Investigating the Reliability and Validity of the Leadership Practices Inventory®

Barry Z. Posner
Leavey School of Business, Santa Clara University, 500 El Camino Real, Santa Clara, CA 95053, USA; bposner@scu.edu; Tel.: +1-408-554-4634

Abstract: This review explains the origins of the Leadership Practices Inventory (LPI) as an empirical instrument to measure The Five Practices of Exemplary Leadership framework, a major transformational leadership model. The essential psychometric properties of the LPI are investigated using both the LPI normative database, with nearly 2.8 million respondents, as well as reviewing pertinent findings of several hundred studies conducted worldwide by scholars utilizing the LPI in their research. Issues of both reliability and validity are considered, with the conclusion that the LPI is quite robust and applicable across a variety of settings and populations.

Keywords: leadership; leadership practices inventory; The Five Practices of Exemplary Leadership; leadership and gender; leadership and ethnicity; cross-cultural leadership; leadership and engagement

1. Introduction

The Leadership Practices Inventory (LPI) is conceptually based on the transformational leadership model of Kouzes and Posner [1]. This model emerged from their analysis of thousands of case studies of people’s personal-best leadership experiences; the times when these individuals accomplished something extraordinary [2–4]. The model postulates that there are five exemplary leadership practices: Model the Way, Inspire a Shared Vision, Challenge the Process, Enable Others to Act, and Encourage the Heart. These practices provide the groundwork for organizational success, by recommending what behaviors and actions people need to do to become effective leaders [2]. The five leadership practices are described in these ways:

1.1. Model the Way

To model the behavior they expect of others effectively, leaders must first be clear about guiding principles. Leaders establish principles concerning the way people (constituents, peers, colleagues, and customers alike) should be treated and the way goals should be pursued. They create standards of excellence and set an example for others to follow [1].

1.2. Inspire a Shared Vision

Leaders passionately believe that they can make a difference. They envision the future, creating an ideal and unique image of what the organization can become. They breathe life into their visions and get people to see exciting possibilities beyond the horizon. They listen to the hopes and aspirations of others, so that by incorporating these, people can enlist in a shared dream about the future [1].
1.3. Challenge the Process

Leaders aspire to improve upon the status quo by searching for opportunities to grow and innovate, many of which are outside of their customary boundaries. They experiment and take risks, and gain momentum through achieving small wins. They look upon setbacks as learning opportunities, for both themselves and their constituents [1].

1.4. Enable Others to Act

Leaders foster collaboration, build trust, and create spirited teams. They actively involve others and understand that mutual respect is what sustains extraordinary efforts; they strive to create an atmosphere of trust and human dignity. They strengthen others, making each person feel capable and powerful [1].

1.5. Encourage the Heart

Leaders make people feel like winners. They keep hope and determination alive by appreciating contributions that individuals make. They recognize that everyone’s contributions are valued, creating a sense of community by celebrating the team’s victories. They establish high expectations and standards, holding people accountable to them by ensuring that rewards and performance are linked [1].

The LPI was created by developing a set of statements describing essential leadership actions and behaviors derived by recording specific one-sentence descriptions of behavior demonstrated in the personal-best leadership cases consistent with The Five Practices. Statements were selected, modified, or discarded following lengthy discussions and iterative feedback sessions with respondents and subject matter experts, as well as through empirical analyses of the behaviorally-based statements [3–5].

Each statement is evaluated on a ten-point Likert-scale. A higher value represents more frequent use of a leadership behavior. The anchors for the scale include: (1) Almost never do what is described in the statement; (2) Rarely; (3) Seldom; (4) Once in a while; (5) Occasionally; (6) Sometimes; (7) Fairly Often; (8) Usually; (9) Very Frequently; and (10) Almost always do what is described in the statement.

The LPI contains 30 statements; six essential behaviors associated with each of The Five Practices of Exemplary Leadership. Both a Self and Observer form of the LPI have been developed. Participating individuals complete the LPI-Self and request five-to-ten people familiar with their behavior to complete the LPI-Observer. The LPI-Observer is voluntary and respondents indicate their relationship to the leader; that is, this individual is my manager, co-worker or peer, direct report, or other. Except for the manager, identification of the respondent is anonymous. The instruments are returned directly to the researchers or seminar facilitator. The LPI (Self and Observer forms) takes approximately eight to ten minutes to complete, and is typically computer-scored but can also be hand-scored. The vast majority of respondents have completed the LPI in English. The online version of the LPI is also available in Spanish, Simplified Chinese, Brazilian Portuguese, German, and Mongolian.

The Five Practices framework is consistent with transformational leadership models [6,7]. Bass and Riggio [8] note that the LPI has been much more widely used by practitioners than it has been applied in published empirical research. However, the Multifactor Leadership Questionnaire (MLQ-X5), the instrument most widely used by researchers, and the LPI appear to be highly correlated. For example, in a study conducted with nurses in China, Chen and Baron [9] found that “The Chinese LPI total and subscales were highly and significantly correlated with the transformational leadership subscale of the Chinese MLQ-X5.” Similarly, Carless, et al. [10] in their study with retail bank personnel found the MLQ (X5) and LPI highly correlated, producing consistent outcomes. Kouzes and Posner [1,3] note that the original purpose of the LPI was for leadership development, helping individuals become more effective leaders through self-assessment and feedback from constituents, rather than for the empirical or scholarly study of leaders.
2. Sample Characteristics

The database for the analyses in this paper is from the LPI normative database, unless otherwise noted. That database includes nearly 2.8 million responses to the LPI online from 2007 through 2015. Seventeen percent of these responses are from leaders (N = 475,891), and the remainder are from “observers” of that leader. These include 333,684 managers, 742,446 direct reports, 906,674 coworkers or peers, and 339,960 others (not one of the previous three categories). Not all respondents provided demographic information, and so the sample sizes vary somewhat on the analyses and tables that follow.

A total of 130,515 men (56.4%) and 100,830 women indicated completing the LPI-Self. Less than one percent were below 24 years of age (N = 2195), with 16% between 24 and 32 years old (N = 37,047), 27.2% between 33 and 40 years old (N = 63,116), 31.2% between 41 and 49 years old (N = 72,284), 20.9% between 50 and 59 years old (N = 48,364), and 3.8% 60 years or older (N = 8887). Most respondents had completed college (N = 87,906; 42.4%), with 32.2% holding a master’s degree (N = 66,891), and 6.4% with a doctoral degree (N = 13,380). High school graduates were 4.2% of this sample (N = 8719), and the remainder indicated completing some college (N = 30,554; 14.7%). The breakdown by hierarchical level was 40,006 executives (19.3%), 78,437 middle managers (37.7%), 41,318 supervisors (19.9%), and 47,981 individual contributors (23.1%). The top five functional areas were management (N = 24,491; 15.8%), finance/treasury (N = 14,446; 9.3%), IT/MIS (N = 13,608; 8.8%), and engineering (N = 13,313; 8.6%). The top five industries they represented were government/military service (N = 32,463; 15.7%), banking/financial services (N = 20,354; 9.9%), education (N = 17,997; 8.7%), healthcare (N = 17,379; 8.4%), and logistics/supply chain (N = 14,326; 6.9%). The majority indicated their ethnicity (asked only of US respondents) as Caucasian (N = 151,886; 72.3%), followed by Asian/Pacific Islander (N = 20,191; 9.6%), Black/African-American (N = 15,001; 7.1%), Hispanic/Latino (N = 10,694; 5.1%), Mixed (N = 6664; 3.2%), and Native American (N = 1160; 0.6%).

Completing the LPI-Observer form were 774,114 men (54.2%) and 654,991 women. Less than two percent were below 24 years of age (N = 22,399), with 16.3% between 24 and 32 years old (N = 331,118), 28.6% between 33 and 49 years old (N = 409,588), 23.9% between 50 and 59 years old (N = 342,069), and 6.4% 60 years or older (N = 92,099). Most respondents had completed college (N = 475,766; 44.7%), with 27% holding a master’s degree (N = 286,957), and 5.1% with a doctoral degree (N = 54,086). High school graduates were 5.9% of this sample (N = 62,577), and the remainder indicated completing some college (N = 184,984; 17.4%). The breakdown by hierarchical level was 270,563 executives (25.4%), 283,269 middle managers (26.6%), 174,294 supervisors (16.4%), and 337,231 individual contributors (31.7%). The top five functional areas were management (N = 172,623; 22.3%), operations (N = 121,147; 15.7%), construction/real estate (N = 85,179; 11%), finance/treasury (N = 70,468; 9.1%), and engineering (N = 67,525; 8.7%). The top five industries they represented were government/military service (N = 151,695; 14.4%), banking/financial services (N = 106,772; 10.1%), logistics/supply chain (N = 92,372; 8.7%), aerospace/airlines (N = 84,133; 7.9%), and healthcare (N = 70,812; 7.4%). The majority indicated their ethnicity (asked only of US respondents) as Caucasian (N = 970,696; 72.8%), followed by Asian/Pacific Islander (N = 120,204; 9.0%), Black/African-American (N = 79,093; 5.9%), Hispanic/Latino (N = 66,751; 5.0%), Mixed (N = 40,036; 3.0%), and Native American (N = 7070; 0.5%).

3. Psychometric Properties of the LPI

The next section examines internal reliability assessments of the LPI from the normative database and findings from other published studies. The second section illuminates the impact of various demographic and contextual factors on LPI responses. Validity is the subject of the third section, particularly the influence on employee engagement levels as a function of leadership behavior and characteristics of individuals and context.
3.1. Internal Reliability

Reliability from a research perspective is about consistency or “repeatability.” This means that the instrument (assessment, survey, questionnaire, etc.) would give the same result over and over again; assuming that what was being measured is not changing. Remember that reliability is a characteristic of a measure taken across individuals and does not speak to the reliability (consistency) of an individual. Scores will seldom, if ever, be 100% reliable (the same) because of random errors (often referred to as “noise”) that cause scores to differ for reasons unrelated to the individual respondent. The fewer errors contained, the more reliable the instrument, and instrument reliabilities above 0.60 are considered good, and above 0.80 to be very strong [11]. The reliabilities for the LPI, as measured by Cronbach alpha coefficients, are consistently strong, and above this criterion, as shown in Table 1.

### Table 1. Internal Reliability (Cronbach alpha coefficients) for the LPI.

<table>
<thead>
<tr>
<th>Respondent Category</th>
<th>Leadership Practice</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model</td>
<td>Inspire</td>
<td>Challenge</td>
<td>Enable</td>
<td>Encourage</td>
</tr>
<tr>
<td>Leaders (Self)</td>
<td>0.810</td>
<td>0.901</td>
<td>0.843</td>
<td>0.825</td>
<td>0.898</td>
</tr>
<tr>
<td>Observers (All)</td>
<td>0.857</td>
<td>0.921</td>
<td>0.877</td>
<td>0.874</td>
<td>0.922</td>
</tr>
<tr>
<td>Managers</td>
<td>0.830</td>
<td>0.917</td>
<td>0.866</td>
<td>0.840</td>
<td>0.913</td>
</tr>
<tr>
<td>Direct Reports</td>
<td>0.873</td>
<td>0.924</td>
<td>0.876</td>
<td>0.869</td>
<td>0.917</td>
</tr>
<tr>
<td>Co-Workers</td>
<td>0.852</td>
<td>0.920</td>
<td>0.881</td>
<td>0.892</td>
<td>0.930</td>
</tr>
<tr>
<td>Others</td>
<td>0.856</td>
<td>0.919</td>
<td>0.878</td>
<td>0.871</td>
<td>0.920</td>
</tr>
</tbody>
</table>

Internal reliability for the LPI has also been found to be quite robust across a very wide range of sample populations. They represent a variety of occupations (fields and disciplines), positions and hierarchical levels, industries and organizations. For example, reliabilities ranged from:

- 0.80 to 0.92 for engineering managers and their constituents [12]
- 0.88 to 0.95 for U.S. Army government employees and support contractors [13]
- 0.78 to 0.86 for NASA scientists [14]
- 0.77 to 0.83 for construction site leaders [15]
- 0.88 to 0.94 for manufacturing personnel [16,17]
- 0.88 to 0.95 for project managers [18]
- 0.82 to 0.94 for hotel managers [19]
- 0.80 to 0.90 for frontline supervisors [20]
- 0.78 to 0.90 for mid-level managers [21]
- 0.90 to 0.96 for investment services directors [22]
- 0.66 to 0.91 for police chiefs [23]
- 0.71 to 0.86 for law enforcement personnel [24]
- 0.77 to 0.86 for sales personnel [25]
- 0.81 to 0.91 for women managers in non-profit organizations [26]
- 0.71 to 0.82 for women in executive positions in banking and higher education [27]
- 0.65 to 0.91 for college presidents [28–31]
- 0.70 to 0.89 for female vice presidents in nonacademic affairs [32]
- 0.70 to 0.91 for chief student affairs officers [33,34]
- 0.75 to 0.85 for chief financial officers at community colleges [35]
- 0.65 to 0.74 for department heads at King Saud University (Saudi Arabia) [36]
- 0.73 to 0.88 for community college faculty [37]
- 0.61 to 0.88 for Southern Baptist pastors [38,39]
- 0.61 to 0.80 for correctional institution leaders [40]
• 0.95 to 0.97 for faculty in a college nursing program [41]
• 0.70 to 0.88 for home health care agency directors [42]
• 0.85 to 0.92 for nursing home directors [43]
• 0.88 to 0.91 for residential treatment staff [44]
• 0.66 to 0.96 for nurses [45–48]
• 0.73 to 0.90 for healthcare managers [49,50]
• 0.89 to 0.92 for emancipated foster care youth [51]
• 0.78 to 0.95 for teachers [52–55]
• 0.73 to 0.85 for professional school counselors [56]
• 0.75+ for college counseling center directors [57]

Internal reliability scores have also been found to be quite strong in research studies outside the United States, involving non-U.S. populations. For example, Cronbach alphas ranged from 0.70 to 0.93 with a sample of bank personnel in Australia [58,59], from 0.94 to 0.97 for lending officers from Thai commercial banks [60], and from 0.64 to 74 for university department heads in Saudi Arabia [36]. They ranged between 0.82 and 0.93 for therapeutic radiographers from Hong Kong [61], from 0.88 to 0.93 for Lebanese hotel managers [62], from 0.88 to 0.95 for Lebanese managers of both Christian and Islamic faiths [63], from 0.77 to 0.92 for school heads in the Philippines [64], from 0.73 to 0.95 for nurses in Thailand [65], and from 0.74 to 0.88 for head nurses and their staff in Ugandan hospitals [66].

Tsend [67], having translated the LPI into Mongolian for use in a study of the leadership practices of higher education leaders, reflected the sentiments of many, in concluding: “The response options on the LPI are fairly straightforward...and there are no statements that directly reflect American cultural values that could potentially confuse respondents from other nations.”

This opinion is supported by other researchers who have used non-English language versions of the LPI. For instance, internal reliability coefficients for a Spanish-language version of the LPI, with Mexican respondents, ranged from 0.81 to 0.89 [68]. Strong internal reliability levels were reported for a Chinese-language version of the LPI [69]. Reliabilities found with an Arabic-language version, used with a sample of Jordanian school teachers, ranged from 0.77 to 0.80 [70]. Table 2 shows that the language which respondents used to complete the LPI online in the normative database makes negligible difference in the internal reliability coefficients of the five leadership practices.

<table>
<thead>
<tr>
<th>Language</th>
<th>Leadership Practice</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model Inspire</td>
<td>Challenge Enable Encourage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>English</td>
<td>0.850</td>
<td>0.919</td>
<td>0.872</td>
<td>0.867</td>
<td>0.918</td>
</tr>
<tr>
<td>Arabic</td>
<td>0.845</td>
<td>0.901</td>
<td>0.871</td>
<td>0.862</td>
<td>0.894</td>
</tr>
<tr>
<td>Brazilian Portuguese</td>
<td>0.838</td>
<td>0.875</td>
<td>0.861</td>
<td>0.861</td>
<td>0.883</td>
</tr>
<tr>
<td>Mongolian</td>
<td>0.903</td>
<td>0.925</td>
<td>0.879</td>
<td>0.899</td>
<td>0.924</td>
</tr>
<tr>
<td>Simplified Chinese</td>
<td>0.892</td>
<td>0.941</td>
<td>0.902</td>
<td>0.910</td>
<td>0.926</td>
</tr>
<tr>
<td>Spanish</td>
<td>0.863</td>
<td>0.912</td>
<td>0.859</td>
<td>0.865</td>
<td>0.924</td>
</tr>
</tbody>
</table>

3.2. Individual and Contextual Factors on the LPI

Another consideration for examining the reliability of any instrument is to examine how it might vary as a result of individual and/or contextual differences. The lack of systematic variations would indicate that the instrument is quite robust and would be safe (reliable) to apply across various sample population dimensions. In the normative database (results not shown), LPI scores were generally independent of various demographic characteristics (e.g., age, marital status, years of experience, educational level) and contextual factors (e.g., company size, functional area, length of service, line versus staff position).
Studies across a wide variety of populations and settings have demonstrated that responses on the LPI are not systematically related to demographic and contextual variables. These have included school administrators [71–81] and teachers [82–84], healthcare personnel [45,66,85–89], corporate managers in pharmaceuticals [90], banking [25], and hospitality [19,62], clergy [91,92], higher education administrators [25,29,93–98], military officers [99,100], law enforcement [101], government and public service personnel [102–106].

Furthermore, leadership practices as measured by the LPI did not systematically vary across different geographical settings (i.e., rural, urban, and suburban school districts) for school principals [107,108] or community college academic guidance counselors [109], or female community college presidents [110]. Neither city population nor number of sworn officers influenced the level of leadership practices for police chiefs [22]. Multiple regression analyses revealed that age, educational level, or work experience had no significant influence on the leadership practices of either male or female Thai managers [111]. No differences were found when comparing the leadership behaviors of those with or without physical disabilities [112]. Such factors as age, ethnicity, age, gender, gender role orientation, work experience, or year in school were not found to systematically influence LPI scores for college students [113–115].

There will be some exceptions, obviously. For instance, with the most experienced principals, their teachers perceived them as engaging more frequently in each of the five leadership practices than did teachers with principals who had less experience [116]. Dietitians who indicated a specialty, had more years of experience, and participated in professional organizations engaged more frequently in the five leadership practices than did their counterparts [117]. Experience, measured by age, years as a registered nurse, years with current unit, and years with current organization were all negatively correlated with the five leadership practices [118].

3.2.1. Self and Observer Comparisons

Means and standard deviations for each LPI scale for leaders (self) and their constituents (i.e., all observers, managers, direct reports, coworker or peers, and others) are presented in Table 3. Based upon mean scores, leaders used the practice of Enable most frequently. This is followed by Model, then Encourage, Challenge, and Inspire. This same pattern is true for all Observers, as well as from the perspective of Managers, Co-Workers, and Others.

<table>
<thead>
<tr>
<th>Respondent Category</th>
<th>Leadership Practice</th>
<th>Model</th>
<th>Inspire</th>
<th>Challenge</th>
<th>Enable</th>
<th>Encourage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leaders (Self)</td>
<td></td>
<td>46.14</td>
<td>42.72</td>
<td>44.00</td>
<td>49.50</td>
<td>45.25</td>
</tr>
<tr>
<td>(N = 446,780)</td>
<td></td>
<td>(8.11)</td>
<td>(10.29)</td>
<td>(8.89)</td>
<td>(7.12)</td>
<td>(9.69)</td>
</tr>
<tr>
<td>Observers (All)</td>
<td></td>
<td>47.12</td>
<td>44.21</td>
<td>45.17</td>
<td>49.57</td>
<td>46.31</td>
</tr>
<tr>
<td>(N = 2,169,866)</td>
<td></td>
<td>(8.93)</td>
<td>(10.84)</td>
<td>(9.63)</td>
<td>(8.32)</td>
<td>(10.43)</td>
</tr>
<tr>
<td>Managers</td>
<td></td>
<td>46.57</td>
<td>42.19</td>
<td>44.38</td>
<td>48.78</td>
<td>45.64</td>
</tr>
<tr>
<td>(N = 309,732)</td>
<td></td>
<td>(7.96)</td>
<td>(10.26)</td>
<td>(8.81)</td>
<td>(7.16)</td>
<td>(9.29)</td>
</tr>
<tr>
<td>Direct Reports</td>
<td></td>
<td>47.26</td>
<td>45.20</td>
<td>45.39</td>
<td>50.33</td>
<td>46.62</td>
</tr>
<tr>
<td>(N = 703,754)</td>
<td></td>
<td>(9.57)</td>
<td>(11.13)</td>
<td>(10.17)</td>
<td>(8.91)</td>
<td>(11.25)</td>
</tr>
<tr>
<td>Co-Workers</td>
<td></td>
<td>46.79</td>
<td>43.60</td>
<td>44.88</td>
<td>49.00</td>
<td>45.85</td>
</tr>
<tr>
<td>(N = 840,290)</td>
<td></td>
<td>(8.73)</td>
<td>(10.73)</td>
<td>(9.48)</td>
<td>(8.23)</td>
<td>(10.20)</td>
</tr>
<tr>
<td>Others</td>
<td></td>
<td>48.25</td>
<td>45.65</td>
<td>46.23</td>
<td>50.15</td>
<td>47.47</td>
</tr>
<tr>
<td>(N = 316,113)</td>
<td></td>
<td>(8.69)</td>
<td>(10.54)</td>
<td>(9.46)</td>
<td>(8.89)</td>
<td>(10.07)</td>
</tr>
</tbody>
</table>

As the sample size increases, the chance of finding statistically significant differences between groups also increases, even if these differences are not, for any one individual respondent, particularly meaningful, practical, or significant. This is true for the LPI, where statistically significant differences (p < 0.001) are found between the average scores of leaders and their observers (all, managers, direct reports, co-workers, and others) on all five leadership practices, but the effect size is rather small.
Given the very large sample sizes, effect size is an important concept because it represents a quantitative measure of the strength of a relationship. A small effect size means that the correlation coefficient between the independent variable (respondent category) and the dependent variables (five leadership practices) is below 0.10; or that the dependent variables are not correlated with the independent variable. From a practical perspective, a small effect size in this case indicates that responses about how frequently people are reported to engage in each of the five leadership practices does not vary very much as a result of the independent variable, in this case whether the respondent is a leader or an observer. A medium effect size is when the correlation is at 0.3 and a correlation at 0.5 would be considered a large effect size [119].

As previously shown in Table 3, the average scores from all observers are higher than those provided by leaders; and the variation in scores (standard deviation), with the exception of managers, is also greater among observers than among self scores. In any specific study, however, comparisons between self and observer reports are typically subjected to empirical scrutiny.

The relationship between self and observer scores on the LPI are mixed. In some studies, self scores are reported to be higher than observer scores, and in other studies, the opposite is found [36,64,71,77,120–136]. Still, in many other instances, no significant differences between self and observer responses have been reported [31,61,96,137–153]. Using a sample of leaders in rehabilitation services, no statistically significant correlations were found between the frequency of leadership behaviors reported by the leaders and their constituents, which led the researcher [154] to conclude that “the leader’s view of their own leadership does not affect perceptions by observers in this study either negatively or positively. In other words, it appears how the leaders of the current study score or view them is not related to how the observers of this study view these leaders on the five practices.”

In an empirical study involving over 1500 respondents, representing leaders, their managers, coworkers, and direct reports, relative “invariance” was found across the raters, which means that the key behaviors of leaders are conceptualized similarly at different levels of the organization [155]. As this researcher noted, “raters from difference sources do not necessarily view leadership performance differently. These findings should be good news to consultants, educators, and others that frequently use ratings from different organizational constituents in the service of leadership development.”

### 3.2.2. Gender (Male and Female) Comparisons

Statistically, the average scores of female leaders are significantly ($p < 0.001$) higher than those from their male counterparts in all five leadership practices, as shown in Table 4. This is also true from the perspectives of observers. The effect size in both cases is rather small ($r < 0.1$), as was also the case for comparisons between self and observer responses. The rank order of the five leadership practices is the same across all four cases.

#### Table 4. LPI Scores by Gender (Means and Standard Deviations).

<table>
<thead>
<tr>
<th>Gender</th>
<th>Leadership Practice</th>
<th>Model</th>
<th>Inspire</th>
<th>Challenge</th>
<th>Enable</th>
<th>Encourage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Males (Self)</td>
<td>(N = 100,824)</td>
<td>45.68</td>
<td>42.29</td>
<td>43.65</td>
<td>49.44</td>
<td>44.46</td>
</tr>
<tr>
<td>Females (Self)</td>
<td>(N = 130,504)</td>
<td>46.36</td>
<td>42.62</td>
<td>44.03</td>
<td>50.04</td>
<td>46.07</td>
</tr>
<tr>
<td>Males (Observers)</td>
<td>(N = 774,072)</td>
<td>46.98</td>
<td>43.80</td>
<td>44.85</td>
<td>49.40</td>
<td>46.15</td>
</tr>
<tr>
<td>Females (Observers)</td>
<td>(N = 654,947)</td>
<td>48.03</td>
<td>45.63</td>
<td>46.26</td>
<td>50.38</td>
<td>47.33</td>
</tr>
</tbody>
</table>

The vast majority of research has found no statistically significant gender differences on the LPI. This has been true across a variety of professional settings, positions, and occupations. For example, in healthcare [66,85,87,156–158], government [13,24], public sector [107,159–161], school teachers and
principals [81, 84, 162–168], higher education administrators [26, 29, 36, 169–172], managers [68, 173–177], scientists [14, 178], black belt martial artists [179], seminary students [180, 181], college students [116, 182], gay men and lesbian women [183], emancipated foster care youth [48], and sales personnel [23].

3.2.3. Government/Military (Public) and Business (Private) Comparisons

Scores in the LPI for respondents working in the public sector (local, state, and federal government, armed forces, and social services) were compared with private sector respondents (employed in aerospace, automobiles, banking/financial services, manufacturing, pharmaceuticals, and telecommunications), as shown in Table 5. For leaders, statistically higher frequency scores were reported in the public sector in Model, Enable, and Encourage than their private sector counterparts. Private sector leaders reported statistically higher frequency scores than public sector leaders in Inspire and Challenge. However, analyzing responses from Observers revealed statistically higher mean scores reported for public sector leaders than for private sector leaders in all five leadership practices. Effect sizes were small in all cases.

Table 5. LPI Scores by Public and Private Sector Employment (Means and Standard Deviations).

<table>
<thead>
<tr>
<th>Gender</th>
<th>Leadership Practice</th>
<th>Model</th>
<th>Inspire</th>
<th>Challenge</th>
<th>Enable</th>
<th>Encourage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public (Self) (N = 38,225)</td>
<td>45.86</td>
<td>41.99</td>
<td>43.38</td>
<td>49.85</td>
<td>45.24</td>
<td></td>
</tr>
<tr>
<td>Private (Self) (N = 54,131)</td>
<td>(7.90)</td>
<td>(10.31)</td>
<td>(8.89)</td>
<td>(6.61)</td>
<td>(9.54)</td>
<td></td>
</tr>
<tr>
<td>Public (Observers) (N = 178,757)</td>
<td>47.83</td>
<td>44.64</td>
<td>45.44</td>
<td>50.27</td>
<td>47.04</td>
<td></td>
</tr>
<tr>
<td>Private (Observers) (N = 324,181)</td>
<td>(9.11)</td>
<td>(11.14)</td>
<td>(10.07)</td>
<td>(8.46)</td>
<td>(10.63)</td>
<td></td>
</tr>
</tbody>
</table>

Research comparing healthcare leaders employed in public or private sector settings found no differences between the two groups [157], as was also true comparing the top staff of human service organizations (non-profit) with a random selection of business managers [184]. No statistically significant results emerged from a study investigating the organizational performance of non-profit organizations between 2006 and 2008 and 2009 and 2010 with the five leadership practices [185]. Studying elementary school principals, Goewey [186] asserted that The Five Practices framework effectively filled the gap in “aligning effective school principal practices with a validated research-based business framework.” Nicewarner [187], in comparing leadership practices in the public, private and nonprofit sectors, concluded: “There appear to be no significant differences in leadership traits as practiced across the three organizational types . . . and I have come to realize that the practice of leadership involves changes in context, not the actual practice.”

3.2.4. Ethnic Background Comparisons

Information about ethnicity in the normative database is only collected for respondents within the United States. Table 6 compares the responses of Caucasian leaders with those of Leaders of Color (combining Asian/Pacific Islanders, Hispanics/Latinos, African-Americans, Native American, and Mixed race categories), revealing statistically significant differences, with the latter reporting statistically more frequent use of the five leadership practices than the former (p < 0.001), although the effect size was small. The results were the same from the perspective of observers (results not shown).

Comparisons between the five groups comprising “persons of color” revealed statistically significant differences (p < 0.001) between the ethnic categories. Most of the differences were between Asian/Pacific Islanders and those of Mixed race with their Black/African American, Hispanic/Latino, and Native American counterparts. A similar pattern was found for observers (results not shown).
Table 6. LPI Scores by Ethnicity (Self only; Means and Standard Deviations).

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>Model</th>
<th>Inspire</th>
<th>Challenge</th>
<th>Enable</th>
<th>Encourage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Caucasians (N = 128,483)</td>
<td>46.02</td>
<td>42.40</td>
<td>43.89</td>
<td>49.74</td>
<td>44.98</td>
</tr>
<tr>
<td>Persons of Color (N = 38,124)</td>
<td>47.20</td>
<td>43.88</td>
<td>44.73</td>
<td>50.51</td>
<td>46.79</td>
</tr>
<tr>
<td>Asian/Pacific</td>
<td>45.95</td>
<td>42.10</td>
<td>44.02</td>
<td>49.75</td>
<td>45.54</td>
</tr>
<tr>
<td>Islander (N = 9593)</td>
<td>47.23</td>
<td>43.88</td>
<td>44.75</td>
<td>50.58</td>
<td>46.78</td>
</tr>
<tr>
<td>African American (N = 13,805)</td>
<td>48.46</td>
<td>45.41</td>
<td>45.39</td>
<td>51.35</td>
<td>48.17</td>
</tr>
<tr>
<td>Hispanic/Latino (N = 9245)</td>
<td>47.23</td>
<td>43.88</td>
<td>44.75</td>
<td>50.58</td>
<td>46.78</td>
</tr>
<tr>
<td>Native American (N = 940)</td>
<td>47.09</td>
<td>44.13</td>
<td>44.69</td>
<td>49.98</td>
<td>46.32</td>
</tr>
<tr>
<td>Mixed Race (N = 4451)</td>
<td>45.96</td>
<td>42.99</td>
<td>44.20</td>
<td>49.52</td>
<td>45.35</td>
</tr>
</tbody>
</table>

Other researchers have investigated the impact of ethnicity on leadership behavior. For example, in a study involving executive directors of community development organizations [188], LPI scores for Caucasians directors were compared with those of Directors of Color (Black, Hispanic, and Asian). The two groups did not differ in Challenge, Enable, or Encourage. Directors of Color reported significantly higher Model and Inspire scores than their Caucasian counterparts. However, assessments provided by their constituents revealed no systematic differences between the leadership practices of managers based upon their ethnic background. (Re-examination of the data by respondent gender also made no difference in the pattern of results).

African-American female community college presidents were found to have similar frequencies in their use of Model, Inspire, Challenge, and Enable compared with their Caucasian counterparts [110]. Faculty at a historically Black university showed significant differences by ethnicity (African-American, Asian, and Caucasian) in their perceptions of their department chairs’ predominant leadership practice [189]. Two separate studies [31,190], comparing African-American and Caucasian female leaders, revealed no significant main or interaction effects on leadership practices by ethnicity. Another study [191] reported that women of color in executive positions within higher education exhibited all five leadership practices at a higher level than that reported in the LPI normative database.

Hispanic school principals did not rate themselves significantly differently in their use of the five leadership practices than non-Hispanic principals [139]. Another study [48], involving emancipated foster care youth, found no differences in how frequently they reported their case workers using the five leadership practices on the basis of ethnicity. LPI scores were also not found to be statistically different in an investigation of Native American and non-Native American secondary school administrators [192]. LPI scores did not vary between Caucasian and non-Caucasian seminary students [181]. In addition, no differences in The Five Practices were found among academic deans based upon their ethnicity [171].

To date, studies of leadership practices by ethnic background have generally been restricted to U.S. sample populations, and comparisons by ethnicity show mixed results. The findings have been mostly descriptive rather than prescriptive across the sample populations.

3.2.5. Cultural (Nationality) Comparisons

Seventy-four percent of the respondents in the normative database are from the United States. Comparisons between U.S. respondents and those from the rest of the world (labeled “international”) in Table 7 show that U.S. leaders report statistically significant ($p < 0.001$) greater use of all five leadership practices than do international leaders, although effect size was small. The results are similar to the perspective of observers (results not shown).
Table 7. LPI Scores from United States and International Respondents (Self only; Means and Standard Deviations).

<table>
<thead>
<tr>
<th>Nationality</th>
<th>Leadership Practice</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model</td>
</tr>
<tr>
<td>United States</td>
<td>46.24</td>
</tr>
<tr>
<td>(N = 174,526)</td>
<td>(7.72)</td>
</tr>
<tr>
<td>International</td>
<td>45.14</td>
</tr>
<tr>
<td>(N = 62,396)</td>
<td>(7.40)</td>
</tr>
</tbody>
</table>

The impact of culture or nationality on leadership practices has been investigated by other scholars. For example: U.S. and European managers, U.S. and Pacific Rim managers, U.S. and Australian managers, U.S. and Mexican managers, and U.S. and Israeli managers. Few differences were found between U.S. and United Kingdom managers working for the same multinational chemical company. Enable was rated most frequently by managers as well as their constituents from both countries. The same consistent pattern was observed for Inspire and Challenge. Within one large high technology firm, no significant differences were found between U.S. managers and their counterparts in England, the Netherlands, or Germany. This was true for both LPI-Self and LPI-Observer scores. A study involving American and Swiss managers found no differences in the leadership practices of Model and Enable; and American managers reported more frequent use of Inspire, Challenge, and Encourage than their Swiss counterparts [175].

Managers from small-sized factories in four Pacific Rim countries (Korea, Philippines, Taiwan, and Malaysia) completed the LPI as part of a multinational semiconductor company’s management development program. LPI-Self scores were significantly higher than those reported by their constituents for all leadership practices, with the exception of Encourage. The rank order for the LPI-Self scores were the same for the Pacific Rim managers as had been found for their U.S. counterparts. This same pattern was true for LPI-Observer scores. Staff nurses in the northern region of Thailand reported significantly more job satisfaction and organizational commitment to the extent that their head nurses engaged in each of the five leadership practices [193]. Comparisons of the LPI scores from MBA students in the U.S., Nigeria, and Slovenia were very similar [194].

Mid-level Australian managers were matched with comparable U.S. managers, and no statistically significant differences between the two groups were found for any of the five leadership practices. While the LPI scores of Mexican managers were, on average, lower than their U.S. counterparts, there were no differences between the two groups in the rank (relative) order of the leadership practices [195, 196]. Studies involving U.S. and Israeli managers have found that nationality was not significantly related with any leadership practice, nor did nationality mediate the relationship between leadership practices and employee commitment levels [197]. The average score for all five leadership practices increased over time as a result of a leadership development program involving nurse leaders in a Swiss hospital, who completed the LPI in German [198]. The Five Practices framework was also found to be quite descriptive of the leadership practices of head nurses from five hospitals in Finland [199]. Researchers comparing the impact of managers’ leadership practices on staff nurses in the U.S. and China reported that the LPI was “easily used across boundaries” [69] and the Chinese-language version of the LPI was labeled “a cultural appropriate instrument” [9]. As Cardador [155] concluded: “transformational leadership may be recognized by individuals at all levels of the organization because of its universality—the fact that regardless of an individual’s station or culture, conceptualizations of transformational leadership will be similar.”

3.3. Validation of the LPI

The previous section examined various individual and contextual factors that potentially differentiate LPI scores between respondents. However, because of the large sample sizes involved, the criterion of statistical significance was beyond chance levels (p < 0.001), and the practical significance
of those differences were typically negligible, as evident by small effect sizes. The critical issue is whether or not such differences have any systematic bearing on the effectiveness of leaders or how their leadership behaviors impact employee engagement levels and productivity.

An instrument (such as the LPI) has validity when, with some degree of reasonable certainty, it truly measures what it purports to measure and, accordingly, its scores have meaning or utility for a respondent. Like reliability, validity is determined in a number of ways. The most common assessment of validity is called face validity, which considers whether, on the basis of subjective evaluation, an instrument appears to measure what it intends to be measuring. Given that the items on the LPI are related to the qualitative findings from interviews with leaders and echo the comments that workshop and seminar participants generally make about their own or others’ personal-best leadership experiences, respondents have found the LPI to have excellent face validity.

Validity is also determined empirically (objectively). Factor analysis is used to determine the extent to which the instrument items measure common or different content areas. The results from various analyses reveal that the LPI contains five factors; the items within each factor correspond more among themselves than they do with the other factors [12, 21, 26, 200–202]. Herold and his colleagues [12], using data primarily from corporate respondents, performed a confirmatory factor analysis using LISREL VII, analyzing a covariance matrix prepared from the raw data by PRELIS. Their conclusion:

Estimating a correlated factors model corresponding to the oblique factor rotation, modified to reflect the inter-correlations among the error items for the LPI items that had correlations with other items exceeding 0.50, resulted in a confirmatory model with acceptable fit (Chi-Square = 399.9, d.f. = 363, p < 0.09). In addition, all of the hypothesized structural coefficients linking the observed variables to the five factors were highly significant with all t values exceeding 7.0, suggesting that when modeled appropriately, the LISREL estimates confirm the LPI factor Model.

Using responses from front-line and middle managers from a community college, Adcock-Shantz [203] ran a confirmatory factor analysis on the LPI, “which yielded five interpretable factors, consistent with Kouzes and Posner’s five factors ... The a priori hypothesis had five dimensions and the scree plot confirmed the five dimensions/factors were correct. The five factors were rotated using Varimax rotation. The rotated solution yielded the following five factors: Model the Way, Inspire a Shared Vision, Challenge the Process, Enable Others to Act, and Encourage the Heart, which accounted for 90% of the item variance.”

Applying a similar methodological approach (LISREL), with a sample of U.S. and Canadian community activists, the analysis confirmed the structural integrity of the LPI framework [112]. A structured interview protocol within a school setting reported that participants validated 81% of the salient principal scores as actual patterns of behavior in their experience [74]. The five factor structure was essentially replicated in a study involving school administrators and teachers [200], and the framework was explained over 72% of the variance in a study of chief faculty officers in Thailand [204].

In still another cross-cultural study [205], the LPI was administered to MBA students (mostly line and middle managers) in six countries from five continents: U.S., India, Nigeria, South Korea, Argentina and Slovenia. The first three were administered in English and the latter three in their native languages. The results showed a high degree of “structural equivalence” indicating that the LPI was measuring the same construct across different culture settings. As the authors summarized,

Results of the multi group confirmatory factor analysis (CFA) performed on the sample showed that the five factor structure emerged in all nations studied and that most of the items that were supposed to load on a particular factor did load on that factor. Out of 180 loadings (6 groups × 30 items), 175 loadings were significantly different from 0.
Furthermore, most of the factor loadings (for 21 out of the 30 items) were equal (showing no statistically significant differences) across cultures studied.

A qualitative study [206] involving pastoral staff, where the researcher anticipated that the data from his case studies might not neatly fit in The Five Practices model proposed by Kouzes and Posner [1], revealed that this “was not the case. The five major categories were clear yet broad enough to serve in the organizing and interpreting of the data.” This conclusion was similar to that reached by Walker and Grey [207] who found that “the practices and behaviors previously identified for persons in business organizations [as represented on the LPI] were mirrored by leaders working in community groups.”

The question of whether the LPI scores are significantly related to other critical behavioral (individual and organizational) performance measures is probably the most important practical matter to participants (leaders and their organizations). This is often referred to as predictive validity. Researchers have shown that leadership scores are consistently associated with important aspects of managerial and organizational effectiveness, such as workgroup performance, team cohesiveness, commitment, satisfaction, and credibility.

For example, the relationship between a six-item Likert scale assessing various aspects of a leader’s effectiveness (Cronbach alpha = 0.98) and their leadership practices (as measured by the LPI) was examined utilizing only the responses from observers. Including only the responses from “other people” about their managers provided relatively independent assessments and thereby minimized potential self-report bias. Regression analysis was performed, with leader effectiveness as the dependent variable and the five leadership practices as the independent variables (F = 318.88, p < 0.0001). Leadership practices explained over 55% (adjusted R^2 = 0.756) of the variance around constituents’ assessments of their leaders’ effectiveness. Organizational commitment, job satisfaction, and productivity were significantly correlated with each of the five leadership practices on the LPI in a study involving directors of an investment services firm and their constituents [22].

The current online version of the LPI provides respondents with an opportunity to voluntarily respond to ten statements about various attitudes and sentiments regarding their workplace. The statements are measured on a five-point Likert scale, with these anchors: “1” Strongly Disagree, “2” Disagree, “3” Neither Agree nor Disagree, “4” Agree, and “5” Strongly Agree. The statements are:

1. My work group has a strong sense of team spirit.
2. I am proud to tell others that I work for this organization.
3. I am committed to this organization’s success.
4. I would work harder and for longer hours if the job demanded it.
5. I am highly productive in my job.
6. I am clear about what is expected of me in my job.
7. I feel that my organization values my work.
8. I am effective in meeting the demands of my job.
9. Around my workplace, people seem to trust management.
10. I feel like I am making a difference in this organization.

Combining responses to these ten statements creates a “positive workplace attitude” scale or what many scholars and practitioners refer to as engagement. Using only direct reports, Cronbach’s alpha for this scale was 0.88. Internal reliability coefficients were quite strong regardless of the language respondents used to complete this scale. For example: English (0.88), Arabic (0.86), Brazilian Portuguese (0.89), Mongolian (0.86), Simplified Chinese (0.93), and Spanish (0.87). This allows for the testing of the hypothesis: Do the leadership behaviors of leaders explain the engagement levels of their direct reports?

The most direct way to study this question is to look at how their direct reports describe the leadership practices of their leaders, and to examine the extent to which these behaviors account for
their level of engagement. Multiple regression analysis, with engagement as the dependent variable, and leadership practices as the independent variables, found that The Five Practices of Exemplary Leadership model, as measured by the LPI, accounted for more than 38% of the variance in the levels of engagement of direct reports ($R = 0.619$, adjusted $R^2 = 38.4$, $F = 44221.21$, $p < 0.0001$).

Regression analysis was also used to examine whether or not variances in engagement levels of direct reports could be accounted for by possible individual differences across respondents. To test this hypothesis, the following nine variables were entered into the regression equation as independent variables, with engagement as the dependent variable: age, educational level, gender, functional area, hierarchical level, industry classification, length of time with company, organizational size (number of employees), and nationality (country of origin). The amount of explained variance in direct reports’ levels of engagement, with all nine variables entered into the equation, was 0.3%. Stepwise regression analysis revealed that length of time with the organization, nationality, and functional area each accounted for about 0.1% of the variance and the contribution of the remaining six variables was zero.

In summary, how frequently people report that their leaders engage in The Five Practices of Exemplary Leadership is directly related to their level of engagement. Moreover, this relationship is not affected by characteristics of the respondents. Demographic factors about direct reports do not illuminate why they are or are not engaged in the workplace, but knowing how they see their leader behaving provides a substantial explanation for their levels of engagement.

Another aspect of validity is determining how the LPI is related with the leader’s effectiveness as reported by their constituents. In the normative database, respondents are asked a global question about the overall effectiveness of the leader, using the same five-point Likert scale described previously. The correlations are all statistically significant ($p < 0.001$); above 0.55 for managers, above 0.57 for colleagues, and above 0.65 for direct reports. Considered together (managers, colleagues, and direct reports), the bivariate correlations between each of the five leadership practices and ratings of the leader’s effectiveness are all above 0.61 ($p < 0.001$).

Looking only at direct reports, which are possibly the most critical constituent group that leaders typically have to influence, those who “strongly agree” that their leader is effective were compared with direct reports which do not strongly agree about how frequently each group observed their leader using The Five Practices. Comparing the mean scores for each of the five leadership practices reveals that the leaders who are reported as most effective by their direct reports are viewed as engaging significantly more often ($p < 0.001$) in The Five Practices than those who view their leader as less effective.

4. Conclusions

The Leadership Practices Inventory has sound psychometric properties. Internal reliabilities for the five leadership practice scales (both the Self and Observer versions) are very good and are consistently strong across a variety of sample populations and situations. The underlying factor structure has been sustained across a variety of studies and settings. Scores from the LPI show construct and predictive validity. Findings are relatively consistent between constituent groups, genders, ethnicity, cultural backgrounds, and national boundaries, as well as across various contextual characteristics (e.g., positions and industry).

After completing a comprehensive review of the LPI, Fornito and Camp [208] echoed this conclusion: “The application of this measure is both practical and efficient: We believe the LPI is a strong measure based on its reliability and validity. Its psychometric properties compounded with its global traits suggest it is a measure we can utilize to compare groups across countries with an unbiased scale.” Lewis [209] in the Mental Measurements Yearbook concurred,

The LPI is one of the most extensively researched management development tools I have encountered. It is a model of sound research design from its initial development and refinement through subsequent concurrent validity studies. The instrument and
instructions are easy to read and follow and the trainer’s guide is logical and clear. I highly recommend it as a developmental tool for new and experienced managers.

The utility of the LPI for leadership development was further underscored by an evaluation of The Royal College of Nursing Institute’s (London) Clinical Leadership Program [210]. The research team reported that among clinical leaders:

- A total of 94% agree or strongly agreed that the LPI was “a useful tool for understanding my leadership development needs.”
- A total of 89% agreed or strongly agreed that the LPI “was useful for developing my professional development plan.”
- A total of 95% agreed or strongly agreed that “it was useful to have a measure of how others perceive my leadership capabilities.”
- A total of 85% agreed or strongly agreed that the LPI “was able to show changes in my leadership capability over time.”

Studies using the LPI normative database, along with scores of scholars utilizing their own research samples across numerous settings, document that the LPI is quite robust in assessing individuals’ leadership capabilities, and demonstrates that The Five Practices of Exemplary Leadership make a difference at the personal, interpersonal, small group, and organizational level. The LPI has proven quite capable of assessing individuals’ leadership behaviors and in providing feedback useful for developing and enhancing leadership capabilities.

Scholars should continue to use the LPI in their research, and further efforts can be made to assess the power of this instrument in comparison with other leadership assessment instruments. The psychometric properties of the LPI should not be taken for granted and future researchers should examine the ongoing reliability and validity of the instrument, along with its basic factor structure and dimensions. The LPI’s validity and utility will also be enhanced by expanding the range and variety of outcome measures associated with leadership practices and behaviors. Efforts by teachers and practitioners involved in leadership development efforts will be enhanced by using the LPI to establish baseline marks, and provide feedback about learning and improvements in leadership effectiveness and capability.

Conflicts of Interest: The author declares no conflict of interest.

References


37. Solis, F. A Profile of Faculty Leadership Behavior at One South Texas Community College. Ph.D. Thesis, Texas A & M University, College Station, TX, USA, May 2011.


133. Haridom, L.J. The Relationship of Principal Leadership to Organizational Learning and Sustained Academic Achievement. Ph.D. Thesis, Texas A & M University, College Station, TX, USA, May 2009.


© 2016 by the author; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (http://creativecommons.org/licenses/by/4.0/).