Traditional versus Consensus Based Assessment (CBA) of Leadership to Predict U.S. Military Academy-West Point Graduates’ Career Intent

Joseph Psotka
US Army Research Institute for the Behavioral and Social Sciences (ARI)
Paul Bartone
Industrial College of the Armed Forces/ National Defense University
ICAF/NDU
Peter Legree
Tim Sherwood, Lynn Milan and Debbie R Robinson
US Army Research Institute for the Behavioral and Social Sciences (ARI)

SUMMARY

The singular challenge of leadership research is to develop overarching theories of leadership that can translate into accurate and powerful assessment instruments. This paper examines some of the best existing instruments in the context of the most comprehensive longitudinal study of leader development extant: the Baseline Officer Leader Development Study (BOLDS), but with a cutting edge new psychometric technology: consensus based assessment (CBA) (Legree, Psotka, Tremble, and Bourne, 2005). CBA applied to the existing measures provides new insights into the Hardiness Scale (HS) and Multifactor Leadership Questionnaire (MLQ), and increases the MLQ’s predictive power. The two goals of this effort are positively demonstrated: the improvement of existing leadership assessment instruments using CBA creates better standards and more perspicacious psychometrics; and the improves the predictive validity in understanding officers’ career intentions. Although Hardiness provided the strongest correlate of career intention using a concurrent design, a regression of Hardiness, a new CBA-based measure, the Balanced Tacit Framework of Leadership (BTFL-MLQ), and Tacit Knowledge of Military Leadership (TKML) proved to be the strongest predictors of Career Intent using a predictive design.

1 Presented at the 2005 annual meeting of the IAMPS – International Applied Military Psychology Symposium, Washington, DC.

The views, opinions, and findings contained in this article are solely those of the authors and should not be construed as an official Department of the Army or DOD position, policy, or decision, unless so designated by other documentation.

2 Joseph Psotka, PhD
Us Army Research Institute For The Behavioral And Social Sciences
2511 Jefferson Davis Highway
Arlington Va 22202-3926
703-602-79452
Fax: 703-602-7737
INTRODUCTION

The Baseline Officer Leader Development Study (BOLDS) represents the most comprehensive longitudinal study of leader development ever undertaken at any U.S. military service academy. With the goal of understanding the factors that contribute to leader development and effective leader performance, cadets who entered West Point in 1994 and graduated in 1998 were studied over their four-year tenure at the academy. Earlier findings from the BOLDS research identified cognitive and personality factors that are associated with leader performance during the academy years (Bartone, Snook & Tremble, 2002; Bartone & Psotka, 2004). A follow-up survey completed in 2003 by 391 Cadets, who had then become Officers, included three self-report predictors of leadership: the Multi-factor Leadership Questionnaire (MLQ) (Bass & Avolio, 1995), the Cadet Performance Report (CPR) (U.S. Corps of Cadets, 1995), and the Hardiness Scale (HS) (Bartone, 1995) that were all included in the five-year earlier BOLDS database. Analyses using improved scoring standards based on Consensus Based Assessment (CBA) means have proven more interpretable than traditional scoring rubrics within the BOLDS database (Psotka, Legree, Sherwood, Bartone, & LaVoie, 2005), and so CBA technologies were used to examine the follow-on Officer’s survey for developmental progress.

Recent trends suggest that junior officers are departing military service at an increasing rate, and this trend portends substantial shortfalls in meeting future military leadership staffing requirements. Given this concern, follow-up analyses explored officer career intentions, and predictors of this self report measure.

Finally, as these officers have now completed their mandatory 5-year service obligation, it is our hope that this paper will provide a baseline for future work to examine actual retention and attrition behaviors and identify associated factors. The overall research program is intended to yield empirical findings on which to base future training and intervention programs aimed at maximizing military leader development, performance, and retention.

Framework of the Research Analysis

Consensus Based Assessment (CBA)

The framework used here to re-examine both the Cadet and follow-on Officers’ data hinges on an information processing perspective of leadership that focuses on knowledge, rather than traits or attributes. Within this framework, leadership knowledge is acquired incrementally from many sources. Leaders must draw upon not only their personal experience, but also the knowledge and

---

3 As was reported in a poster presented at the July 2004 convention of the American Psychological Association, Hawaii.
experience of others through study, synthesis, and reflection (Ruvolo, Petersen, & LeBoeuf, 2004). This knowledge, including procedural knowledge of skills and declarative knowledge of values and attitudes, represents and reflects an overall organizational culture in the Army (Gerstner & Day, 1994). Although much of this knowledge is explicit and codified in doctrinal training and field manuals, a significant portion is tacit (Hedlund, Williams, Horvath, Forsythe, Snook, Wattendorf, McNally, Sweeney, Bullis, Dennis, and Sternberg, 1998) and unwritten, but still shared through everyday practices, common situations, and overarching consensus of opinions and beliefs.

The fact that so much of this knowledge resides in a culture of opinions and beliefs creates a unique opportunity for assessing the enculturation, the knowledge acquisition levels, of officers by comparing their judgments against experts, more senior ranks, or simply a large representative sample of peers (Legree, Psotka, Tremble, and Bourne, 2005). Even though there is no absolute or objectively correct answer to a question such as “How important is it for a leader to re-examine critical assumptions about questions versus seeking differing perspectives when solving problems?” Army officers as a whole can be shown to have a highly reliable consensus on this question. Given Likert scale alternatives of these questions, groups of Army officers of all ranks, with reliability correlations very close to 1.0, will show very great agreement in their judgments. These group averages provide potential CBA standards for assessing the knowledge acquisition level of any individual. The correlation coefficient between an individual’s pattern of responses and the CBA mean of a relevant group provides a direct measure of his or her understanding of Army organizational culture. Since Army culture is reflected in all of its shared knowledge and meaning systems (Hanges, Lord, & Dickson, 2000) the assessment of cultural knowledge acquisition should have many topics and dimensions for scaling; and very broad generalization of its findings. Potentially, many of the existing scales that assess opinions of others or self assessments could tap broadly into cultural knowledge acquisition through this technology of CBA standards. Theoretically, these measures should provide an index of an officer’s leadership effectiveness.

Leadership effectiveness and level of cultural knowledge acquisition are intimately and causally related in many ways. Organizational culture is the important supporting framework for leader growth and development. Organizational culture influences how prototypical a leader is seen to be, on the basis of prevailing knowledge schemas and standards of categorization (Lord, Foti, & DeVader, 1984) and leader prototypicality has been shown to be a determinant of effectiveness in many ways, including self sacrifice, influence, and emergence (Hogg & van Knippenberg, 2003). Organizational culture provides the overarching standards and norms of behavior and is an essential component for building an optimal leadership development system (Avolio, 1999). Leader development initiatives, like most organizational change efforts, demand a clear understanding of organizational culture and the powerful manner in which it
shapes and maintains behavior in almost invisible ways (Kotter, 1996). Furthermore,

Survey of Officers

In September 2002, a survey was mailed to all West Point Class of 1998 graduates (N=871) for whom addresses could be located (out of 883 graduates, 12 were unknown). The timing was intentional. We considered it important to wait some years after graduation to allow these young officers to accumulate professional and leadership experience. At the time of the survey, most respondents were in the rank of Captain with about four years of post-commissioning experience. Of the 871 surveys mailed, 79 individuals were never located, leaving a final pool of 792 officers who we presume received the survey. Of these, N=391 responded with completed surveys, for a response rate of 49%. This is an excellent response rate for such a highly mobile population.

In addition to an informed consent statement, the survey included a standard measure of “transformational” leadership style (MLQ), a modified version of the West Point cadet performance report (CPR), and the short Hardiness (dispositional resilience) Scale (HS), which were all available in the existing BOLDS database. In addition there were career and job assignment history questions, a career intentions item, and a general health item.

Before we analyze these results, we will provide some BOLDS background information on a few measures of the Cadets meaningfully related to the survey results of a subgroup of these Cadets, now Officers.

Class of ’98 at USMA

The USMA class of 1998 was initially composed of 1143 individuals who arrived at West Point to participate in the summer 1994 activities. Of the 1143 students who were admitted in 1994, 883 (77%) graduated in 1998. Data collection for the BOLDS project began in the summer of 1994, using either normal administrative collections or else special paper-and-pencil measures. A complete listing of the measures and when they were administered can be found in Milan, Bourne, Zazanis, and Bartone (2002)

Tacit Knowledge of Military Leadership

Tacit knowledge refers to the practical, action-oriented knowledge that Army leaders typically acquire from experience. The Tacit Knowledge for Military Leaders – Platoon Leader Questionnaire (TKML-PLQ) is a survey designed to measure this knowledge. The instrument consists of 16 scenarios, each
describing a situation encountered by military leaders. Following each scenario were several options (7 to 15) for how to handle the situation. Using a scale of 1 (“extremely bad”) to 9 (“extremely good”), cadets rated the quality of each option for achieving the goal or solving the problem. Hedlund, Williams, Horvath, Forsythe, Snook, Wattendorf, McNally, Sweeney, Bullis, Dennis, and Sternberg (1998) reported a coefficient alpha of .68 for this 16-item instrument.

BOLDS researchers administered the 16-item TKML-PLQ to 327 cadets during their 7th term at USMA. Each cadet’s response pattern (i.e., ratings for all the options) was correlated with the mean ratings given by a CBA standard (Legree, Psotka, Tremble, and Bourne, 2002) created by “experts” (i.e., experienced Army majors). The resulting correlations between experts and cadets’ ratings are the TKML scores that measure how closely each cadet’s understanding resembles that of experienced officers. The mean correlation was .64 (see Table 1). The coefficient alpha for 320 Cadets with complete responses was 0.935 (Nunnally, 1967).

Table 1.

<table>
<thead>
<tr>
<th>Administraton Date</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Range of Correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Term 7 - Fall ’97</td>
<td>307</td>
<td>.64</td>
<td>.15</td>
<td>.05 – .82</td>
</tr>
</tbody>
</table>

** p < .01; (From Milan et al., 2002)

In the analyses below, the Cadets’ TKML-PLQ scores correlated significantly with the MLQ, HS, and career intent measures in both Cadets and Officers data, but this instrument was not given in the Survey.

**Hardiness**

As a personality dimension, hardiness refers to a “pervasive and steady sense of commitment, control, and challenge” (Bartone, 1999a) and has been shown to moderate the ill effects of stress (Bartone, 2000). Compared to individuals who are low on hardiness, hardy persons have a higher sense of life and work commitment, a greater feeling of control, and more openness to change and challenges in life (Bartone, 1999a).
During the 8th term (4th year), a short hardiness scale developed by Bartone (1995) was disseminated to a sample of cadets through e-mail. The 15-item scale included both positively and negatively keyed items associated with the three conceptual facets of hardiness: commitment, control, and challenge. Each of the 15 items was a general statement about life (e.g., “Changes in routine are interesting to me”). Cadets indicated on a four-point scale how true they felt each statement was (“not at all true” to “completely true”).

According to Bartone (1999b), Cronbach’s coefficient alpha for the total HS measure was .82 (.77 for commitment, .68 for control, and .69 for challenge) in a sample of 787 Army National Guard and reservists in medical units mobilized for the Gulf War. Among 105 USMA cadets, the three-week stability coefficient for the total HS was .78 (Bartone, 1999b).

The HS scale was completed and returned through e-mail by 430 cadets. As a group, their hardiness scores ranged from 11 to 42 (out of a possible range of 0 to 45), with a mean of 30.8 ($SD = 4.65$) and a median of 31.0 (see Table 2). The coefficient alpha for 430 Cadets with complete responses was 0.705 (Nunnally, 1967).

### Table 2.
**Cadets’ Hardiness Scores**

| Administrati
on Date | $N$ | Mn  | $SD$ | Range of Scores |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Term 8 - Spr '98</td>
<td>430</td>
<td>30.8</td>
<td>4.65</td>
<td>11 – 42</td>
</tr>
</tbody>
</table>

(From Milan et al., 2002)

**Cadets’ Leadership Styles (MLQ)**

Beginning in the second year, the Multifactor Leadership Questionnaire (MLQ; Bass & Avolio, 1991) was administered by BOLDS researchers to a sample of cadets each subsequent term. This instrument measures at least three leadership styles (transformational leadership (TF), transactional leadership, and laissez-faire leadership) and two outcomes of leadership (extra effort by followers and satisfaction with leader).
The MLQ was first administered in the fourth term using a shorter 39-item version (Avolio, Bass, & Jung, 1996). The respondents used a five-point scale (ranging from “frequently if not always” to “not at all”) to indicate their perceptions of how frequently the person being rated displayed the behavior described. Different forms of the instrument exist, such that individuals can rate themselves, their leaders, or their subordinates. Although data were collected from self, supervisor, subordinates, peers, and USMA instructors, only the self-assessment scores and peer-rated scores are discussed below, since the others have much smaller number of raters for any semester.

On the short MLQ, each of the five scales described above was represented by a different number of items. TF contained 20 items (five sub-scales of four items) more than half the questionnaire. (Higher scores indicate a more frequently displayed behavior; specifically: 1 = not at all, 2 = once in a while, 3 = sometimes, 4 = fairly often, 5 = frequently, if not always.) Reliability coefficients for each scale (see Table 3) are slightly lower than the range (.73 to .89) cited by Bass (1996) for eight of the scales (excluding the attributed charisma scale) derived from the 70-item MLQ-Form 5. However, the overall reliability for the Self-rating of 370 Cadets (who also took the TKML and CPR) was .867; and their peers’ ratings had a reliability of 0.778.

<table>
<thead>
<tr>
<th>MLQ Scale</th>
<th>Cronbach’s Alpha (Range) for Self Ratings</th>
<th>Cronbach’s Alpha for Peer Ratings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Idealized Influence</td>
<td>.61 - .69</td>
<td>.75</td>
</tr>
<tr>
<td>Inspirational Motivation</td>
<td>.66 - .75</td>
<td>.83</td>
</tr>
<tr>
<td>Intellectual Stimulation</td>
<td>.61 - .72</td>
<td>.84</td>
</tr>
<tr>
<td>Individualized Consideration</td>
<td>.54 - .67</td>
<td>.69</td>
</tr>
<tr>
<td>Attributed Charisma</td>
<td>.59 - .72</td>
<td>.81</td>
</tr>
<tr>
<td>Contingent Reward</td>
<td>.49 - .67</td>
<td>.76</td>
</tr>
<tr>
<td>Mgmt-by-Exception (Active)</td>
<td>.59 - .67</td>
<td>.74</td>
</tr>
<tr>
<td>Mgmt-by-Exception (Passive)</td>
<td>.65 - .73</td>
<td>.72</td>
</tr>
<tr>
<td>Laissez-Faire Leadership</td>
<td>.62 - .75</td>
<td>.80</td>
</tr>
</tbody>
</table>

(From Milan et al., 2002)
A total score for TF was computed by calculating the mean of the five behavior scales representing transformational leadership (see Table 3). Past research has found these behaviors to be highly correlated, so a total transformational score is meaningful (Bullis, Kane, and Tremble, 1997).

The BOLDS database includes both self and peer ratings from the eighth term. The correlations between these two raters’ scores are listed in Table 4.

Table 4.
Correlations Between Cadets’ Self Ratings and Peer Ratings During Term 8

<table>
<thead>
<tr>
<th>MLQ Scale</th>
<th>N</th>
<th>Spearman’s Rho^1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transformational Leadership</td>
<td>220</td>
<td>.24***</td>
</tr>
<tr>
<td>Contingent Reward</td>
<td>223</td>
<td>.16*</td>
</tr>
<tr>
<td>Active Mgmt-by-Exception</td>
<td>222</td>
<td>.40***</td>
</tr>
<tr>
<td>Passive Mgmt-by-Exception</td>
<td>228</td>
<td>.29***</td>
</tr>
<tr>
<td>Laissez-Faire Leadership</td>
<td>224</td>
<td>.34***</td>
</tr>
<tr>
<td>Extra effort by those led</td>
<td>225</td>
<td>.18**</td>
</tr>
<tr>
<td>Satisfaction with leadership abilities</td>
<td>226</td>
<td>.20**</td>
</tr>
</tbody>
</table>

*p < .05. **p < .01. ***p < .001. (From Milan et al., 2002)

RESULTS AND DISCUSSION

Table 5 contains correlations among the main scales used in the Officers’ Survey. These include the HS facets (Commitment, Control, and Challenge), the MLQ composites (Extra Effort, TF, Transactional leader style), the Career Intentions measure, the Current Health measure. The career intentions item adopted a standard format used on annual officer personnel surveys, with response options ranging from 1=Definitely will stay until retirement, to 6=Definitely will leave upon completion of present obligation. However, this format was slightly different than the questions asked previously of the Cadets so the comparison between Cadets’ and Officers’ intentions is a little ambiguous, but the predictability of Officers’ Career Intentions from other measures was expected to be improved by these changes. The Current Health item contained used a four point scale ranging from 1=Poor to 4=Excellent.
Table 5.
Officers’ Correlations Among Several Measures in the 2002-3 Survey.

<table>
<thead>
<tr>
<th>Variable</th>
<th>HS</th>
<th>CM</th>
<th>CO</th>
<th>CH</th>
<th>EE</th>
<th>TRF</th>
<th>TRA</th>
<th>CI</th>
<th>HE</th>
</tr>
</thead>
<tbody>
<tr>
<td>HS Hardiness</td>
<td>1</td>
<td>.76*</td>
<td>.67*</td>
<td>.67*</td>
<td>.35*</td>
<td>.48*</td>
<td>-.09</td>
<td>-.18*</td>
<td>.18*</td>
</tr>
<tr>
<td>CM HS Commitment</td>
<td>.76*</td>
<td>1</td>
<td>.42*</td>
<td>.17*</td>
<td>.22*</td>
<td>.33*</td>
<td>.055</td>
<td>.285</td>
<td>.12+</td>
</tr>
<tr>
<td>CO HS Control</td>
<td>.67*</td>
<td>.42*</td>
<td>1</td>
<td>.15*</td>
<td>.33*</td>
<td>.45*</td>
<td>.051</td>
<td>.027</td>
<td>.19*</td>
</tr>
<tr>
<td>CH HS Challenge</td>
<td>.67*</td>
<td>.17*</td>
<td>.15*</td>
<td>1</td>
<td>.21*</td>
<td>.26*</td>
<td>.15*</td>
<td>.04</td>
<td>.09</td>
</tr>
<tr>
<td>EE Extra Effort</td>
<td>.35*</td>
<td>.22*</td>
<td>.33*</td>
<td>.21*</td>
<td>1</td>
<td>.69*</td>
<td>.04</td>
<td>.05</td>
<td>.13+</td>
</tr>
<tr>
<td>TF TransFormational Leadership</td>
<td>.48*</td>
<td>.34*</td>
<td>.45*</td>
<td>.26*</td>
<td>.69*</td>
<td>1</td>
<td>.09</td>
<td>-.07</td>
<td>.13+</td>
</tr>
<tr>
<td>TRA Transactional Leadership</td>
<td>-.09</td>
<td>-.05</td>
<td>.05</td>
<td>-.15*</td>
<td>.04</td>
<td>.09</td>
<td>1</td>
<td>.02</td>
<td>.04</td>
</tr>
<tr>
<td>CI Career Intentions</td>
<td>.18*</td>
<td>.28*</td>
<td>.03</td>
<td>.04</td>
<td>.05</td>
<td>.07</td>
<td>.02</td>
<td>1</td>
<td>.06</td>
</tr>
<tr>
<td>CH Current Health</td>
<td>.18*</td>
<td>.11+</td>
<td>.19*</td>
<td>.09</td>
<td>.13+</td>
<td>.13+</td>
<td>.04</td>
<td>.06</td>
<td>1</td>
</tr>
</tbody>
</table>
Notes: Due to missing data, N ranged from 369 to 388. *: Correlation is significant at the 0.01 level (2-tailed); +: Correlation is significant at the 0.05 level (2-tailed).

The Hardiness (HS) facets of commitment, control and challenge show the expected pattern of interrelations and correlations with total HS. More interestingly, HS correlates strongly with Transformational Leadership (TF) style ($r = .48$, $p<.001$). Earlier work with this cohort also found this correlation in the cadet data, so the present finding replicates the earlier result. Furthermore, the earlier work showed that although HS and TF are correlated, they are both independent predictors of leader performance (Bartone, 2003). This suggests that HS may be an important personality precursor in the development of transformational leadership.

TF is also strongly correlated with the MLQ subcomponent, “obtaining Extra Effort from subordinates” ($r = .69$, $p<.001$), a fairly common finding in studies utilizing the MLQ. While this is noteworthy, the correlation is likely inflated somewhat by “method variance,” since both variables are based on the same rating scale. Hardiness also correlates with Extra Effort ($r = .35$, $p<.001$), and this correlation is less vulnerable to the problem of same-source variance, although the use of a similar Likert scoring system makes some overlap in psychophysical use of numbers and Likert categories likely.

**Career Intentions**

Upon beginning their West Point careers, one-fifth of cadets planned to stay in the Army until retirement (21%), whereas two-fifths were undecided about staying beyond their five-year obligation (41%). Cadets whose parents were on active duty or retired from the military were more likely than others to indicate they planned to stay in the Army until retirement, as shown in Table 4. For a comparison of cadets’ career intentions with the intentions expressed by 2LTs and 1LTs in 1992 and 1996, see Table 5.

**Table 6.**
**Difference in Cadets’ Career Intent Based on Having Parents in the Military**

<table>
<thead>
<tr>
<th>Career Intentions Upon Entering USMA</th>
<th>% Cadets with Military Parents ($N = 202$)</th>
<th>% Other Cadets ($N = 713$)</th>
<th>% All BOLDS Cadets ($N = 915$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plan to stay in Army until retirement</td>
<td>29</td>
<td>18</td>
<td>21</td>
</tr>
<tr>
<td>Plan to stay beyond obligation but undecided about staying until</td>
<td>28</td>
<td>28</td>
<td>28</td>
</tr>
<tr>
<td>------------------------------------------------</td>
<td>------</td>
<td>------</td>
<td>----------------</td>
</tr>
<tr>
<td>Undecided about staying beyond 6-year obligation</td>
<td>35</td>
<td>43</td>
<td>41</td>
</tr>
<tr>
<td>Probably leave Army upon completion of obligation</td>
<td>6</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>Definitely leave Army upon completion of obligation</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

(From Milan et al., 2002)

Table 7.
Career Intentions of Cadets in 1994 1998 and 2002-3

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Plan to stay in Army until retirement</td>
<td>16</td>
<td>31</td>
<td>21</td>
<td>20</td>
<td>20</td>
<td>8</td>
<td>18</td>
</tr>
<tr>
<td>Plan to stay beyond obligation but undecided about staying until retirement</td>
<td>24</td>
<td>22</td>
<td>28</td>
<td>22</td>
<td>25</td>
<td>21</td>
<td>5</td>
</tr>
<tr>
<td>Undecided about staying beyond 6-year obligation</td>
<td>28</td>
<td>19</td>
<td>41</td>
<td>28</td>
<td>21</td>
<td>51</td>
<td>20</td>
</tr>
<tr>
<td>Probably leave Army upon completion of obligation</td>
<td>19</td>
<td>12</td>
<td>7</td>
<td>16</td>
<td>18</td>
<td>9</td>
<td>14</td>
</tr>
<tr>
<td>Definitely leave Army upon completion of obligation</td>
<td>13</td>
<td>16</td>
<td>3</td>
<td>14</td>
<td>17</td>
<td>11</td>
<td>24</td>
</tr>
</tbody>
</table>

The survey results of LTs and CPTs match traditional career intentions of LTs more closely than they do the first survey of BOLDS Cadet career intentions, although the correlation with original career intentions in 1994 is significant ($r = 0.22, N = 338; p < .001$). Because this question was not asked in the same format on the Survey as in the BOLDS database, the three categories of “probably leaving before retirement but after completion of obligation” cannot be
matched consistently. The visible trend is a decline in their intent to stay until retirement and an increase in their intent to leave the Army upon completion of their obligation. Since these officers were near the end of their obligations, these intentions may be much firmer than their earlier intentions.

Finally, looking at retention as informed by career intentions, TF is unrelated to career intentions, but HS is \((r = -0.18, p<.001)\). The higher one’s HS levels, the less inclined one is to leave the service (stated intentions, not actual behaviors, although other studies do show that intentions are strongly predictive of behaviors in this regard). Examining the HS facets, it appears that HS Commitment is most strongly linked to career intentions \((r = -0.29, r=.001)\). In the HS conceptual framework, commitment represents something more global and deep-seated than organizational commitment. HS commitment is rather an encompassing and abiding belief that the world is an interesting and meaningful place, and the high-hardy person tends to extend this outlook to his or her social world, physical world, and world of the self. Relating this to a military officer career, the high hardy person is likely to see great intrinsic meaning and value in the officer role, beyond any financial or material benefits. This can also be understood in terms of high level of commitment to the military as a profession (Forsythe, Snook, Lewis & Bartone, 2002).

Reliability of the Officers’ Measures

The coefficient alphas for the 387 officers with complete responses were 0.840 for the MLQ, 0.853 for the CPR, and 0.896 for the HS. Combined with the high reliabilities for the same measures as Cadets, there is every reason to expect good reliability across the five-year gap between administrations (Nunnally, 1967).

Stability of Measures Between Cadets and Officers

We next examined the five–year stability of the main measures taken in both the Cadets’ BOLDS database and the Officer Survey.

**MLQ**
Figure 1. Stability of the means across the 5 years from testing in Term 8 (1997) at USMA to responses from the field in 2002 – 3 is very strong.

The Officers’ survey results are predicted better by Cadets’ Self ratings than Cadets’ Peer ratings ($r = 0.99$ and $0.82$, respectively). The very impressive stability of the means on each of the 39 MLQ questions (See Figure 1) for both the self - and peer - rated scales suggests an improved strategy for creating a standard to score the MLQ. A CBA standard may be appropriate.

An Alternative Interpretation of the MLQ Ratings

In any real leader, it is probably impossible to engage in all varieties of excellent, transformational leadership behaviors at all times. Compromises are necessary. At times it is even necessary to engage in non-leadership, or Laissez-faire inactivity. The best leadership is then a working compromise among all the different behaviors assessed by the MLQ. This golden mean or balance among all the behaviors is exactly the objective measure indexed by the profile of MLQ means for all the questions. We call this standard created by the means on each question, the Consensus Based Assessment (CBA) standard. Our research (Legree, Psotka, Tremble, and Bourne, 2005; Legree and Psotka, 2005) shows that novices and journeymen offer excellent CBA standards when experts are unavailable, correlating almost perfectly with experts and senior officers when the sample size is reasonably large, as it is here. This CBA
standard is often superior to traditional standards that are based simply on the maximum rating possible.

The implications of the CBA standard are somewhat at odds with the instructions on the MLQ. The MLQ instructs people to rate their perceptions of how frequently they display a behavior, and then scores that rating on a scale that essentially says that certain questions are more transformational, and the higher the rating on those questions, the more transformational that person is going to be. Yet, it is patently impossible to perceive that you are doing so many possibly good and effective leadership behaviors all the time, so somehow each person must judge which behaviors are relatively more frequent. It may be more plausible that each person is judging the relative frequency of these various behaviors in accord with their notions of how important or effective they are. If the Cadets and Officers are using their tacit knowledge and experience gained by their years at USMA, and in the field, their enculturation into the Army’s norms and standards, to judge their own profile of actions, they may well be judging this optimal balance of effective behaviors in the same way as more experienced leaders and the culture of the Army dictates. The high correlations (0.99) of their means as Cadets with their means as LTs and CPTs 5 years later appears to offer excellent evidence that this may be the case. This view of their ratings is somewhat at odds but overlaps the traditional view of the MLQ, so we have chosen to give it a new name - Balanced Tacit Framework for Leadership (BTFL – MLQ) (Psotka, Legree, Lavoie, Matthews, and Sherwood; 2005.)

Balanced Tacit Framework for Leadership (BTFL)

In a Balanced Tacit Framework for Leadership (BTFL) leaders acquire with experience an understanding of their actions in a situated context of the Army, within the shared meaning and culture of the Army, that some actions are more important and effective than others. Their goals and plans are mediated by actions that are generally more or less effective, but that act in concert to reach desired effects. These actions reflect the culture of practice of the Army that arises in complex ways from experience, training, and the official doctrine of the Army. Since much of the Army’s culture is tacit, the knowledge, skills, and attitudes grow slowly. Some of this tacit framework reflects understanding and knowledge that has an empirical basis, a kind of ground truth about the Army, while other components are simply beliefs about practices that are too complicated for anyone to understand thoroughly. The behaviors surveyed in the MLQ arise both from those generally observed in Army leaders and from a theoretical framework that has its roots in broader analysis of leadership thoughts and actions. Transformational behaviors are generally thought to result in better leadership, while Laissez Faire behaviors are theoretically less effective. When cadets and officers rate themselves and their peers on the MLQ items they
appear to be using an ideal standard for Army leaders that correlates highly with actual means observed in any representative group of leaders (Psotka et al., 2005) rather than the extreme maximum or minimum for the items. This leads to the BTFL. Transformational behaviors will be observed more frequently in a better leader, but if they are too frequent compared to the BTFL consensual standard, they may indicate an imbalanced leader, incapable of dealing competently with all of the complex aspects of leadership. Similarly, the negative non-leadership behaviors will in general be seen less often in a good leader, but they may be necessary at times, and if they are taken too infrequently, the BTFL consensual standard may once again be broken. This then, in sum, provides a viewpoint and argument for asserting that the BTFL - MLQ CBA standard may in fact be a better measure of good leadership than the traditional scoring method.

The requirement for expert knowledge, balance and wisdom (Sternberg, 2004) in the application of leadership behaviors suggests that the MLQ consensus score based on the BTFL (BTFL-MLQ) will show improvement with experience, and will be more strongly related to the TKML-PLQ score and to other measures of the convergent validity of these instruments, such as the CPR and Career Intent measure. Of course, the critical component of a good leadership assessment scale, the development of questions that are meaningfully related to leadership, especially charismatic or transformational leadership, as part of the MLQ remains unchanged by this change in assessment standard in the BTFL–MLQ.

Comparison of the Traditional MLQ and the BTFL-MLQ Scores.

The TF component is the largest component of the MLQ (20 out of 39 questions) and the Term 8 score shows good stability over the five – year period between Cadets and Officers’ testing \( r = 0.41; N = 121 \). The stability is slightly improved if TF is averaged across Terms 5, 6, and 7: \( r = 0.43; N = 275 \). However, the BTFL-MLQ measure yields much better stability \( r = 0.59; N = 122, p< .01 \). This stability also holds for the mean ratings of each group on each question in the scale: The stability across all 15 questions’ means was \( r = 0.987, (37 \text{ df}, p < .001) \).

At first glance, this improved stability of the Consensus – based BTFL-MLQ over the Traditional MLQ and TF might be because it does not capture the change in leadership over the years and so does not show the re-ordering of which Cadets are better leaders. However, in fact, the reverse was true. The traditional MLQ means did not change significantly over this period: (TF cadets = 4.16; TF officers = 4.28; n.s.) The experience of the Cadets in the field should be expected to increase their leadership skills considerably, including all the main components. Specifically, Avolio and Bass (1991) proposed that transformational leadership would be most highly correlated with effectiveness followed by transactional and non-transactional styles of leadership, which together,
constituted what they referred to as “a full range of leadership”. If the BTFL-MLQ is a more accurate assessment of effective leadership skills, knowledge and attitudes than the MLQ measure, this increase should be evident in the scores. In fact, none of the traditional MLQ subscale components showed significant changes between 1997 and 2003. On the other hand, the BTFL-MLQ measure (the average of each person’s Pearson correlation coefficient against the consensus based standard) showed a significant increase: (BTFL-MLQ cadets = 0.72; BTFL-MLQ officers = 0.82; t = 6.1; 121 d.f.; p < .001). In fact, the high mean score of 0.82 for the officers and the distribution of scores (see Figure 2) provides an indication that in the BTFL-MLQ there is a real ceiling effect: the MLQ questions appear to be too easy to capture the full range of expertise of these officers. Converting the Pearson coefficients to z-scores did not change this ceiling effect skewness appreciably, although it is sometimes recommended (Nunnally, 1967). Multi- Factor or Transformational Leadership is posited to be a set of skills, knowledge, and values that increase with experience, not inborn or fixed at an early age. Several efforts to teach and train the “Full Range of Leadership Development” have shown modest success (Avolio and Bass, 1994).

**Figure 2:** A ceiling effect in Officers’ self-rated BTFL-MLQ scores on the 2002-3 survey, in spite of which there is an increase above their scores as Cadets. The modal response for Officers is in the two categories below the ceiling of 1. These are actual frequencies, not cumulative frequencies.
The implications of these differences between the BTFL-MLQ and the traditional MLQ will next be examined in their interrelationships with the other measures on a subset of Cadets who were members of both the BOLDS and Survey datasets.

**Hardiness Scale (HS)**

The 5-year stability of the HS measure was 0.54 (n = 283; p< .01). Intriguingly the means for the two groups were not significantly different: 31.2 for Officers and 31.6 for the matched Cadets. Neither commitment, challenge, nor control changed even though their work experience might be expected to have had an effect on their commitment or control. However, these are enduring attributes that may be developed in childhood and are resistant to change. This stability also held for the mean ratings of each group on each question in the scale: The stability across all 15 questions' means was r = 0.95, (13 df, p < .01).

Table 8.
The intercorrelations between the Cadets’ and Officers' HS and BTFL-MLQ and MLQ.

<table>
<thead>
<tr>
<th>Pearson Correlations</th>
<th>Officers’ Hardiness</th>
<th>Cadets’ Hardiness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cadets’ Commitment</td>
<td><strong>0.37</strong></td>
<td><strong>0.75</strong></td>
</tr>
<tr>
<td>N</td>
<td>217</td>
<td>428</td>
</tr>
<tr>
<td>Officers’ Commitment</td>
<td><strong>0.76</strong></td>
<td><strong>0.40</strong></td>
</tr>
<tr>
<td>N</td>
<td>376</td>
<td>215</td>
</tr>
<tr>
<td>Cadets’ Control</td>
<td><strong>0.36</strong></td>
<td><strong>0.69</strong></td>
</tr>
<tr>
<td>N</td>
<td>219</td>
<td>428</td>
</tr>
<tr>
<td>Officers’ Control</td>
<td><strong>0.68</strong></td>
<td><strong>0.33</strong></td>
</tr>
<tr>
<td>N</td>
<td>376</td>
<td>215</td>
</tr>
<tr>
<td>Cadets’ Challenge</td>
<td><strong>0.33</strong></td>
<td><strong>0.65</strong></td>
</tr>
<tr>
<td>N</td>
<td>218</td>
<td>428</td>
</tr>
<tr>
<td>Officers’ Challenge</td>
<td><strong>0.66</strong></td>
<td><strong>0.42</strong></td>
</tr>
<tr>
<td>N</td>
<td>376</td>
<td>215</td>
</tr>
<tr>
<td>Cadets’ BTFL-MLQ</td>
<td><strong>0.24</strong></td>
<td><strong>0.32</strong></td>
</tr>
<tr>
<td>N</td>
<td>121</td>
<td>156</td>
</tr>
<tr>
<td>Cadets’ TF (MLQ)</td>
<td>0.18</td>
<td>0.11</td>
</tr>
<tr>
<td>N</td>
<td>103</td>
<td>108</td>
</tr>
<tr>
<td>Cadets’ extra effort (MLQ)</td>
<td>0.13</td>
<td>0.12</td>
</tr>
<tr>
<td>N</td>
<td>103</td>
<td>109</td>
</tr>
</tbody>
</table>

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

The Cadets’ BTFL-MLQ scores correlate significantly with Officers’ HS ratings, whereas the traditional MLQ measures in Term 8 do not. It appears that whatever enduring attributes or knowledge are measured by HS and overlap with
the questions of the MLQ, the attributes or knowledge are better assessed with a CBA standard than with the traditional extreme rubrics.

_Cadet Performance Report (CPR)_

The CPR is a rating instrument consisting of 12 leadership dimensions: namely, duty motivation, military bearing, teamwork, influencing others, consideration for others, professional ethics, planning and organizing, delegating, supervising, developing subordinates, decision-making, and oral and written communication. A rated cadet is given a score of 1 (“needs much improvement”) to 5 (“excellent”) on each dimension. In addition, two dimensions are selected as the cadet’s relative strengths and two as his or her relative weaknesses. Lastly, cadets are given an overall ranking that indicates whether their performance is in the upper 10%, upper 25%, middle 30%, lower 25%, or lower 10% of cadets in that particular duty position.

CPRs are primarily completed by cadets, from superior, peer, and subordinate positions. Cadet leaders are required to complete CPRs on those in subordinate duty positions. Peer and subordinate raters, on the other hand, nominate and rate the cadets having the highest and lowest performance relative to others in a particular group (for details, see Schwager & Evans, 1996). Academic instructors follow similar procedures, by nominating and rating the highest and lowest performers enrolled in their courses. Hence, it is possible for any cadet to receive multiple peer, subordinate, and instructor CPRs or none at all, depending on whether he or she is considered a high, low, or average performer. Based on the contents of the peer and subordinate CPRs, cadets receive developmental counseling from their Tactical Officer. For the BOLDS research, cadets were asked during their third and fourth years to complete CPRs on themselves. While nearly all cadets received a CPR from at least one type of rater each term, relatively few cadets consistently received a CPR from each type of rater. Accordingly we compared only the self – ratings of the CPR as Cadets and Officers.

The 5-year stability of the measure was 0.37 (n = 283; p< .01) for their ratings and just 0.23 (n = 283; p< .05) for their CBA scores. Intriguingly the mean self -ratings for the two groups were not significantly different: 4.07 (N=376) for Officers and 4.05 (N=687) for Cadets even though their work experience might be expected to have had an effect on their leader performance self-ratings. A consensus-based measure did show a marginally significant increase: 0.34 for Cadets and 0.42 for Officers (t = 2.8; N=283; p < .01). However, these scores are quite low. This stability also held for the mean ratings of each group on each question in the scale: The stability across all 15 questions’ means was r = 0.85, (12 df, p < .01), again significantly lower than the stability for the BTFL-MLQ and Hardiness scales.
Career Intent (CI)

The career intent measures’ stability between the 338 Army respondents as Cadets and Officers was quite low: $r = 0.22$. However those Officers who intended to leave upon completion of their present obligation came proportionately from all the other Cadet categories of intent. Several measures correlated significantly with the survey career intent. Hardiness as Cadets correlated as highly with it ($r = -0.17; N = 210$) as did Officers’ hardiness. The Cadets’ self-rated BTFL-MLQ correlated more highly ($r = -0.28; N = 119$) with Officers’ CI than the Cadets’ self-rated MLQ traditional measure of TF ($r = 0.12; N = 269$) although both were significant ($p < .01$). The Cadets’ TKML measure also was predictive of CI: $r = -0.15$, although with only 77 matched respondents it just failed significance ($p < .10$). The CPR was not predictive of career intent ($r > 0.05$) although the consensus – based measure of CPR was ($r = -0.13, N=281$, $p < .05$). In the combined linear regression of these measures CPR did not add any significant predictability to the other three measures, and Hardiness, BTFL-MLQ, and TKML together correlated $r = 0.45$ with career intent ($3, 40 \text{ df}, p < 0.05$).

DISCUSSION

Although there is some evidence that a consensus-based standard (Legree, Psotka, Tremble, and Bourne, In Press) proved to be superior psychometrically for the MLQ and CPR, it showed no real improvement for the Hardiness scale. A rationale for the superiority of the BTFL-MLQ was given by pointing out that the instructions for the MLQ suppose that all behaviors, either good or bad, should be implemented all or none of the time, but this implication is impractical if not self destructive for good leaders in real practice. Instead of trying to maximize or minimize behaviors, officers should seek out the culturally defined balance for each behavior. If there is a balanced, optimal level for each behavior and the CBA means provide the best index for that optimal performance, then the groups means for each item should be the best standard for scoring the scale. No such rationale was provided for the Hardiness scale or the CPR, although a similar line of reason appears appropriate for the CPR

Consensus Based Standards and the CPR

In deciding how to rate oneself on the CPR, in deciding how well one takes care of one’s soldiers, motivates them, and engages in all the other duty motivation to decision-making skills and attributes assessed by the scale; the respondent must have a standard of some kind that arises from his or her understanding of the Army standards in the prevailing community. It is easy to imagine that this shared meaning created within Army organizational culture is a moving standard that changes with the officer’s knowledge and experience; and since it is a standard applied to oneself by oneself it seems readily
understandable that it is derived from the consensus of the surrounding community. However, since all the items were selected to be roughly equally important from the perspective of the Army’s organizational culture, the instrument has a much lower stability of the groups means for all items (0.85), and consequently lower correlation between Cadets’ and Officers’ CBA scores (0.23).

The fact that both the CPR ratings and CBA scores were so low suggests that there is a rating difficulty with the CPR. Perhaps the Army organizational culture cannot set accurate standards for these items, so that both officers and cadets find it difficult to assess whether they are performing poorly or well against those very ambiguous standards. Also, the scale items are very broad, ranging over the entire domain of leadership, so it may not be surprising that the reliability and stability of the scale are comparatively low and the scale ratings or CBA scores do not correlate highly with the other measures.

**Consensus Based Standards and Hardiness**

Given the high stability of the means on each of the 15 Hardiness scale questions (the correlation between the Cadets’ Means and the Officers’ Means on the 15 items was 0.95) it might be construed that there was an underlying consensual standard for these questions as well as for the MLQ. However, the correlation between the CBA and traditional scoring versions of the HS (0.684, N = 376 p<.01) appears to be so high as to preclude much of a difference. And indeed, an examination of the relationships among the other measures and a CBA score of Hardiness revealed entirely parallel and equivalent correlations. The CBA standard appeared to be more or less equivalent to the existing HS standard, which expects a perfect 45 out of 45 Total Score for the hardiest officers. However, an examination of the histograms for the CBA HS, in contrast with the traditional measure of HS, again showed no significant differences between cadets and officers (Officers’ Mean = 0.70 and Cadets’ Mean = 0.69), and a small ceiling effect for Officers’ CBA score (cf. Figures 3 and 4).
Figure 3: No sign of a ceiling effect in Officers’ self-rated traditional HS scores on the 2002-3 survey, and no significant difference above their responses as Cadets. The modal response for Officers is still several standard deviations below the ceiling of 45. These are actual frequencies, not cumulative frequencies.
Figure 4: A small ceiling effect in Officers’ self-rated Hardiness CBA scores on the 2002-3 survey, but no significant increase above their responses as Cadets. The modal response for Officers is not in the category immediately below the ceiling of 1. These are actual frequencies, not cumulative frequencies.

However, a closer examination of the scalar words on the HS questions, revealed a careful balancing of their degree. For instance, some questions were extreme and asked about "most of my life" or even "always" achieving a goal; while others were more moderate, asking “not very much”; and still others left the evaluative word unmodified and moderate, as "like" or "enjoy". Six psychologists rated the degree of these evaluative words on a 0 to 4 scale, from never or not at all to always, or a great degree. The rated degree of each question against the overall, combined means of Cadets and Officers for each question (see Fig. 5) revealed a powerful correlation ($r = 0.71$, 13 d.f.; $p < .01$). The means of Cadets and Officers for each question is, of course, the CBA standard. So it seems that the careful preparation of the HS scale empirically created a standard that to some degree already mimics the CBA standard.
Figure 5: A possible explanation for the lack of difference between CBA scoring of HS and traditional scoring of HS. The evaluative words of each question (e.g. most, really, some, like, dislike, etc.) were scored for their evaluative degree by 6 Professionals and correlated with the means of each question.

Culture versus Facts

If the CBA standard is to be effective in creating an assessment of leadership that empirically predicts officers’ knowledge, skills, and attitudes and a successful career, it would seem important that it accurately reflects some sort of empirical facts about leadership, not simply a cultural consensus without an underlying ground truth. In previous writings (Legree et al., 2000, 2004, 2005) we have proposed an argument, that (greatly simplified) suggests that consensus arises in these instruments by averaging out noise, removing it and revealing an accurate image of the measured components underneath. Our metaphor for this process is like the averaging that astronomers use to get an accurate image by superimposing multiple images on each other (e.g. Figure 6.)
Figure 6. Clarifying Detail by Averaging. These three pictures were created from a sequence of images of Titan. They illustrate how the details of Titan's surface can be revealed through averaging. The picture on the left is a single image that has undergone only basic cleaning of corrupted pixels and imperfections in the camera's charge coupled device, a light-sensitive detector similar to those found in digital cameras. In the middle frame, multiple images were used to enhance the contrast detected from Titan's surface and to reduce the blurring effect of atmospheric haze. The picture on the right has added even more images.

However, the consensus that Cadets and Officers achieve is not as easy to analyze as the images of Titan. The distorting effects of our atmosphere may correspond to many different sources of distortion in the Cadets' and Officers' opinions, and it is not necessary at this point to define those sources in detail, as long as they are effectively removed, but the real question is whether there is a common reality under the noise, or whether there is only a consensus based on the Army's organizational culture, consensus and communication. In other words, are the behaviors effective even if they are effectively implemented? Are they effectively implemented even if they are often used? For example, an officer may realize that it is important to provide a vision statement of a unit's mission, and even promote that vision frequently, but does he or she have the knowledge, skill, and attitudes to do it effectively? And even if it is done effectively, is this truly one of the most important activities to engage in, given the limited time and resources available, or is there something even more important that needs doing?

The simplest answer seems to be that given the current state of the art and science of leadership, it is impossible to know accurately, but the intelligent, dedicated efforts of the profession that have created the Army's organizational culture offer the best possible approximations to the most effective methods available, in the absence of an accurate science of leadership. Some of those efforts are undoubtedly based on empirical trial and error that reflects accurately how leadership, in all its diversity, from interpersonal relationships to organizational structure and function, actually works. Others may be based on
much less knowledge in new and unpredictable circumstances, and may be dead wrong.

So the existence of consensus itself is not a guarantee of effectiveness, but may simply be our best index of effectiveness currently available until the science of leadership catches up to practice.

**Level of Detail**

Comparing the various instruments, the MLQ, HS, CPR, and the TKML on their level of detail is instructive. Of the four, only the TKML delves more precisely into the thoughts, actions, and values of leaders to elicit differences that would be elided in the general questions of the others. The TKML begins to fill in the detail in the general categories outlined by the others. Where they ask general questions that leave some doubt whether or not the leader actually knows how to carry out the intentions described so generically, the TKML explicitly poses alternatives that test an officer's capabilities against the wisdom of his or her peers and superiors. Of course, this process could be continued practically indefinitely to ever finer grain size of issues and knowledge. At some point it might simply be evaluating whether or not an officer had any relevant experience with a particular situation. Specific situations and greater level of detail may well assess an officer's understanding of diffuse causal relationships among actions and consequences within an Army environment. More general scales may assess the broader understanding of the meaning of words that have more specific meanings within an Army organizational culture but also draw broadly on the meanings derived from experiences in the broader culture. Which of these levels tap more effectively into the knowledge, skills, and attitudes required for effective leadership may not be discovered until a thorough longitudinal analysis of Army leaders progresses throughout entire careers. However, it would seem that a good argument could be made that more specific levels of detail in instruments should tap more cleanly into the particulars of Army organizational culture, and so be more predictive of Army leadership effectiveness.

It is also possible that the more specific level of detail may be more empirically tied into causal relationships than the more general knowledge levels. Although grand generalizations are the hallmarks of scientific achievement in well established science, they may be more difficult to achieve and untrustworthy in the social sciences. And they may be particularly difficult to discern in assessment scales that are very general.

However, at this point in psychometric development, it is not clear how we would use the vast knowledge base of information such a program of detailed assessment would produce. Nevertheless, we could create an encyclopedic compendium on all the knowledge, attitudes, and values an officer possesses, from the perspective of the culture of the Army at this point in history and use it
as an assessment instrument. The possibility of doing this is awesome, and forces us to reflect carefully on the best way to proceed.

One of the real strengths of CBA standards is that the procedures for creating accurate assessment are just as applicable at the more specific knowledge levels of detail as at the abstract levels in the MLQ, HS, or CPR. CBA standards make it possible to compare the efficacy of assessment scales that include self assessment on the same continua as objective tests or peer assessment. Using CBA scales it is possible to examine opinion – like items in self assessment in two ways; one simply at face value as an expression of a metric that uses someone’s implicit understanding of the underlying scale directly and at face value (with perhaps some contingencies for errors from pressure to fake and look good); and the other simply as a test of the respondents' understanding and knowledge of what precisely it means to be the best, to look good. By taking advantage of the cultural influences that act on an officer to be the best, the CBA procedure stands “faking” and “looking good” on its head and challenges the respondent to use all the knowledge available about Army organizational culture to decide just what the best possible opinion is. So, for instance, in the CPR, just how willing should one be to listen to subordinates? Or in the MLQ, just how often should one seek differing opinions when solving problems? Or in the HS, how true should it be that you know when to seek help? By examining these scales both with traditional scoring, and with scores based on CBA, it may be possible to extend our insights about leadership knowledge, skills, and attitudes in new ways that improve our overall understanding.

CONCLUSION

This paper adds to the gradually accumulating evidence that CBA assessment technology provides a truly revolutionary capability to psychometrics. It has explored the implications of using CBA standards within several highly respected assessment scales, the MLQ, HS, and CPR and discovered a broad range of outcomes. With the MLQ, the CBA standard appears to offer insight into an officer’s understanding of the factors that make for an “ideal” effective leader, and the accuracy of that understanding (compared to the Army organizational norm as expressed by the entire group of Army officers assessed) appears to be a useful measure of each officer’s leadership knowledge, skill and values. Using the CBA standard transforms the traditional MLQ into a new scale, the BTFL - MLQ that appears to have more in common with the TKML and appears to assess an understanding of “prototypical leadership” that in other contexts, has been demonstrably related to leadership effectiveness (e.g., Hogg & Knippenberg, 2003; Gerstner & Day, 1994) In conjunction with astonishing advances in automated knowledge representation, retrieval, and manipulation techniques (cf. Streeter, Lochbaum and Psotka, 2005) CBA standards make it possible to expand the assessment of knowledge and facts to probabilities, possibilities, inferences, attitudes, and values, at all levels, from the most abstract and general semantic relations to the most detailed assertions available.
Not only is the CBA standard a powerful assessment tool, it is remarkably economical. Existing self-assessment rating instruments, such as those focused on in this paper, can easily be re-analyzed from the CBA perspective, with sometime surprising results.

The truly surprising result in these findings is the apparent ceiling effects of the BTFL-MLQ on scores of officers and even cadets (cf. Figure 1). The ceiling effect on Officer's scores is most apparent, but the skewness of the Cadet scores suggests that even they found the BTFL-MLQ too easy. This is a somewhat radical finding, that a self-assessment instrument is too easy! But the consensus-based CBA standard leads convincingly to this conclusion. Even the Hardiness scale, measured with the CBA standard, gave some sign of being too easy. Since the CBA standard scoring converts an opinion scale into a kind of maximal knowledge test, it provides a novel perspective on opinion-like rating scales, particularly self-assessment scales. How these two perspectives relate to each other is not at all self-evident, and this paper only begins the exploration of these relationships.

**Prototypical Leadership**

In the case of the MLQ, we have made a case that the relationship is mediated by knowledge of the Army organizational culture and its understanding of what ideal or prototypical leadership is. We have argued that the BTFL-MLQ using the CBA standard essentially assesses an officer's understanding of what leadership best practices are within Army culture. In fact, it may be that the traditional MLQ ratings are not doing much more than that and should only be interpreted as a BTFL-MLQ. If an officer is performing at his best, and doing all the things that he or she thinks an ideal leader should be doing, then in fact the self-assessment should be nothing more than a description of the ideal leader from the perspective of Army culture, as far as that officer understands it. If the officer, however, feels that he or she is an inadequate leader in some ways, then the self-assessment might deviate from ideal leadership along the dimensions of inadequacy. If so, then the differences between the CBA standard and self-ratings could reflect either a misunderstanding of the Army cultural perspective on leadership or a self-appraisal that honestly believes it falls short in some ways. At the moment we have no way of distinguishing these alternatives, but perhaps another question or two might clarify the issue. Further research might make this clear.

The notion that a leader's typicality or nearness to an Army ideal is a basis for leadership effectiveness by no means implies that prototypical leaders can only behave like "the average group member", even if the CBA standard is described by group means. The Army organizational prototype describes leadership actions, knowledge and values that are generally effective, but understanding these ideals not only requires an internalization of those standards
but also constructing an entire complex of shared meanings. These shared meanings cannot be purely definitional and inert; they have to be integrated into each officer’s semantic knowledge system in ways that they can be applied in the proper contexts and with the right goals and effects, with some shared meanings more central to the group's identity than others. The prototypical group member is thus, in fact, closer to a representation of the ideal group member than to the typical or average group member. It is therefore also possible that the prototypical leader is, for instance, rather average in some respects and exceptional or unconventional in others. Prototypicality provides leaders with more leeway in their behavior (in the eyes of their followers, peers and superiors) and thus positions them to effectively engage in behavior that may lead the group or organization in new directions. Being representative of the group's identity as a basis of leadership effectiveness is thus not at odds with engaging in special, unusual, or distinctive behavior; indeed, it sets the stage for effectively engaging in such behavior. In sum, displaying unusual and unconventional behavior, which is sometimes seen as a basis for leadership effectiveness is perfectly consistent with maintaining (or even actively enhancing) prototypicality.

In fact, the Hardiness scale appeared to have implemented an approximate CBA standard in the selection of each question’s degree, expressed by the evaluative words in the question. A correlation between each HS question’s mean and the degree of these evaluative words was positive and significant. In a way this provides an indirect support for the procedures of creating and using a CBA standard, since the HS questions were developed pragmatically by winnowing the most effective ones from a much larger pool of similar items (e.g., Bartone, 1999).

These findings suggest that a systematic exploration of other assessment measures using CBA may continue to give us new insights into the complex mixtures of knowledge, skills, and values that underlie good leadership.

REFERENCES


Hedlund, J., Williams, W. M., Horvath, J. A., Forsythe, G. B., Snook, S., Wattendorf, J., McNally, J. A., Sweeney, P. J., Bullis, R. C., Dennis, R. C.,


