	-.221	.192	.696
plis72	.324	.439	.286	.382	.413	.395	.512	.355	.483	.446	.595	.104	.421	.423	.547	.398	.511	.460	.534	.618	.456	.465	.483	.708	.600	1.000	.455	.555	.791	-.191	.204	.508
plis75	.553	.646	.566	.586	.608	.541	.603	.479	.600	.580	.606	.254	.680	.708	.582	.466	.638	.726	.674	.602	.605	.729	.617	.445	.703	.455	1.000	.651	.501	-.213	.230	.589
plis77	.366	.515	.464	.554	.543	.552	.636	.484	.465	.506	.646	.221	.535	.544	.628	.362	.556	.608	.649	.666	.520	.586	.592	.694	.749	.555	.651	1.000	.683	-.159	.235	.669
plis79	.318	.454	.327	.444	.480	.440	.579	.366	.431	.504	.625	.160	.421	.469	.519	.341	.476	.533	.560	.637	.439	.482	.549	.740	.651	.791	.501	.683	1.000	-.156	.204	.530
plis21	-.212	-.201	-.190	-.242	-.240	-.227	-.199	-.106	-.204	-.177	-.195	-.370	-.245	-.245	-.211	-.278	-.279	-.275	-.253	-.263	-.162	-.175	-.202	-.123	-.221	-.191	-.213	-.159	-.156	1.000	-.725	-.192
plis80	.198	.230	.241	.342	.302	.184	.211	.180	.151	.212	.231	.430	.213	.279	.180	.209	.257	.305	.236	.266	.193	.209	.218	.154	.192	.204	.230	.235	.204	-.725	1.000	.161
plis81	.450	.519	.453	.507	.580	.594	.643	.481	.567	.533	.612	.196	.583	.599	.636	.479	.625	.608	.704	.615	.592	.651	.629	.521	.696	.508	.589	.669	.530	-.192	.161	1.000

Appendix D

Item Parameter and Standard Error Estimates

Table D1

Item Parameter and Standard Error Estimates of U.S. and N.Z. comparison group.

	b_1 (SE)		b_2 (SE)		b_3 (SE)		a (SE)	
Item	NZ	U.S.	NZ	U.S.	NZ	U.S.	NZ	U.S.
1	.564(.06)	.144(.07)	.647(.06)	.816(.10)	.767(.06)	1.220(.13)	3.105(.32)	4.121(.36)
5	.728(.06)	.171(.07)	.768(06)	.766(.10)	.919(07)	1.056(.12)	3.046(.30)	4.182(.38)
8	.704(.07)	.247(.07)	.768(.07)	.927(.11)	.847(.07)	1.159(.15)	2.839(.28)	4.233(.38)
10	.354(.07)	-.044(.06)	.487(.07)	.517(.07)	.716(.08)	.848(.10)	1.888(.17)	5.242(.41)
13	.815(.07)	.480(.09)	.891(.07)	1.173(.15)	.980(.07)	1.443(.15)	3.040(.29)	3.745(.40)
14	.989(.10)	.697(.08)	1.041(.25)	1.146(.15)	1.097(.22)	1.436(.18)	3.505(1.54)	4.608(.58)
15	.612(.07)	.500(.08)	.691(.07)	1.039(.13)	.792(.07)	1.327(.16)	2.542(.22)	4.203(.43)

	b_1 (SE)		b_2 (SE)		b_3 (SE)		a (SE)	
Item	NZ	U.S.	NZ	U.S.	NZ	U.S.	NZ	U.S.
19	.785(.04)	.747(.09)	.814(.04)	1.177(.13)	.875(.05)	1.375(.16)	4.241(.89)	4.341(.55)
23	.632(.05)	.149(.06)	.685(.05)	.857(.10)	.751(.05)	1.154(.14)	3.652(.34)	4.752(.44)
25	1.083(.09)	.549(.09)	1.140(.10)	1.024(.15)	1.217(.11)	1.235(.19)	2.962(.37)	3.833(.44)
27	.942(07)	.308(08)	.988(.07)	.919(.13)	1.070(.07)	1.358(.18)	3.338(.36)	3.608(.31)
34	.786(.05)	.546(.08)	.834(.05)	.994(.10)	.890(.05)	1.364(.15)	5.079(.57)	5.202(.63)
36	.910(.06)	.232(.07)	.950(.06)	.839(.10)	1.023(.08)	1.200(.15)	3.902(.39)	4.655(.63)
42	.974(.07)	.909(.10)	1.049(.07)	1.531(.22)	1.114(.08)	1.719(.28)	3.568(.39)	4.176(.67)
46	3.575(1.16)	1.094(.19)	4.325(1.35)	1.707(.30)	5.314(1.58)	1.921(.37)	.695(.22)	2.875(.50)
48	.894(.04)	.547(.07)	.926(.05)	1.114(.10)	.988(.05)	1.447(.17)	5.496(.67)	5.582(.72)
52	.870(.05)	.156(.05)	.932(.05)	.708(.08)	.999(.05)	1.001(.10)	5.097(.62)	6.282(.59)

	b_1 (SE)		b_2 (SE)		b_3 (SE)		a (SE)	
Item	NZ	U.S.	NZ	U.S.	NZ	U.S.	NZ	U.S.
59	1.212(.06)	.955(.05)	1.281(.06)	1.358(.20)	1.346(.10)	1.461(.21)	5.190(96)	4.441(.86)
63	.620(.06)	.091(.06)	.691(.06)	.627(.08)	.799(.06)	1.078(.09)	3.193(.31)	5.494(.50)
65	.655(.07)	.306(.07)	.709(.07)	.885(.11)	.809(.07)	1.362(.17)	2.490(.25)	4.210(.44)
67	1.423(.14)	1.098(.20)	1.454(.14)	1.460(.25)	1.510(.16)	1.640(.29)	3.037(.49)	2.987(.54)
68	.946(.05)	.515(.07)	.996(.05)	1.146(.12)	1.068(.05)	1.535(.19)	5.188(.61)	6.204(.77)
72	1.464(.15)	1.278(.21)	1.518(.08)	1.755(.35)	1.579(.17)	2.179(.58)	3.593(.76)	3.616(.67)
75	.922(.08)	.276(.07)	.975(.08)	.754(.12)	1.094(.09)	1.019(.14)	2.825(.27)	3.758(.37)
77	.938(.06)	.693(.11)	.981(.06)	1.306(.19)	1.066(.07)	1.491(.22)	4.007(.45)	3.205(.42)

Note. The equated item parameters for the U.S. sample has been reported in this table. Constants used to derive the transformation coefficients were A = 0.7627 and K = -0.1233.

Table D2

Item Parameter and Standard Error Estimates of the US-NZ and Mexico comparison group

	b_1 (SE)		b_2 (SE)		b_3 (SE)		a (SE)	
Item	US-NZ	Mexico	US-NZ	Mexico	US-NZ	Mexico	US-NZ	Mexico
1	1.084(.06)	.821(.09)	1.422(.07)	1.385(.09)	1.640(.07)	2.070(.15)	2.306(.17)	2.349(.24)
5	1.241(.06)	1.111(.08)	1.530(.07)	1.610(.09)	1.772(.08)	2.269(.15)	2.278(.18)	3.159(.32)
6	.655(.08)	.253(.08)	1.521(.12)	.960(.09)	2.045(.17)	1.608(.12)	2.305(.28)	2.290(.21)
8	1.240(.06)	.745(.07)	1.559(.08)	1.382(.10)	1.693(.08)	2.097(.14)	2.264(.17)	2.738(.25)
10	.581(.06)	.948(.06)	1.316(.07)	1.430(.08)	1.710(.10)	2.000(.11)	4.017(.42)	3.580(.34)
13	1.481(.07)	1.173(.09)	1.796(.09)	1.860(.12)	1.951(.09)	2.448(.18)	2.183(.18)	2.595(.27)
14	1.883(.07)	1.417(.07)	2.130(.09)	1.966(.11)	2.270(.10)	2.427(.15)	3.308(.41)	4.005(.46)
15	1.275(.08)	.705(.10)	1.895(.12)	1.541(.14)	2.217(.15)	2.457(.23)	3.579(.53)	1.902(.20)
19	1.555(.08)	1.619(.09)	2.025(.12)	2.171(.13)	2.240(.15)	2.658(.16)	3.968(.53)	2.978(.36)
23	1.143(.05)	1.224(.08)	1.470(.06)	1.720(.11)	1.603(.06)	2.283(.14)	2.388(.20)	3.090(.37)
25	1.694(.08)	1.522(.10)	1.951(.10)	2.084(.13)	2.089(.11)	2.569(.18)	2.388(.24)	2.988(.32)

	b_1 (SE)		b_2 (SE)		b_3 (SE)		a (SE)	
Item	US-NZ	Mexico	US-NZ	Mexico	US-NZ	Mexico	US-NZ	Mexico
27	1.512(.07)	.081(.25)	1.826(.09)	1.720(.37)	2.036(.10)	2.935(.64)	2.305(.18)	.708(.14)
34	1.461(.05)	.844(.06)	1.669(.05)	1.495(.08)	1.804(.06)	2.042(.11)	3.565(.32)	3.537(.33)
36	.946(.07)	1.220(.07)	1.672(.10)	1.747(.08)	2.092(.12)	2.288(.12)	3.760(.49)	4.019(.44)
42	1.778(.08)	1.427(.08)	2.013(.09)	1.917(.09)	2.113(.10)	2.419(.16)	2.697(.25)	3.979(.38)
46	3.319(.46)	1.277(.11)	4.116(.59)	1.998(.15)	4.882(.67)	2.558(.22)	.874(.14)	2.038(.24)
48	1.572(.05)	1.174(.07)	1.805(.06)	1.700(.09)	1.947(.07)	2.171(.11)	5.098(.37)	3.792(.42)
52	.861(.05)	1.018(.06)	1.534(.07)	1.549(.08)	1.864(.09)	1.989(.10)	5.098(.61)	3.951(.40)
53	1.542(.10)	1.395(.07)	2.147(.17)	1.737(.08)	2.579(.24)	2.330(.12)	2.816(.47)	4.551(.52)
59	2.064(.08)	1.365(.08)	2.289(.11)	1.888(.10)	2.398(.13)	2.444(.15)	3.328(.46)	3.429(.40)
61	1.411(.07)	1.143(.07)	2.080(.14)	1.804(.11)	2.405(.17)	2.252(.16)	3.929(.74)	2.856(.29)
63	1.107(.06)	1.066(.06)	1.409(.06)	1.592(.08)	1.640(.07)	2.066(.12)	2.447(.18)	3.733(.39)
65	1.257(.07)	1.216(.07)	1.522(.08)	1.678(.10)	1.719(.09)	2.183(.13)	1.915(.15)	3.324(.39)
67	2.270(.14)	2.239(.12)	2.441(.15)	2.573(.18)	2.552(17)	2.911(.24)	2.235(.29)	3.595(.72)

	b_1 (SE)		b_2 (SE)		b_3 (SE)		a (SE)	
Item	US-NZ	Mexico	US-NZ	Mexico	US-NZ	Mexico	US-NZ	Mexico
68	1.606(.05)	1.568(.07)	1.888(.07)	2.160(.10)	2.054(.07)	2.571(.16)	3.648(.34)	4.851(.66)
72	2.451(.17)	2.193(.13)	2.668(.20)	2.575(.19)	2.807(.23)	2.930(.26)	2.461(.41)	3.474(.67)
75	0.991(.07)	1.128(.07)	1.577(.11)	1.630(.08)	1.892(.13)	1.990(.11)	3.044(.42)	3.696(.38)
77	1.633(.07)	1.682(.09)	1.901(.08)	2.212(.13)	2.044(.09)	2.622(.18)	2.673(.24)	3.541(.51)

Note. The equated item parameters for the Mexico sample has been reported in this table. Constants used to derive the transformation coefficients were A = 0.8968 and K = 0.6645.

Appendix E

Table E

NCDIF Results by comparison groups

PLIS Items	NZ and U.S.	US-NZ and Mexico
1	.007 (.0053)	.003 (.0095)
5	.019 (.0000)	.012 (.0000)
6		.114 (.0000)
8	.001 (.0000)	.009 (.0000)
10	.063 (.0000)	.040 (.0000)
13	.010 (.0000)	.002 (.0000)
14	.002 (.0000)	.014 (.0000)
15	.082 (.0000)	.113 (.0000)
19	.057 (.0000)	.007 (.3546)
23	.011 (.3282)	.051 (.0000)
25	.022 (.0000)	.006 (.0000)
27	.036 (.0000)	.476 (.0000)
34	.017 (.1436)	.046 (.0000)
36	.061 (.0000)	.019 (.0000)
42	.039 (.0000)	.001 (.4027)
46	.018 (.0000)	.106 (.0000)
48	.017 (.4354)	.014 (.0000)
52	.148 (.0000)	.005 (.0000)
53		.018 (.1155)
59	.006 (.0000)	.056 (.0000)
61		.035 (.0000)
63	.014 (.0000)	.029 (.0000)
65	.022 (.0000)	.045 (.0000)
67	.003 (.0000)	.013 (.0000)
68	.016 (.0530)	.021 (.0000)
72	.001 (.0352)	.001 (.0000)
75	.072 (.0000)	.010 (.0000)
77	.012 (.4181)	.030 (.0000)

Note. Value in parenthesis is the significance level of the χ^2 test corresponding to each NCDIF value for the given item.

Figure 1. Scree Plot of Eigenvalues for 28-item English version of the Perceived Leader Integrity Scale used by the United States group (n = 329).

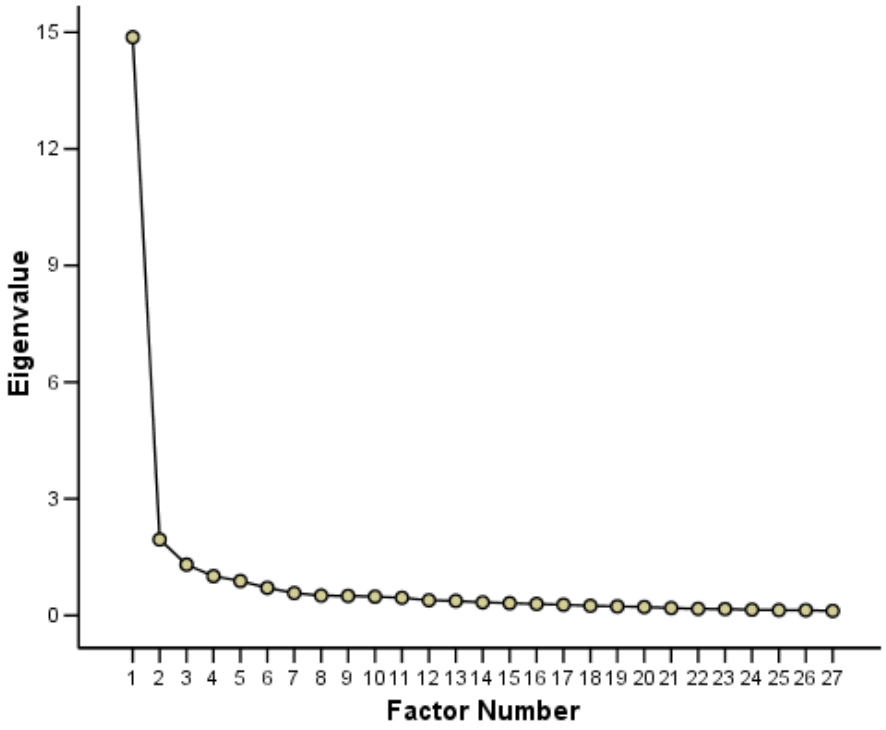

Figure 2. Scree Plot of Eigenvalues for 25-item English version of the Perceived Leader Integrity Scale used by the New Zealand group (*n* = 1,354).

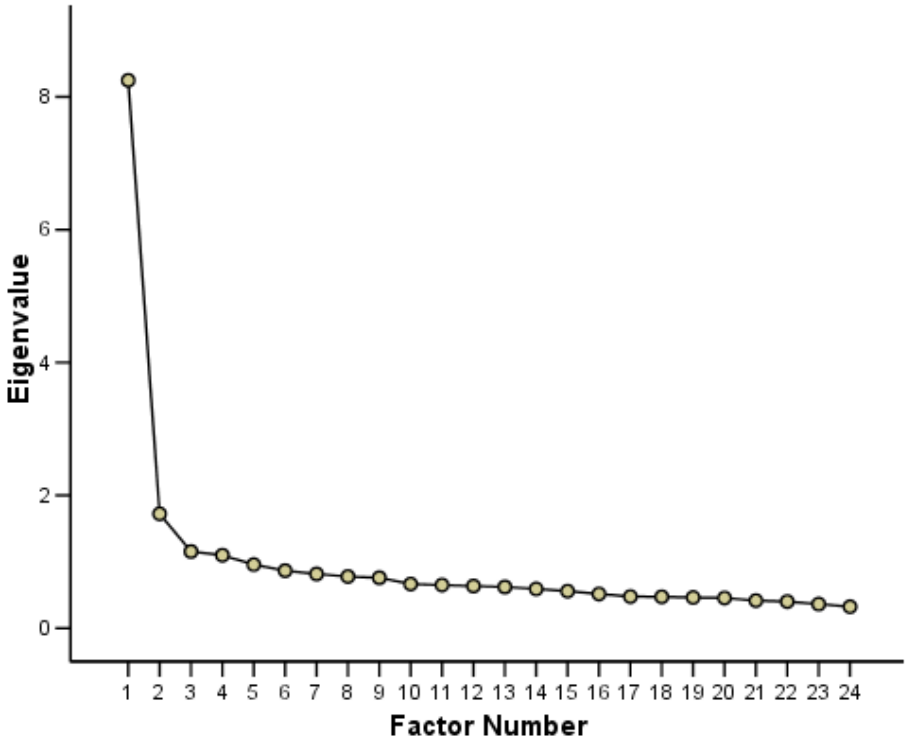

Figure 3. Scree Plot of Eigenvalues for 28-item Spanish version of the Perceived Leader Integrity Scale used by the Spanish speaking (Mexico) group (*n* = 381).

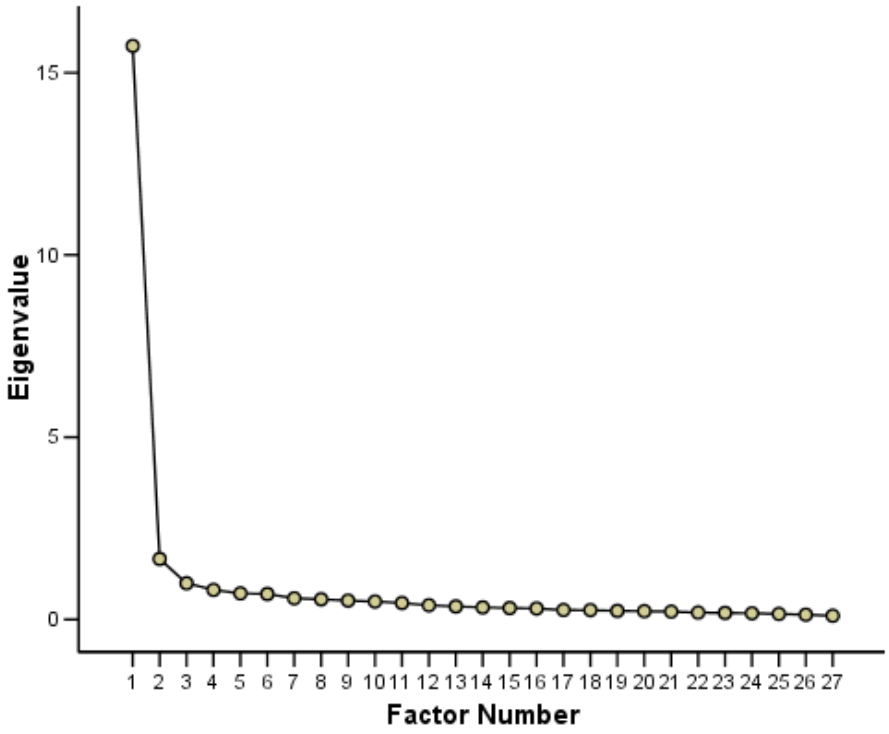

Figure 4. Scree Plot of Eigenvalues for 28-item English version of the Perceived Leader Integrity Scale (not including items 6, 53, and 61) used by the English speaking (US-NZ) group (n = 1,537).

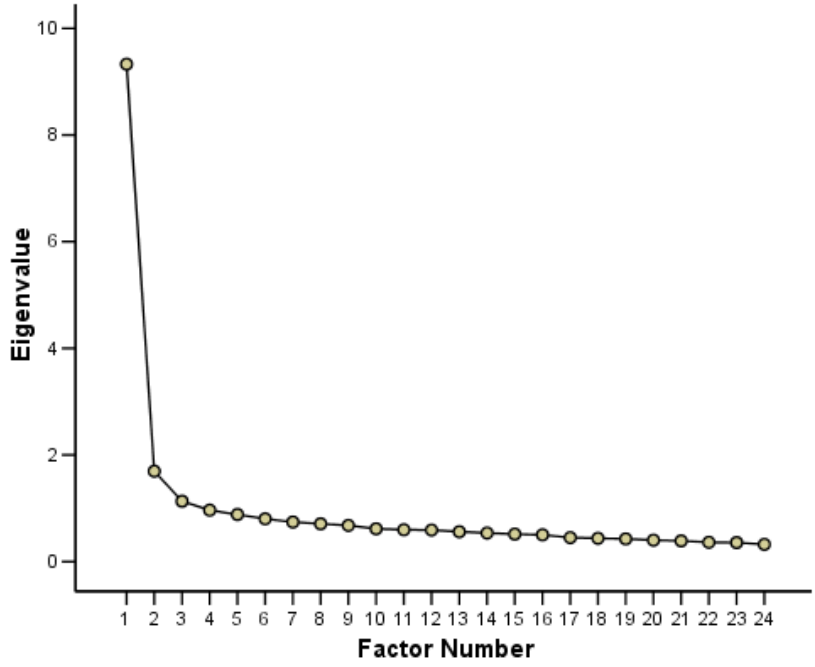

Figure 5. The expected item true score as a function of different levels of theta for Item 6.

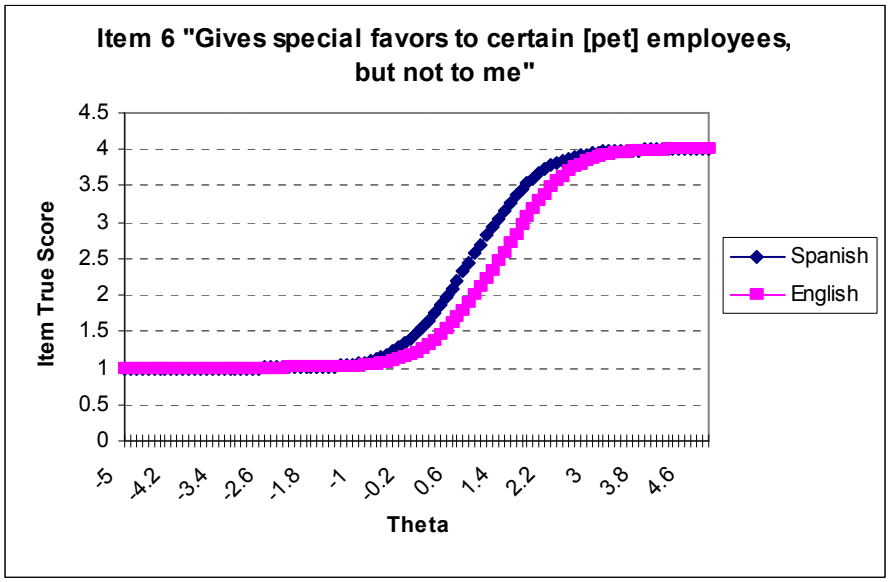

Figure 6. The expected item true score as a function of different levels of theta for Item 15.

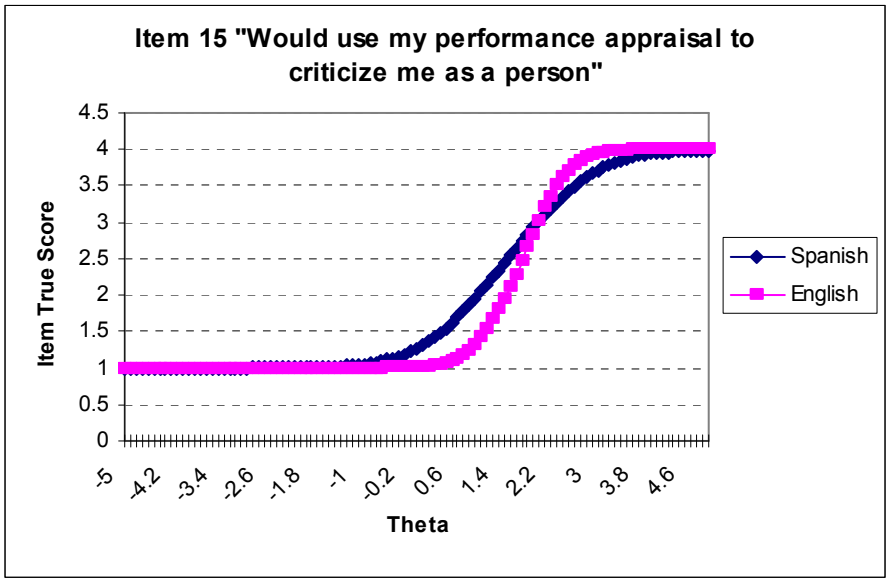

Figure 7. The expected item true score as a function of different levels of theta for Item 27.

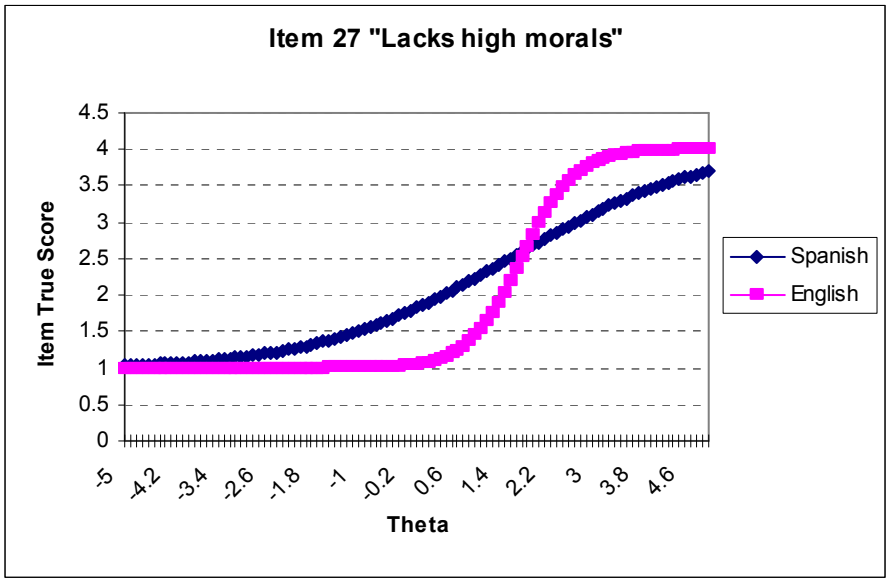

Figure 8. The expected item true score as a function of different levels of theta for Item 46.

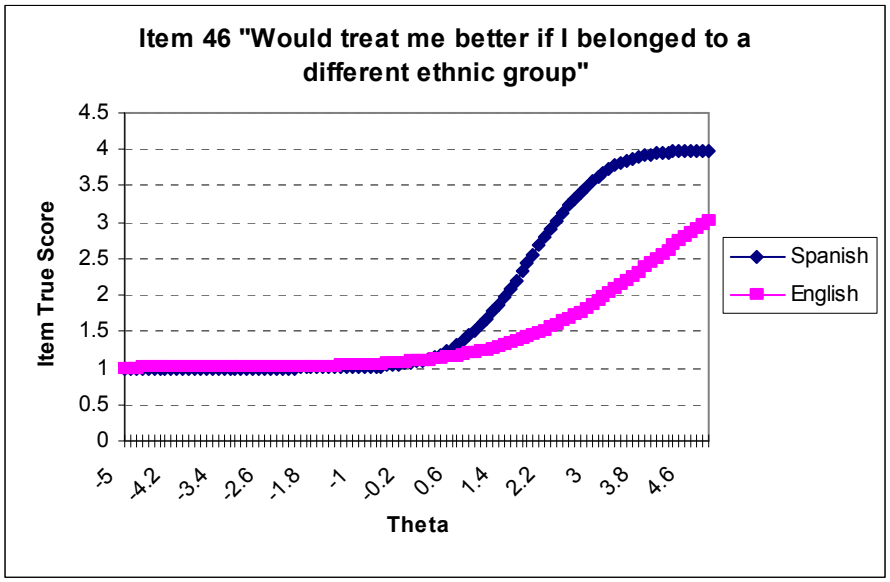

Figure 9. The expected item true score as a function of different levels of theta for Item 59.

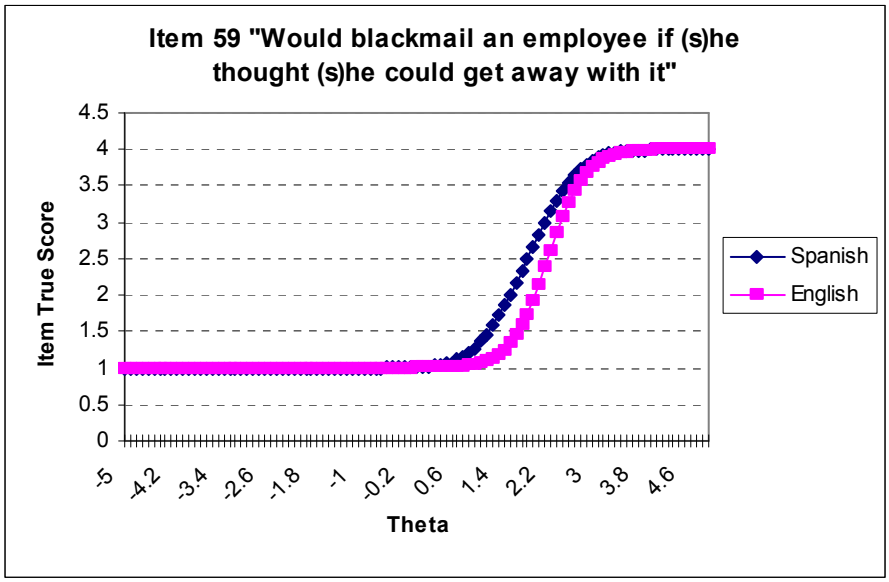

Figure 10. Mean Perceived Leader Integrity as a function of the different groups using the versions of the Perceived Leader Integrity Scale.