ARI Research Note 98-06

# Evaluation of ARI Leader Assessment Measures

**John E. Mathieu**
Pennsylvania State University

**Richard J. Klimoski**
George Mason University

**Cathy E. Rouse and Wendy M. Marsh**
Pennsylvania State University

**Research and Advanced Concepts Office**
**Michael Drillings, Chief**

April 1998


U.S. Army Research Institute
for the Behavioral and Social Sciences

**U.S. Army Research Institute**
**for the Behavioral and Social Sciences**

**A Directorate of the U.S. Total Army Personnel Command**

**EDGAR M. JOHNSON**
**Director**

Research accomplished under contract
for the Department of the Army

Pennsylvania State University

## NOTICES

**DISTRIBUTION:** This Research Note has been cleared for release to the Defense Technical Information Center (DTIC) to comply with regulatory requirements. It has been given no primary distribution other than to DTIC and will be available only through DTIC or the National Technical Information Service (NTIS).

**FINAL DISPOSITION:** This Research Note may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

**NOTE:** The views, opinions, and findings in this Research Note are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision unless so designated by other authorized documents.

# REPORT DOCUMENTATION PAGE

| 1. REPORT DATE (dd-mm-yy) | 2. REPORT TYPE | 3. DATES COVERED (from. . . to) |
|---|---|---|
| April 1998 | Final | |

| 4. TITLE AND SUBTITLE | 5a. CONTRACT OR GRANT NUMBER |
|---|---|
| Evaluation of ARI Leader Assessment Measures | MDA903-93-C-0005 |
| | 5b. PROGRAM ELEMENT NUMBER |

| 6. AUTHOR(S) | 5c. PROJECT NUMBER |
|---|---|
| John E. Mathieu (Penn. State Univ.), Richard J. Klimoski (George Mason Univ.), Cathy E. Rouse (Penn State), and Wendy M. Marsh (Penn State) | B74F |
| | 5d. TASK NUMBER 4901 |
| | 5e. WORK UNIT NUMBER C77 |

| 7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) | 8. PERFORMING ORGANIZATION REPORT NUMBER |
|---|---|
| Pennsylvania State University<br>George Mason University | |

| 9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) | 10. MONITOR ACRONYM |
|---|---|
| U.S. Army Research Institute for the Behavioral and Social Sciences<br>5001 Eisenhower Avenue<br>Alexandria, VA 22333-5600 | ARI |
| | 11. MONITOR REPORT NUMBER<br>Research Note 98-06 |

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

Companion data base (RN 98-07) is on file in ARI Library.

14. ABSTRACT *(Maximum 200 words)*:

This project grew out of a need for a cataloging, synthesis, and review of measures designed to predict and/or assess leader effectiveness developed and/or used by the U.S. Army Research Institute over the past 10 years. The purpose of this report is to review featured ARI leadership measurement initiatives and compare them to benchmarks in nonmilitary research. The objectives of the effort were to (a) identify and describe major themes and initiatives by ARI leadership labs over the past ten years, (b) critically analyze resulting instruments according to specific and common evaluative criteria, (c) compare ARI initiatives against external benchmarks, and (d) to offer suggestions and guidance for future leadership research endeavors.

15. SUBJECT TERMS

Leadership, leader assessment, Army Research Institute, measurement

| SECURITY CLASSIFICATION OF | | | 19. LIMITATION OF ABSTRACT | 20. NUMBER OF PAGES | 21. RESPONSIBLE PERSON (Name and Telephone Number) |
|---|---|---|---|---|---|
| 16. REPORT | 17. ABSTRACT | 18. THIS PAGE | | | |
| Unclassified | Unclassified | Unclassified | Unlimited | 231 | Michael Drillings (703) 617-8641 |

CBCS 97-3

# Evaluation of ARI Leader Assessment Measures

by

John E. Mathieu
Pennsylvania State University

Richard J. Klimoski
George Mason University

Cathy E. Rouse and Wendy M. Marsh
Pennsylvania State University

Final Report
Contract No. ARI050-GMU97-2

October 15, 1997

Submitted to:

The Consortium of Universities of the Washington Metropolitan Area
for the U.S. Army Research Institute for Behavioral and Social Sciences
5001 Eisenhower Avenue
Alexandria, VA 22333

Submitted by:

Center for Behavioral and Cognitive Studies
Department of Psychology
George Mason University
Fairfax, VA 22030-4444

## ACKNOWLEDGEMENTS

We thank Mary Lee Peterson and Connie Moore for their administrative support and assistance throughout this project and the preparation of this report. We also thank the ARI Research Scientists for making vast amounts of materials available to us and for their willingness to share with us the results of their research.

# Evaluation of ARI Leader Assessment Measures

## Executive Summary

### Project Overview

Currently there exists a vast array of research on variables related to Army leader effectiveness. Concomitant with this, however, there has been a proliferation of measures designed to predict and/or assess leader effectiveness. This project grew out of a need for a cataloging, synthesis, and review of such measures developed and/or used by the U.S. Army Research Institute for Behavioral and Social Sciences (ARI) over the past ten years. Accordingly, the purpose of this report is to review featured ARI leadership measurement initiatives and compare them to benchmarks in nonmilitary research. The objectives of the effort were to (a) identify and describe major themes and initiatives by ARI leadership labs over the past ten years, (b) critically analyze resulting instruments according to specific and common evaluative criteria, (c) compare ARI initiatives against external benchmarks, and (d) to offer suggestions and guidance for future leadership research endeavors.

This report examines measures employed in ARI leadership research over the period of 1987-1997. Due to the sheer number of constructs, measures, and variables researched over those ten years, the focus of our review needed to be narrowed to become manageable. For example, our initial review of general ARI-supported leadership research included over 30 technical reports, 13 research notes, 20 research reports and briefing slides, and 17 other miscellaneous documents from the ARI archives.

These numbers support a need to narrow the focus of the evaluation project. In order to accomplish this, only the most prominent and productive themes and initiatives were included. We used four primary means for narrowing our focus. First, we met with all ARI research lab directors and asked them to nominate which of their projects they considered to be most central to the purpose(s) of our project. Second, together with the lab directors, we considered the relative time and attention devoted to various projects and highlighted those that had garnered the greatest emphasis. Third, we considered the quantity and quality of documentation available. Finally, we considered the applicability of the initiatives to the larger area of leader effectiveness.

Once the major themes were identified it was necessary to create an evaluation template against which they could be gauged. Six dimensions were established to characterize instrument development and use. First, brief descriptive information is presented, such as the purpose of the construct/measure, the target population, scales, authors, publishers, etc. Second, the development and theoretical grounding of the construct/measure are identified, followed by the frequency and nature of reported use. The psychometric characteristics of instruments are reviewed, as related to reliability and validity. These include internal consistency estimates, test-retest reliabilities, and some

interrater reliabilities. Validity indices include construct and content, as well as predictive and concurrent studies. The fifth criterion is the generalizability of a measure. This identifies the various contexts in which the instrument has been used. The final criterion deals with the specific use of the instrument and how it "looks." We made judgments regarding the face validity of the items, the ease of use, and the apparent transparency of the measures based on past literature and our direct examination. Both the features ARI products and benchmarks are evaluated using these same criteria, permitting direct comparisons.

Once the ARI initiatives were identified, it was necessary to identify external benchmark measures for comparison purposes. Benchmarking essentially describes a practice of comparing research or systems of interest against similar types from outside of the immediate context. Comparisons between the ARI work and these benchmarks, along with references to the larger leadership research domain, drive the conclusions and recommendations offered at the end of the report.

## Conceptual Foundations

**Conceptualizing the measurement domain.** In light of the vast array of leader assessment measures we encountered in our literature search of ARI documentation, it became clear that some sort of organizational scheme was necessary to place efforts in perspective. For purposes of this report, therefore, we used a general Input-Process-Output framework to organize material. The input component includes **individual resource variables** (e.g., background information and demographics), leader **knowledge, skills, abilities**, as well as **other** individual difference constructs, such as attitudes, motivation, and personality (KSAOs). The process component covers approaches to assessing leader **behaviors**, such as interaction, communication, and problem-solving styles. Finally, the output deals with the **effectiveness of various leader actions** as indexed by such things as evaluations of the leader's effectiveness, unit performance, and/or subordinate followers' behaviors and reactions. In addition, following the logic of contingency theories, variables that may have the potential to moderate the impact of individual differences on leadership processes on effectiveness, such as **subordinates' attributes** and the immediate **operational environment** (i.e., task characteristics) are depicted. These are intended to illustrate the point that different combinations are likely to be most effective in different circumstances. Finally, the reader should notice that the entire framework is nested within an appreciation of a larger **contextual environment**. This is intended to illustrate that what constitutes effective leadership will likely differ depending on the organizational subculture and circumstances within which it is imbedded. For example, attributes of effective military leadership will likely differ depending on whether one is leading a drafted vs. volunteer corps, in peace time vs. peace keeping vs. combat situations, the perceived moral imperatives related to actions, etc.

**Leadership theory.** Many ARI researchers appear to subscribe to stratified systems theory (SST) of leadership. The basic assumption behind SST is that the basis of effective leadership changes across career stages and hierarchical levels and becomes

more complex at higher ranks. For example, platoon leaders primarily focus on interpersonal issues (motivating subordinates, establishing personal credibility). Company commanders are concerned about issues of coordination and the balancing of subordinate and institutional interests. At the battalion levels, commanders' tacit knowledge is primarily focused on the larger system (protecting the organization and managing organizational change).

SST outlines how different types of leader knowledge are thought to be critical at different career stages/hierarchical levels. Similarly, the different levels place a premium on different facets of leaders' personalities. Moreover, what constitutes effective leader behaviors differs across hierarchical levels. Accordingly, these three facets became central to our review and this report. In addition, however, we have featured biodata as unique measurement strategy. Biodata is a description of a measure approach that usually includes many substantive areas, including the three noted above. This comprehensiveness, however, is both an attribute and liability, as it is difficult to neatly place constructs that biodata addresses into substantive areas. Because of this property and the amount of attention devoted to biodata as a measure tool by ARI researchers, we have featured it in a separate section, but also discuss it in the more substantive sections, as appropriate.

**Describing and Assessing Inputs to Effective Leadership**

**Personality.** ARI researchers have focused greater attention on the role of leaders' personalities and, in particular, on a theme called "proclivity" from SST. They have primarily used three measures: (1) biodata, (2) the Subject-Object Interview (SOI), and (3) the Myers-Briggs Type Indicator. In addition, we reviewed the general research community's approach to personality, known as the "Big 5," and three benchmark personality inventories: (1) Hogan Personality Inventory (HPI), (2) NEO Personality Inventory (NEO-PI), and (3) California Personality Inventory. Our summary of this section suggested that while debate continues as to the precise number and composition of these factors, in general the Big 5 has been adopted as the basic structure and measurement framework of personality. The ARI researchers, however, have tended to employ measures of proclivity in an effort to operationalize features of SST. Unfortunately, they have not done so in a consistent manner. Furthermore, no study to date has attempted to fully assess the proclivity domain as articulated by SST. Consequently, it is difficult to draw firm conclusions about the role of proclivity either within or across investigations.

We suggest that some further foundation work is in order. We would also suggest that during the course of such development, one or more measures of the Big 5, such as the benchmark instruments reviewed here, be administered.

**Knowledge.** Based on the representation of knowledge-related variables in our database, as well as the nominations by the ARI research scientists, we concluded that leader knowledge was an important area to feature in this report. We should note, however, that for convenience we used the term "knowledge" fairly loosely to include

variables that are sometimes considered to be cognitive abilities or skills. While distinctions between knowledges and cognitive abilities and skills are often important in practice, taking this latitude allows us to use Fleishman's taxonomy of cognitive abilities as an organizing framework for this section. The taxonomy is arranged into five higher-order cognitive abilities, but because it still does not capture the range of knowledge-related assessments we encountered in our review of ARI work, we have added a few additional entries, including tacit knowledge and mental models. Five ARI assessment initiatives are reviewed in this section: (1) ARI Background Data, (2) ARI Critical Incidents, (3) Mental Models, (4) The Career Path Appreciation (CPA) protocol, and (5) Tacit Knowledge for Military Leadership Inventory (TKLMI). A corresponding range of external benchmark measures are featured in this section, including (1) the Watson-Glaser Critical Thinking Appraisal; (2) Concept Mastery Test; (3) Consequences; (4) a low-fidelity simulation by Motowidlo, Dunnette, and Carter (1990); (5) Leatherman Leadership Questionnaire (LLQ); (6) PathFinder (PF) analyses of paired-comparison mental model ratings (Stout, Salar, & Kraiger, 1997); and (7) tacit knowledge (Wagner, 1987).

We concluded that the ARI instruments were essentially parallel to the selected benchmarks. The development of the ARI instruments and benchmarks are comparable, with moderate to strong development efforts supporting them. (The LLQ is the exception, with a fairly weak instrumental development.)

We noted an important distinction between general vs. specific forms of knowledge. Naturally, there is an implicit tradeoff here between measurement fidelity for any given application vs. generalizability and widespread use. Accordingly, it is important for researchers to articulate what type(s) of knowledge is(are) important in their applied research context. We could easily envision applications where either, or both, general and specific knowledge assessment would prove valuable. Different research questions and applications will call for different strategies, but, in general, it makes sense to have a battery of general cognitive ability measures available for general use across future studies. Such batteries are readily available in the commercial market. This would still leave a need, in many applications, to assess more specific forms of knowledge, such as tacit or mental models. The approach adopted by ARI for these measures has been sound, in that the researchers have sought to strike a balance between sensitivity to the knowledge requirements of individual assignments, yet maintain a limited range of generalizability. Such development strategies, combined with a comprehensive job analysis of leadership positions, would help to align specific knowledge assessments with the requirements of different positions.

**Biodata.** Whereas the other three sections represent substantive variables, biodata really describes a method of measurement. Three ARI instruments are featured: (1) Civilian Supervisors, (2) Special Forces, and (3) Background Data Inventory (BDI). For comparison purposes, two benchmarks are included: (1) LIMRA's Assessment Inventory for Managers (AIM) and (2) Owens' Biographical Questionnaire (BQ).

In terms of the theory behind the ARI biodata instruments and the benchmarks, they are all based on the same concept of past behavior predicting future behavior. The differences among the instruments lie more in terms of the specific models of leader effectiveness they are based on and the dimensions that they include.

In terms of direct comparisons, four of the instruments reviewed are similar in format, with approximately the same number of items. All measures include some scales related to management skills and personality, but the diversity of the specific dimensions selected for inclusion is striking. Finally, the SF version shares more with Owens' biographical questionnaire in terms of addressing physical abilities, a lie or social desirability check, and outside interests, as compared to the other instruments.

We argued that biodata does, however, present a bit of a paradox, as it simultaneously appears to be "everything" and "nothing." Attempting to classify what biodata is proves to be very difficult. As so eloquently stated by Owens (1976, p. 623), "It is entirely appropriate to wish to allocate biodata to some position within the network of variables which constitutes the measurement domain. The task, however, it not singular but plural, since biodata is not one measure of one dimension but multiple measures of multiple dimensions. Thus, one must first decide the essential dimensions and then decide how each relates to some key variables in the domain (emphasis in original)." Following Owens' advice, we recommend that future biodata efforts adopt a more a priori framework. The prototypical procedure followed to date has been to generate a lengthy list of potential items, to reduce them using rational and empirical methods, and to derive a new set of dimensions for each application. What is needed, we suggest, is a more theory-guided approach, where specific underlying dimensions are articulated initially, items written to address those specific dimensions, and then confirmatory analyses be conducted to determine how well those dimensions were assessed. Moreover, we believe that a "core set" of leadership effectiveness-related dimensions likely exists that could be generalizable, at least across Army classifications. In other words, we believe that a core set of dimensions could be constructed and included in virtually all leader effectiveness studies where biodata predictors are warranted. Naturally, these could be supplemented with additional scales to the extent warranted by the research design, criteria addressed, sample population, etc. However, there should definitely be some (relatively large) degree of carry-over across studies.

**Describing and Assessing Process Correlates of Effectiveness**

**Leader behaviors.** Based on a review of the ARI leadership projects and discussions with ARI research scientists, leader behavior was an area that has received considerable attention. A scan of the ARI leadership database showed that 83 of the 243 variables categorized related to some aspect of leader behavior. The ARI methods of measuring leader behavior vary widely, and multiple constructs tend to be tapped. The featured ARI products for this section were (1) Multifactor Leadership Questionnaire (MLQ), (2) Cadet Performance Report (CPR), and (3) Leader Azimuth Check/Strategic Leader Development Inventory (Azimuth/SLDI). These three measures were compared to the following benchmarks: (1) Leader Behavioral Description Questionnaire (LBDQ),

(2) Leader Practice Inventory (LPI), (3) Benchmarks, (4) Campbell Leadership Index (CLI), (5) Profiler, and (6) Prospector.

We concluded this section by suggesting that both in terms of what they do and how well they do it, the ARI measures of leader behavior are comparable to the benchmark ones. Whereas the ARI instruments are essentially comparable to those available in the private sector, all, in our opinion, lack a clear focus. Some instruments focus on leader behaviors, others largely on personality-type dimensions, and most include a variety of skill assessments. This "mixed-bag" limits the extent to which these indices can be unequivocally employed as predictors or criteria in any given study. It also presents difficulties when it comes to establishing clear frames of reference for raters and targeted feedback for ratees. In short, there is a need to refocus ratings of leader behaviors on behaviors per se, not on leader attributes. We submitted that it would be advantageous to develop a 360 rating system for Army leaders that closely attends to the purpose, content, sources, and process issues. We also argued that it would be advantageous to identify a core set of leader behaviors that would apply across settings and others that would have more limited applicability.

## General Summary and Recommendations

This document chronicles the development and use of a vast array of leader assessment measures. Moreover, the number of measures reviewed here are but a subset of the ones that have been used by ARI research scientists over the past 10 years. In this section we will attempt to identify some common themes running throughout the body of work that we reviewed. In addition, we offer some recommendations for future research.

We caution the reader to appreciate, however, that the following comments must be tempered in terms of the objectives and goals for any assessment effort. In fact, we had begun this project with the hopes of classifying clearly the intended purpose(s) of each assessment device we reviewed. Unfortunately, such clarify did not exist. Some measures are used for predicting leader effectiveness, some as indices of leader effectiveness, some as both, yet others as neither. Therefore, our following comments are framed more in terms of reactions and recommendations regarding the utility of leader assessment procedures and measurement tools in general rather than with an appreciation for the intended purposes of each.

**Theory.** In terms of the theoretical background driving the ARI work, it is fair to say that a wide spectrum of theories has been utilized, even if only in a post-hoc manner. However, Stratified Systems Theory (SST) is, perhaps, the most widely cited. As outlined earlier, SST suggests that different leader knowledges and personal orientations (i.e., proclivity) are important as individual progress through their careers and organizational hierarchies. This suggests that measures of different types of leader knowledge and personal characteristics must be articulated, defined, and assessed with context in mind. It also suggests that criteria indices of leader effectiveness must be chosen appropriately in order to test the validity of the theory. This places a premium on the kinds of measures included in this review.

**Existing measures.** Several promising ARI measurement tools do exist. In terms of personality assessments, specific facets of the SST proclivity theme have been identified and assessed (e.g., SOI, Biodata). However, it is also fair to say that the proclivity construct has not yet been fully articulated and thoroughly assessed by the efforts and measures that we reviewed. Moreover, the commercial benchmark measures that we reviewed have long track records of successfully assessing facets of the Big 5 personality framework. We would strongly encourage the incorporation of these types of assessments in efforts designed to examine the role that personality plays in leader effectiveness.

ARI assessment of leaders' knowledge shows some promise. Recall that we differentiated between general types of cognitive abilities, such as problem solving and information processing, and more specific types of knowledge, such as tacit or mental models. In terms of the general cognitive abilities, the ARI biodata measures yield several useful indices. As compared to the Fleishman and Quaintance (1984) taxonomy, the biodata indices still lack coverage of 35% of the areas. Accordingly, targeted development of additional subscales would be warranted if a complete sampling of the ability taxonomy is desired. Alternatively, commercial analogues exist that have proven histories of assessing these abilities that should be considered.

As for assessments of more focused types of knowledge, both the ARI tacit knowledge and mental model measures that have been developed show promise. These types of assessments require a substantial investment in the development stage because of two concerns. First, as compared to more generic approaches, these types of knowledges are more embedded in the specific job requirement and organizational settings. In other words, they are grounded more specifically in job conditions and, therefore, require development efforts that delve more deeply into job nuances. Second, there are no objective right-or-wrong answers to these types of assessments, so they require either reference against an "ideal response profile" derived from a consensus of experts, or must be evaluated individually by experts. Here, too, one must either devote a substantial amount of time initially to develop the expert template(s) or absorb the ongoing cost associated with ratings of responses. In any case, we should note that we believe that both the tacit knowledge and mental models measures developed by ARI have struck a nice balance in terms of grounding vs. generalizability. Both development efforts constructed multiple forms for use with leaders at different organizational levels. While falling short of the "core" dimensions theme with supplemental scales that we have advocated, this limited generalizability approach has enabled the researchers to both focus their assessment efforts while not overly confining the use of the measures.

The ARI assessments of leader behaviors (e.g., CPR, AZIMUTH) have been designed for limited applications. As we discussed in Section 5, we believe that the framework or infrastructure for gathering 360-type ratings of leader behaviors could be developed in a fairly generic fashion, allowing for more customized applications in terms of what dimensions are evaluated, by whom, for what purpose(s), in any given application. Whereas the MLQ instrument affords widespread comparability across

settings, it is not designed to hone in on specific requirements of Army leadership positions nor to direct developmental feedback efforts. It (or comparable assessments) is useful for research purposes and for making comparisons across settings, hierarchical levels, etc., but that comparability comes at the expense of applicability to any given circumstance.

**Research protocols.** We found that most ARI efforts followed a common research approach. First, most started with a good foundation in theory and a description of the larger framework within which the specific effort was targeted. Then, whether it was a prediction or assessment effort, some attention was devoted to identifying the underlying dimensions of leadership to be focused upon. Next, a large number of potential items, observations, etc. (i.e., indicators) of the relevant domain were generated and distilled. Herein lies a weakness of the prototypic method. There was typically a disconnect between the a priori specification of intended underlying dimensions, the indicator generation, and the indicator confirmation. The modal strategy appears to be to generate a large number of potential indicators and then to employ both judgmental techniques and exploratory quantitative data reduction analyses to "reveal" underlying dimensions. In contrast, an a priori approach would first specify the intended dimensions and then generate indicators of those specific dimensions. Next, depending on the number and potential redundancy of indicators, expert judgments could be solicited to combine, refine, and focus the preliminary set of items as related to their intended underlying dimensions. Finally, data can be collected from a preliminary sample that represents the intended boundaries of generalizability for use of the assessment device. Confirmatory analytic techniques can then be applied to test the extent to which the indicators map to their underlying dimensions. No doubt some revision will be necessary, and the stability of the resulting structure can be evaluated using additional developmental samples.

In fairness to the ARI researchers, we believe that they often try to accomplish "too much" in any particular study. That is, there is often an attempt to develop or refine measures while addressing more substantive relations with other variables of interest. While laudable, this dual focus tends to detract from both aims. The inclination is to "shotgun" the measurement effort in order to ensure that adequate coverage of the domain will be achieved. But this approach, combined with the use of exploratory data reduction techniques, yields instruments that are not comparable from one study to the next and limits the evolution of knowledge. Now we fully recognize that different research questions, field applications, and so forth, imposed demands on every research investigation. What we advocate, however, is the development of more standardized assessments that can be used intact in a number of different investigations. To achieve this, we recommend the following. First, a theory or common framework for conceptualizing the antecedents of leader effectiveness needs to be adopted. This is not to say that every study needs to subscribe to a particular theoretical position, but it would hasten the evolution of knowledge if all ARI studies of leadership could at least be described in terms of how they represent certain facets of a given theory. While, naturally, the theory that researchers believe best fits the U.S. Army of the 21st Century is

the best candidate for this function, what is more important is the some common yardstick be adopted. Such a theory would be grounded in reality.

Second, an updated job analysis of Army leadership positions is warranted for the identification of dimensions that are common across positions and those that have more limited representation. Third, an analysis of the important knowledge, skills, abilities, and other attributes important for performing those dimensions should be conducted. Fourth, criteria measures of effective performance of those dimensions should be developed. Given the multiple uses of feedback, a 360 rating framework focused on leader behaviors would likely pay high dividends here. However, other indices of effectiveness should also be considered and incorporated (see below). Fifth, there is a need to move beyond exploratory data analytic methods to more confirmatory techniques. Perhaps the biggest advantage of doing so lies not so much in the statistical tests and model fit indices as it does in the demands it places on investigators. These analyses require that researchers formulate an a priori framework for the measures they are testing. Sixth, additional explanatory variables should be incorporated to identify the limits of generalizability and potential moderators of relations.

The recommendations in the paragraphs above are not new, grand insights nor are they revolutionary. Rather, they hearken to a call for getting back to the basics before moving forward. Research scientists are intrinsically and extrinsically rewarded for developing new measures, testing new or innovative ideas, and essentially for moving forward into uncharted territory. However, if each study in a program of research introduces a new twist or "refinement" of an assessment technique, then progress is actually stunted, not enhanced. As we have mentioned repeatedly, if attention were devoted to establishing measures of core dimensions of Army leadership (whether those be predictors or assessments), along with more specific dimensions for given applications, in the aggregate, ARI research would be facilitated as each new study would have a better foundation from which to begin. This approach would, then, free resources for expanded inquiries incorporating other factors.

**Expanding the framework.** Our review of the ARI literature from the past 10 years revealed that most work focused on leader KSAOs and behaviors. Only a few studies addressed other influences shown, such as the task and operational environments, follower characteristics, or effectiveness (i.e . outcome) measures. Tenets of SST suggest that different variables will be important for leader effectiveness depending on the leaders' career stages and level in the organization. Beyond that focus, however, very few studies have considered situational influences on leader effectiveness. Moreover, follower characteristics have been virtually ignored. Clearly the Army of the 21st Century will differ from what we have seen in the past. The shear number of troops and officers will diminish, yet the demands on them will increase. While the number of men and women serving will decrease, their average abilities and expectations will surely go up as compared to previous generations. Technological sophistication has changed, and will continue to change, how battles are fought in the future. While some features of effective leadership are timeless, such as the ability to inspire and motivate troops, history has demonstrated that technology changes the nature of warfare and what makes

for effective leadership. These factors warrant far more attention as ARI works to understand and enhance leadership in the Army of the future.

There is also a serious need to develop the criteria side of ARI research investigations. Far too many of the leader assessment studies "validated" some measure of, for example, leader knowledge, by correlating scores on it with participants' responses on a different type of test (e.g., a situational exercise). Whereas such studies do provide evidence of construct validity for the measure in question, they do not substitute for criterion-related validity coefficients. Furthermore, when actual criteria measures have been employed, they have been limited to ratings of leaders' behaviors. As illustrated in the report, a vast number of effectiveness criteria, such as unit performance and subordinates' reactions, have yet to be incorporated. We caution to add that using some of these indices, such as unit performance, may impose limits on the research designs that can be employed and the applicable generalizations, but they better approximate ultimate criteria and are of great interest to line units.

**Army HR practice and leadership research.** In times of diminishing budgets and demands to do more with less, it is important leverage leadership research with ongoing human resource (HR) programs in the Army. This alignment should highlight two factors. First, it is widely accepted that different leader attributes are important at different career stages and hierarchical levels. ARI research that samples across these stages can inform practice as to what specific features are most critical at which times. In terms of the research implications of this approach, it also suggests that some variables are rendered moot for some purposes. For example, Zaccaro's (1996) summary of SST theory suggest that acute cognitive abilities skills are presumed to be possessed by all high-ranking officers such that what differentiates effective and ineffective executive leadership is attributable to other factors, such as proclivity. Note that this would suggest that indexing leaders' attributes, such as cognitive capacity, would be important if one was interested in predicting who would rise to senior officer levels, but would be far less informative if one were interested in predicting effectiveness among executive officers. Therefore, there is a natural synergy between what the focus of certain research investigations should be, given their purpose, and how they can inform practice in terms of providing developmental focus, critical feedback dimensions, and so forth.

The second theme linking ARI leadership research and practice involves the imbeddedness of investigations. Many of the efforts we reviewed had clear linkages with ongoing Army activities (e.g., the CPR, AZIMUTH, Special Forces, and Biodata). Embedding research investigations in ongoing activities always necessitates some compromises due to administrative demands and constraints and multiple data purposes. However, it also enhances the relevance of the research, both to the line units and to the participants. We see numerous benefits from making ongoing research investigations relevant to the units providing the data, including enhancing the ease with which it is collected and the quality of the resulting indices. Having said the above, we realize that many more basic research investigations simply cannot be woven into the fabric of ongoing activities, at least not in their developmental phases. We submit, however, that gaining access for these more basic and developmental activities will be easier in the

context of ongoing efforts that are valued by the line and training units. Such a demarcation of efforts would also clarify the value of different studies for the Army units.

A third theme to be pursued relates to the dynamics of leader effectiveness and the developmental processes that support it. In particular, more insight regarding how individual difference factors are changed or improved as a result of training or experience and specific career path sequences is needed. An emphasis on dynamics would also reveal the impact of context and job assignment on the shifting utility of input or process factors. For example, the leader behavior pattern required and the skills needed to display them may be more or less predictable, depending on the individual difference factor selected, with general cognitive ability being more transitional, but job knowledge context bound.

In summary, this report has chronicled a great deal of ARI leadership assessment work from the past 10 years. Much has been developed and learned. We suggest, however, that ARI is at a critical juncture and should pause to consider its strategic directions for future leadership research. In one sense, we advocate a more limited focus and integrated "back to the basics" emphasis. On the other hand, we encourage an expansion to consider a wider array of variables, such as situational and follower attributes, that moderate the effectiveness of leader behaviors in different circumstances. We also recommend greater embedding on research activities in ongoing Army activities and a cross-fertilization between research and practice.

# Table of Contents

| Section | Page Number |
|---|---|

| Section | Page Number |
|---|---|

# List of Tables and Figures

# Section I: Introduction

## Project Overview

The study of leadership dates back to at least antiquity. More recently, however, systematic research on the predictors of, processes related to, and consequences of, leadership has been conducted. In particular, research on what makes for effective leadership has been a focus of attention in the United States Army since the two world wars earlier in this century. Paralleling the larger arena of leadership research, work in the U.S. Military, and the Army in particular, has investigated individual traits and behaviors, situational and follower moderators, and a host of other variables related to leader effectiveness. Prior to the 1980s, much of the military research focused on generic dimensions of leadership, with most attention being devoted to the lower grade levels (Zaccaro, 1996). This focus expanded in the early 1980s to include research on the nature of leadership at higher grades. Particular interests included leader performance requirements, requisite skills, and developmental interventions targeting these executive leadership skills (Zaccaro, 1996).

Currently there exists a vast array of research on variables related to Army leader effectiveness. Concomitant with this, however, there has been a proliferation of measures designed to predict and/or assess leader effectiveness. The present project grew out of a need for a cataloging, synthesis, and review of such measures developed and/or used by the U.S Army Research Institute for Behavioral and Social Sciences (ARI) over the past ten years. Accordingly, the purpose of this report is to review featured ARI leadership measurement initiatives and compare them to benchmarks in nonmilitary research. The objectives of the effort were to: a) identify and describe major themes and initiatives by ARI leadership labs over the past ten years; b) critically analyze resulting instruments according to specific and common evaluative criteria; c) compare ARI initiatives against external benchmarks; and d) to offer suggestions and guidance for future leadership research endeavors.

This report examines measures employed in ARI leadership research over the period of 1987-1997. Due to the sheer number of constructs and variables researched over those ten years, the focus of our review needed to be narrowed to become manageable. For example, our initial review of general ARI leadership research included

34 technical reports, 17 research notes, 29 research reports and briefing slides, and 21 other miscellaneous documents from the ARI archives. Focusing on the specific leadership labs, 283 variables were identified from the initial briefing held in September 1996 and follow-up documentation on research initiatives. We categorized these variables into a database on the basis of with 11 features related to each variable. Examples of the descriptive data points for each variable included: a) psychometric characteristics; b) the projects that include each variable; c) the purpose of the project; d) the target and sample populations; e) stage of instrument development; and f) and potential uses of the instrument, such as evaluation or prediction. As a result, there are 3,679 database entries containing information on leadership variables examined by ARI labs. Further details regarding this database are available in an accompanying report (i.e., Marsh, Rouse, Mathieu & Klimoski, 1997).

These numbers support our need to narrow the focus of this evaluation project. In order accomplish this, only the most prominent and productive themes and initiatives were included. We used four primary means for narrowing our focus. First, we met with all ARI Research Scientists and asked them to nominate which of their projects they considered to be most central to the purpose(s) of our project. Second, together with the lab directors we considered the relative time and attention devoted to various projects and highlighted those that had garnered the greatest emphasis. Third, we considered the quantity and quality of documentation available. Finally, we considered the applicability of the initiatives to the larger area of leader effectiveness.

Once the major themes were identified it was necessary to establish an evaluation template against which they could be gauged. Six criteria were adopted to describe and evaluate the instrument development and use, and are listed in Table 1. First, brief descriptive information is presented, such as the purpose of the measure, the target population, scales, authors, publishers, etc. Second, the development and theoretical grounding of the measure are identified, followed by the frequency and nature of use. The psychometric characteristics of the instrument are reviewed as related to reliability and validity. Reliability indices include internal consistency estimates, test-retest

Table 1

Evaluation Criteria Used for Featured Instruments and Benchmarks.

1) Theory

2) Descriptive Information

3) Development & Empirical Use

4) Psychometrics

5) Generalizability

6) Face Validity, Ease of Use and Transparency

reliabilities, and some interrater reliabilities. Validity indices include construct and content, as well as predictive and concurrent studies. The fifth criterion is the generalizability of the instrument. This identifies the various contexts in which the instrument has been used and might be used. The final criterion deals with the specific use of the instrument and how it "looks." We made judgments regarding the face validity of the items, the ease of use, and the apparent transparency of the measures based on past literature and our direct examination. Both the featured ARI products and benchmarks are evaluated using these criteria permitting direct comparisons.

Once the ARI themes were identified, it was necessary to identify external benchmark measures for comparison purposes. Benchmarking essentially describes a practice of comparing research or systems of interest against similar types from outside of the immediate context. Benchmarking allows a comparison of issues such as content, cost, methods of administration, and effectiveness. These external benchmarks were obtained through extensive literature searches, electronic searches, electronic bulletin boards, web pages, and contact with external research organizations. Comparisons between the ARI work and these benchmarks, along with reference to the larger leadership research domain, drive the conclusions and recommendations offered at the end of the report. We should emphasize, however, that the benchmarks included here are representative of alternatives that are available in the literature and not necessarily exemplary measures. Indeed, as will become clear, in many instances there does not yet exist a measure that could be considered exemplary.

**Report Organization**

One cannot conduct any type of organizational assessment in the absence of some theoretical or organizational framework (Hausser, 1980; Mathieu & Day, 1997). Accordingly, below we provide a very brief overview of leadership theories with a particular emphasis on one, stratified systems theory, that appears to have guided much of the recent ARI leadership research. Next, we outline an integrative framework for leadership research. This framework is not designed or offered as a sine qu non, or "the" view of leadership; rather it is presented simply as one way to organize an abundance of research and to provide placeholders for later discussion. After we establish this foundation, we review the more specific themes identified from the ARI work.

## A Brief Review of Leadership Theory

There are many theories that have been proposed over the years to explain leadership and what makes a leader effective. Early in the history of research, attention was directed toward identifying the most valuable set of leader traits and skills related to leader effectiveness (Yukl, 1994; Yukl & Van Fleet, 1992). However, researchers were unable to agree on any one set of traits as being most necessary for effective leadership (Yukl, 1994). Leaders with different traits could feasibly be effective in the same situation. In addition, an individual leader who possessed certain traits that made him or her effective in one situation did not necessarily guarantee success in other situations.

Following the search for effective leadership traits was an emphasis on leader behaviors. In other words, research changed its focus from one that sought to know "what is it about leaders that makes them effective" to one that asked "what is it that effective leaders do?" This line of inquiry was exemplified by the research conducted during the 1950s and 1960s at Ohio State University by Fleishman, Stogdill, and others (Yukl & Van Fleet, 1992).

More recently, these leader-centered theories have garnered renewed attention, although emphasizing a more limited domain of leader traits/behaviors, such as transformational vs. transactional behaviors. Both of these approaches focus on the effects that leaders' behaviors have on followers, but in different ways. Bass (1985) proposed that transformational leadership is comprised of four components: 1) charisma; 2) inspirational leadership; 3) individualized consideration; and 4) intellectual stimulation. The notion here is that leaders behave in such a way as to empower subordinates and to motivate them to realize their full potential. In contrast, transactional leadership suggests that leaders attempt to motivate employees by explicitly tying rewards and punishments to certain types of behaviors. Whereas the former approach is empowerment-based, the latter is an exchange-based approach. While much was learned by both the trait and behavior focused lines of research, soon it became evident that few (if any) universally effective traits or behaviors exist. In other words, it became clear that characteristics of the situation and subordinates would dictate, in part, what constituted most effective leadership. Thus, contingency theories were born.

Contingency theories of leadership include the path-goal theory as articulated by Evans (1970) and by House (1971), leader substitutes theory (e.g., Kerr & Jermier, 1978), and Fielder's (1967, 1978) Least Preferred Co-worker (LPC) theory. The common theme running throughout this line of inquiry is that what makes for effective leadership depends on the aspects of the situation, subordinates, time pressures, resources, etc. Whereas these theories do offer a certain appeal and "middle ground" for leadership research, they are often stated and tested in so general a way as to provide little guidance to practicing managers. One of the more thoroughly developed and articulated contingency theories, however, is stratified systems theory.

Stratified Systems Theory. A fairly new theory that many ARI researchers subscribe to in their research is the stratified systems theory (SST) of leadership. This framework focuses on top executives instead of lower-level managers, although its basic premises apply across levels and career stages. The basic assumption behind SST is that the basis of effective leadership changes across career stages and hierarchical levels, and becomes more complex at higher ranks. For example, platoon leaders primarily focus on interpersonal issues (motivating subordinates and establishing personal credibility). Company commanders are concerned about issues of coordination, and the balancing of subordinate and institutional interests. At the battalion level, commander's tacit knowledge is primarily focused on the larger system (protecting the organization and managing organizational change).

Based on this assumption, Jaques (1976) formulated a theory that specifies parameters for vertical differentiation for hierarchical organizations. Using these specifications, Jacobs and Jaques (1987) began the task of specifying the leadership performance requirements of managers at these differentiated levels, with the ultimate objective of understanding how cognitive maps are developed. Leadership in this environment is the process of giving purpose or meaningful direction to collective effort, and causing willing effort to be expended to achieve a purpose (Jacobs & Jaques, 1990). There are three important elements to this definition of leadership. The first is the process of decision discretion. Leadership occurs when position incumbents are able to make choices or decisions. Based on this concept, leadership will in a large part reflect a cognitive or problem solving process, which becomes more complex across levels. The

second element is the effectiveness of the leader's direction-setting efforts of adaptiveness to the environment. The complexity of the organizational environment interaction will have short term and long term effects, so that organizational adaptation at the executive level requires more proactivity and planning within longer time frames. One of the most critical elements is the frame of reference or conceptual model for the collective action. This frame of reference provides the basis for a leader's understanding and interpretation of information and events encountered in the organizational operational environment (Jacobs & Lewis, 1992).

According to SST, there is an orderly progression of complexity from one level to the next higher level. This progression is marked by increasing time span, as well as by increasing complexity of the cognitive processes required of the incumbent. Given that the formal organizations to which SST applies are defined as accountability hierarchies, it follows that objectives generally are defined at the top most levels. In addition, performance required to achieve the objectives are given successively more concrete definition as one moves down the organizational hierarchy, until the level of direct output is reached (Jacobs & Jaques, 1987). The hierarchies in this theory are as follows:

**Stratum I** - this is the level of organizational production where employees work on the immediate process of operation with high levels of certainty. A central issue at this level is the pacing of work by the individual so that tasks can be completed.

**Stratum II** - this is the first level of management where their task is reducing uncertainty and defining the task for Stratum I. A central issue at this level is maintaining the output capability of the work group by balancing the requirements for individual development against the requirements for immediate production

**Stratum III** - this is the second level of management where the manager will have several subordinate section leaders or foremen. The work remains concrete in the sense that tasks are specifically given. However, the central issue at this level is balancing improvement of the system to meet current quotas or goals against the improvements deemed necessary to meet future predicted requirements.

**Organizational Domain**

**Stratum IV** - this is the first level of general management where the work is no longer concerned with concrete realities. A central concern of this level is the existing submits and submit processes for which the manager is accountable.

**Stratum V** - this is highest level of general management corresponding to the present of an organization. These managers are responsible for the adaptation of their organizations to the external environment.

**Systems Domain**

**Stratum VI** - this is the level at which managers validate the profit and loss objectives of subordinate companies within the context of the corporation as a whole and the environment.

**Stratum VII** - this is the level in which the CEO's primary responsibility is the development, construction, and fielding of new business units. The CEO opens avenues for Stratum VI managers, validates their judgment that their business units are viable, and creates opportunities for them by forming coalitions that result in the creation of resource bases necessary for the development or acquisition of new businesses (Jacobs & Jaques, 1987).

Overall, leadership skills are required at all levels of the hierarchy. However, each level requires a different array of knowledge, skills, abilities, and other characteristics. Lower levels are typified by direct leadership, which concerns the accomplishment of specific tasks, and direct interaction with the subordinates responsible for completing the tasks. The next higher level is characterized by organizational leadership. Here, leaders coordinate and facilitate the accomplishment of a broader range of specific tasks, and interact only indirectly with those responsible for carrying out the tasks. The highest level is executive leadership and is exemplified by leaders being concerned with establishing and communicating a broad vision, as well as setting a context within which meaning and direction are given to activities at lower levels (Laskey, Leddo, & Bresnick, 1990).

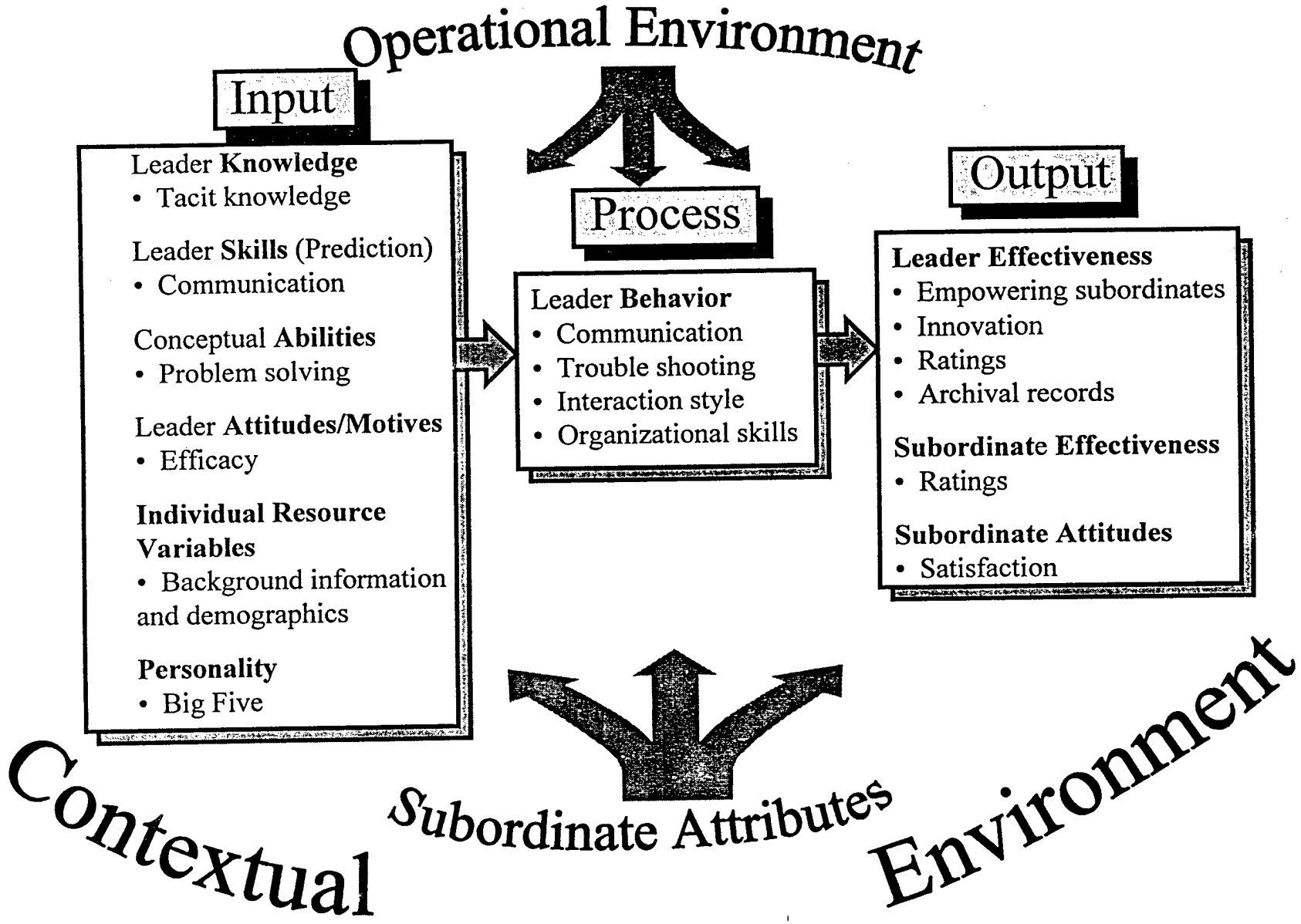
The progression from lower to higher levels of leadership is accompanied by several shifts in emphasis. The following are examples of these shifts (Jacobs & Jaques, 1987):

1) technical skills to abstract analytical to abstract integrative thinking skills;
2) from shorter to longer time horizons;
3) from direct to less direct forms of control;
4) from system component to system to multi-system perspective.

**An Integrative Framework**

Given the basic tenets of SST together with the vast array of leader assessment measures we encountered in our literature search of ARI documentation, it became clear that some sort of organizational scheme was necessary to place various research efforts in perspective. Following the logic of SST theory, leadership is not a unidimensional concept that can be assessed in a vacuum. Therefore, some organizational framework for reviewing leader assessment measures is warranted. For purposes of this report, therefore, we used an Input-Process-Output framework to organize the many dimensions of leadership. As shown in Figure 1, the input component includes **individual resource variables** (e.g., background information and demographics), leader **knowledge**, **skills**, **abilities**, as well as **other** constructs, such as attitudes, motivation, and personality (KSAOs). The process component covers leader **behaviors** such as their interaction, communication, and problem-solving styles. Finally, the output section deals with the **effectiveness of various leader actions** as indexed by evaluations of them, unit performance, and/or subordinates followers' behaviors and reactions. In addition, following the logic of contingency theories, variables that may moderate leadership processes such as **subordinates' attributes** and the immediate **operational environment** (i.e., task characteristics) are depicted. These are intended to illustrate the point that differing combinations are likely to be most effective in different circumstances. Finally, the reader should notice that the entire framework is nested within a larger **contextual environment**. This is intended to illustrate that what constitutes effective leadership will likely differ depending on the larger organizational culture and circumstances within which it is imbedded. For example, attributes of effective military leadership will likely differ depending on whether one is leading a drafted vs. volunteer corps, in peace time vs.

**Figure 1. Leadership Conceptualized as Input, Process, and Output Components**



peace-keeping vs. combat situations, the perceived moral imperatives related to actions, etc.

SST outlines how different types of leader knowledge are critical at different career stages/hierarchical levels. Similarly, the different levels place a premium on different facets of leaders' personalities. Moreover, what constitutes effective leader behaviors differs across hierarchical levels. Accordingly, these three facets became organizing mechanisms for our review and this report. In addition, however, we have featured biodata as unique measurement strategy. Biodata is really more of a description of a measurement approach that usually includes many substantive areas, including the three noted above. This comprehensiveness, however, is both an attribute and liability as it is difficult to neatly place constructs that biodata addresses into substantive areas. Because of this property, and the amount of attention devoted to biodata as a measure tool by ARI researchers, we have featured it in a separate section, but also discuss it in the more substantive sections as appropriate.

The remainder of this report is organized into four substantive sections, one covering each of the measurement themes noted above. The first part of each section introduces the main ARI initiatives, followed by a general literature review of the topical area. Important constructs will be defined and placed within the context of the theme. The next part of each section will contain more in depth information and an evaluation of the ARI featured measures in terms of the criteria identified above. Then, a parallel review of selected benchmark measures is provided. Each of the four sections then concludes with a critique of the ARI measures against the benchmarks, and summary statements are presented. At the conclusion of this report, overall conclusions and recommendations for organizing and directing future leadership research are offered.

**Personality** is the first substantive section covered in this report. ARI researchers have begun to focus on a new approach to personality called 'proclivity.' This approach to leadership will be described as will the following three measures ARI used to tap personality: 1) biodata; 2) the Subject-Object Interview (SOI); and 3) the Myers-Briggs Type Indicator. The general research community's approach to personality, known as the "Big 5" will be defined. This taxonomy clusters personality into five main areas: 1) extroversion; 2) neuroticism; 3) agreeableness; 4) conscientiousness; and 5) openness.

Three external personality inventories: 1) Hogan Personality Inventory (HPI); 2) NEO Personality Inventory (NEO - PI); and 3) California Personality Inventory will be presented as the benchmarks. Following the evaluation of all measures, the overlap between ARI measures of proclivity and standard personality inventories is explored and constructs are identified.

The second substantive section considers leader **knowledge**. Fleishman's taxonomy of cognitive abilities is used as an organizing framework for this section, which includes a very wide range of variables. This taxonomy is used to organize and compare the variables tapped by the featured ARI products and external benchmarks. The taxonomy is arranged into five higher-order cognitive abilities, yet still does not capture the range of knowledge-related assessments we encountered in our review of ARI work. Therefore, we have added a few additional entries including tacit knowledge and mental models. Five ARI assessment initiatives are reviewed in this section: 1) ARI Background Data; 2) ARI Critical Incidents; 3) Mental Models; 4) The Career Path Appreciation (CPA) protocol; and 5) Tacit Knowledge for Military Leadership Inventory (TKLMI). A corresponding range of external benchmark measures are featured in this section including: 1) the Watson-Glaser Critical Thinking Appraisal; 2) Concept Mastery Test; 3) Consequences; 4) a low-fidelity simulation by Motowidlo, Dunnette, and Carter (1990); 5) Leatherman Leadership Questionnnaire (LLQ); 6) PathFinder (PF) analyses of paired-comparison mental model ratings (Stout, Salas, & Kraiger, 1997); and 7) tacit knowledge (Wagner, 1987).

The third section of the report covers **biodata** in particular. Biodata measures tap a variety of constructs that fall under different themes. In the past, biodata relied on more experiential and behavioral information. Today, the focus has expanded to include personality, attitudes, and knowledge in this domain. Three ARI instruments are featured: 1) Civilian Supervisors; 2) Special Forces; and 3) Background Data Inventory (BDI). For comparison purposes, two benchmarks are included: 1) LIMRA's Assessment Inventory for Managers (AIM); and 2) Owens' Biographical Questionnaire (BQ).

The final substantive section of this report focuses on **leader behavior**. This theme falls under the process dimension of the IPO model. The methods of measuring leader behavior vary widely, and multiple constructs tend to be tapped. The featured ARI

products for this theme are: 1) Multifactor Leadership Questionnaire (MLQ); 2) Cadet Performance Report (CPR); and 3) Leader Azimuth Check/Strategic Leader Development Inventory (Azimuth/SLDI). These three measures will be compared to the following benchmarks: 1) Leader Behavior Description Questionnaire (LBDQ); 2) Leader Practice Inventory (LPI); 3) Benchmarks; 4) Campbell Leadership Index (CLI); 5) Profiler; and 6) Prospector.

# Section II: Personality

Based on a review of the ARI leadership projects, personality was one area that has received considerable attention. In fact, a scan of the database that we created revealed that over 32% of the variables categorized were related to personality. The most prominent personality attributes found were related to the concept of proclivity. The nature of the proclivity system highlights personality constructs that are thought to reflect temperamental characteristics that direct an individual's desire or inclination to engage in reflective thinking or cognitive model building. It also implies the degree to which an individual feels intrinsically rewarded by the cognitive activity of organizing complex experience. Leaders who are high in proclivity find mental effort intrinsically rewarding. In applied settings, it is personality attributes that are likely to influence leader performance by promoting a willingness and energy to solve problems in an ambiguous performance setting, providing the cognitive flexibility to acquire, encode, and manipulate information in such settings. In addition, personality allows a sense of individualism that is resilient in the face of uncertainty and potential failure (Mumford, Zaccaro, Harding, Fleishman, & Reiter-Palmon , 1991; Zaccaro, 1996). We should note, however, that the notion of proclivity advanced by SST is not entirely a personality construct, but also embodies some aspects of cognitive capacities and knowledge structures. For purposes of this Section, however, we will direct our attention to strictly the personality / orientation features of proclivity, and consider the more cognitive and knowledge facets in Section IV.

Proclivity has been assessed using three approaches in the ARI research that we reviewed: 1) the Subject–Object-Interview (SOI); 2) Biodata; and 3) the Myers Briggs Type Indicator (MBTI). These will be contrasted with a number of commercially available benchmark measures. The next section contains a literature review of the "Big 5," a personality taxonomy used outside of the military, and proclivity, the popular approach followed by many ARI researchers. Following this review, Table 2 is presented which provides a summary of the ARI measures and the three benchmarks used for comparison on the six evaluation criteria. The information in the table is expanded upon in the text for the six instruments. Next, our evaluation of ARI measures as compared to

the benchmarks, and recommendations for future research are offered. This section concludes with Table 3, which presents the overlap of the specific facets measured by all of the instruments for comparison purposes.

The study of personality has a long and tumultuous history. This may be due, in part, to a lack of agreement on a simple definition of personality. One reason for this is that there have always been a variety of competing systems that claimed to offer the best representation of personality structure (Cattell, 1946; Hogan, 1990). Thinking about personality went from an emphasis on the trait to the situation. Now, most investigators recognize an interaction approach that highlights the trait and the situation. Even with this development, the debate shifted to which taxonomy provided the most useful perspective. For example, Eysenck focused on specific traits of interest (e.g., extroversion), whereas, Cattell argued for the value of 16 factor scales. In a more contemporary treatment, Hogan (1990) advocated six dimensions. Presently, there seems to be general agreement among researchers concerning the number of dimensions of personality that might best summarize the available evidence into five factors. The development of the five-factor model is based on 50 years of factor analytic research on the structure of peer ratings. Even on this point, there is disagreement about the factors' precise meaning (Briggs, 1989; John, 1989; Livneh & Livneh, 1989).

During the past decade, literature has accumulated that provides evidence for the robustness of the five-factor model using different instruments, in different cultures, with different rating sources, and with a variety of samples (Bond, Nakazato, & Shiraishi, 1975; Costa Jr. & McCrae, 1987, 1988). The five-factor model provides parsimony in studying personality, as hundreds of individual scales may be pooled at a higher level. In addition, the model provides a framework for integrating the results of diverse research programs (Mount & Barrick, 1995). Consequently, we have employed the Big 5 as an organizing scheme for this section. This provides us with a common framework against which to gauge the various measures of personality that have been employeed both within and beyond the ARI research.

**Big Five System**

The first dimension of the Big Five is analogous to Eysenck's concept of Extroversion/Introversion. This construct implies sociability (preferring large groups and gatherings), being gregarious, assertive, talkative, and active (Hogan, 1990; John, 1989; McCrae & Costa, Jr., 1985; Smith, 1967). Those that score high on this factor tend to like excitement and stimulation, and generally have a cheerful, optimistic disposition. Those who score low are labeled introverts. Introverts may be seen as reserved, independent, even-paced, and even sluggish. They prefer to be alone, but they are not pessimistic or unhappy (Barrick & Mount, 1991; Costa, Jr. & McCrae, 1992). The specific facets measured by this dimension are: 1) activity/energetic; 2) assertiveness; 3) excitement seeking; 4) gregariousness; 5) positive emotions; and 6) warmth.

The second dimension is labeled variously Emotional Stability, Stability, Emotionality, or Neuroticism. This construct implies the tendency to experience negative affects such as fear, embarrassment, sadness, anger, and disgust (Hakel, 1974; John, 1989; McCrae & Costa, Jr., 1985; Smith, 1967). Individuals who score low on emotional stability are also prone to have irrational ideas and cope more poorly with stress. Individuals who score high are emotionally stable, usually calm, even tempered, relaxed, and may handle difficult and stressful situations better (Barrick & Mount, 1991; Costa, Jr. & McCrae, 1992). The specific facets encompassed by this construct are: 1) angry hostility; 2) anxiety; 3) depression; 4) discretion; 5) ego control; 6) emotional control; 7) impulsiveness; 8) self consciousness; and 9) vulnerability.

The third dimension has generally been interpreted as Agreeableness or Likability. This dimension is primarily a dimension of interpersonal tendencies. Traits associated with a high score on this dimension include being sympathetic, courteous, flexible, trusting, good-natured, cooperative, forgiving, soft-hearted, and tolerant (Hakel, 1974; McCrae & Costa Jr., 1985; John, 1989; Hogan, 1990). Lower scorers tend to be disagreeable or antagonistic persons who are ready to fight for their own interests (Barrick & Mount, 1991; Costa Jr., & McCrae, 1992). The specific facets included in this dimension are: 1) altruism; 2) caring; 3) cheerful; 4) compliance; 5) cooperative/competitive; 6) flexible; 7) good-natured; 8) modesty; 9) not jealous; 10) straightforwardness; 11) tender mindedness; and 12) trust.

The fourth dimension has most frequently been called Conscientiousness or Conscience, Conformity, and Dependability. There is some disagreement in terms of the essence of this dimension. Some say it reflects dependability, carefulness, responsibility, organization, and planfulness (Hakel, 1974; Hogan, 1990; John, 1989; McCrae & Costa, Jr., 1985). Others say it also incorporates hardworking, achievement-orientation, and persevering (Digman, 1990). A conscientious person is purposeful, strong-willed, and determined. Those who score low tend to be less exacting in applying themselves and working toward their goals (Barrick & Mount, 1991; Costa, Jr. & McCrae, 1992). The specific facets implied by this dimension are: 1) achievement striving/oriented; 2) cautious; 3) competence; 4) deliberation/planful; 5) dutifulness; 6) orderly; 7) responsible; and 8) self-discipline. In terms of the impact of personality on performance in applied settings, conscientiousness has sometimes been referred to as the "Big 1." In other words, whereas the role of various personality attributes on job performance tends to depend on aspects of the situation, conscientiousness has exhibited a more universal linear positive influence across situations. Therefore, this would naturally be a candidiate for inclusion in any system seeking to link personality with job performance.

The last dimension has been the most difficult to conceptualize. It has most frequently been identified as Intellect or Intelligence (Borgatta, 1964; John, 1989; Hogan, 1990). It has also been labeled Openness to Experience or Culture (Hakel, 1974; McCrae & Costa, Jr., 1985). Traits commonly associated with this are being imaginative, cultured, curious, original, broad-minded, intelligent, and artistically sensitive. Open individuals are: 1) curious about both their inner and outer worlds; 2) are willing to entertain novel ideas; 3) engage in more divergent thinking; and 4) experience both positive and negative emotions stronger than closed individuals. Those who score low on measures of openness tend to be more conservative, and prefer familiar rather than new stimuli (Barrick & Mount, 1991; Costa Jr., & McCrae, 1992). The specific facets covered by this dimension are: 1) actions; 2) aesthetics/artistically sensitive; 3) curious; 4) fantasy; 5) feelings; 6) ideas/original; 7) independent; 8) intellectual; 9) imaginative; and 10) values.

There are several other specific personality concepts that are often considered important, but do not easily fit into the "Big Five" taxonomy. These appear to reflect

aspects of orientation (e.g., dominance). However, the five constructs reviewed above would seem to be the most efficient way to organize our thinking about personality and the mainstream research in this domain.

**Proclivity**

As mentioned above, the number of potential constructs in the personality domain is overwhelming, and not always relevant to every situation. Army research on leadership appears to involve efforts at capturing what has been termed proclivity. As mentioned before, proclivity is thought of as reflecting the temperamental characteristics that direct an individual's desire or inclination to engage in reflective thinking or cognitive model building (Mumford et al., 1991; Zaccaro, 1996).

There are many different components that make up a proclivity profile. These are grouped into three main themes with specific personality variables under each one. The first theme is adaptability/ego resilience, the second is openness/curiosity, and the third is self-awareness.

The adaptability or ego resistant component is comprised of characteristics that foster motivation to work hard in uncertain, difficult/variable performance settings. Adaptable individuals exhibit resilience in the face of risk, uncertainty, and potential failure. Overall, it implies the degree to which a person appears calm, self-critical, and self-reflective. This factor is composed of six facets of personality. The first three, emotional control, risk taking, and self esteem represent a sense of ego strength and self-assurance that allows the leader to take chances in solving organizational problems, while having the confidence to perform in sometimes difficult interpersonal or social situations. A fourth facet is performance motivation, which reflects the disposition to work hard, persist, and adapt to changing environmental factors. The fifth facet is emotional control, which is similar to the variable under neuroticism in the Big Five. The sixth is energy level, which reflects the performance motivation or activity level of an individual (Mumford et al., 1991; Zaccaro, 1996).

The second sub-component of proclivity is curiosity, openness, or curiosity. It is comprised of seven facets. The first, intellect, is the cognitive and interpersonal style that causes people to be perceived as bright. This facet measures the degree to which a person is perceived as bright, creative, and interested in intellectual matters. A second

facet is openness to experience, often called the "ideas facet." In this facet, the individual actively pursues intellectual interests for their own sake, with open-mindedness and willingness to consider new and perhaps unconventional ideas. A third facet is cognitive complexity, followed by the facets of thinking, flexibility, tolerance for ambiguity, and investigation or curiosity (Mumford et al., 1991; Zaccaro, 1996).

The self-awareness component is the third theme in proclivity. This is defined as being able to promote problem solutions with little or no initial social support. One facet of this dimension is discretion or ego control, which reflects a self-concept of independence in the problem-solving process. Leaders possessing this personality attribute can make decisions when initial social support is lacking, and evaluate themselves in relation to established plans and goals. A second facet is internal locus of control, which is a person's tendency to take full responsibility for his or her achievement outcomes and to believe one's "life chances" are under personal control. A third facet is the tolerance for failure, defined as a sense of resiliency and encouragement after the occurrence of failure. The fourth facet is the ability of self-appraisal (Mumford et al., 1991; Zaccaro, 1996).

In summary, the notion of proclivity is fairly broad and encompasses many facets of personality identified in other research. By way of comparison, the proclivity dimension of Curiosity parallels the Openness to Experience dimension of the Big 5. Direct parallels, however, end there. The remaining two dimensions of proclivity, adaptability and self-awareness, contain elements of the Big 5 dimensions of conscientiousness and agreeableness. However, the proclivity subdimensions are not nearly as coherent, robust, or empirically validated, as are the Big 5 themes.

The next feature this section contains is the evaluation of ARI's instruments and the benchmarks. First a summary table outlining the evaluation is presented (see Table 2). Following this is the in depth review of the instruments on the six criteria identified in the introduction. This section concludes with our evaluation and recommendations of the ARI instruments, along with a table containing a breakdown of the specific variables assessed by each instrument (see Table 3).

Table 2
ARI Personality Measures and Benchmarks by Criteria.

| Criteria | Subject-Object Interview | Biodata | Myers-Briggs Type Indicator | Hogan Personality Inventory | NEO-Personality Inventory | California Personality Inventory |
|---|---|---|---|---|---|---|
| Theory | stratified systems theory | past behavior/ experiences predicts future behavior | Jung's Theory of Psychological Types | Socioanalytic Theory | Big 5 | 20 folk concept scales |
| Descriptive Information | three tasks: phrase selection (cards), symbol sorting (cards), work history (interview) | 467 items with 21 scales cognitive, self-confidence, motivational, management skills, and social skills variables | identify psychological types forced choice multiple forms | assess observable personality 7 primary scales 206 items | assess observable personality 120 items (short form) 5 dimensions 30 facet scales 5 point scale | 20 scales 4 categories 462 items |
| Development | strong | strong | weak | strong | moderate | moderate |
| Empirical Use | moderate use | widespread use in military | widespread use, few constant relationships found | widespread use | widespread use | widespread use |
| Psychometrics | high reliability strong construct validity moderate criterion-related validity | mixed reliability strong construct validity strong criterion-related validity | moderate to high reliability low to moderate criterion-related validity poor construct validity | moderate to high reliability moderate criterion-related validity strong construct validity | high reliability strong construct validity | moderate to high reliability strong construct validity |

| Criteria | Subject-Object Interview | Biodata | Myers-Briggs Type Indicator | Hogan Personality Inventory | NEO-Personality Inventory | California Personality Inventory |
|---|---|---|---|---|---|---|
| Generalizability | high | high(civilian army) | high | moderate | high | high |
| Face Validity | moderate | moderate | low | low to moderate | low to moderate | low |
| Ease of Use | low | high | moderate | moderate | high | high |
| Transparency | low | moderate | low | moderate | moderate | moderate |

## Subject-Object Interview

| | |
|---|---|
| Purpose | Assess specific constructs of personality |
| Population | USMA cadets and TACs |
| Acronym | SOI |
| Scores | 1) Extroversion; 2) Achievement; 3) Cooperation; 4) Imaginative; 5) Sensing; and 6) Locus of Control (LOC) |
| Administration | Individual Interview |
| Price | N/A |
| Time | 90 minutes |
| Authors | Kegan (1982) Center for Leadership Research (1996; briefing slide) |
| Publishers | ARI |

### Theory

This instrument is based on Jaques' (1975) SST. This theory postulates that the core of the psychological experience of doing work is "the exercise of discretion" (Stamp, 1988). This exercise of discretion is concerned with choices that must be made and the psychological processes of choosing an action. This exercise is seen to be that of imagination, formulation, and execution of a course of action which is not prescribed (Stamp, 1988). One of the characteristics of discretion is the extent to which an individual is capable of making a choice and following it through.

### Development and Empirical Use

This interview protocol is used at the USMA and was developed by the research team there. The protocol is as follows:

Part 1: The interviewee is handed ten cards and is requested to record memory joggers about events tied to the subject. The ten subjects are: 1) angry; 2) anxious/nervous; 3) success; 4) strong stand/conviction; 5) sad; 6) torn; 7) moved/touched; 8) lost something; 9) change; and 10) important.

Part 2: The interviewer spends about one hour with the interviewee discussing the experiences he or she recorded on the cards. The interviewee is allowed to pick the cards he or she wishes to discuss, and it is not necessary to get through all the cards. The interviewer asks follow-up questions, and probes to elicit further information.

At the conclusion of the interview, the trained interviewer extrapolates personality themes from the situations given by the interviewee, and additional information is collected from the follow-up questions.

**Psychometrics**

Psychometric information is in progress for this measure. Construct validity is strong due to the strong development of the SST. The interrater reliability has been high when highly skilled, trained coders were used to evaluate tapes.

**Generalizability**

The protocol of the measure would seem to generalize to a variety settings. However, the complexity of scoring and expertise required of the interviewer may limit its use.

**Face Validity/Ease of Use/Transparency**

The interview protocol is time-consuming and difficult to score. The interview process is also not seen as very face valid, because individuals are asked to describe a situation. Then, the interviewer determines what the experience means in terms of personality. The cards, follow-up questions, and probes do not appear to be transparent. The use of this assessment approach is, however, quite intensive both in terms of contact hours and data acquisition. Highly skilled interviewers are needed to conduct the interviews, and extensively trained coders are needed for reviewing audio or video-tapes. In other words, each 90 minute individual interview requires at least five hours of effort on the part of skilled professionals to yield quantifiable indices.

## ARI Civilian Supervisor/Special Forces Biodata

| | |
|---|---|
| Purpose | Assess personality constructs (along with other factors) |
| Population | Special Forces, Civilian Supervisors |
| Acronym | N/A |
| Scores | 1) Dominance; 2) Achievement; 3) Energy level; 4) Consideration; 5) Stress tolerance; 6) Dependability; 7) Flexibility/adaptability; 8) Agreeableness; 9) Cooperation; 10) Openness; 11) Extroversion; 12) Ego control; 13) Emotional control; 14) Conscientiousness; 15) Locus of control (LOC) |
| Administration | Paper and pencil, individual |
| Price | N/A |
| Time | 20-40 minutes |
| Authors | Kilcullen, White, Mumford, & O'Connor (1995) |
| Publishers | ARI |
| Comment | Personality items are a small part of the entire biodata instruments featured in the biodata section of the report |

### Theory

Biodata can best be described as past behaviors and experiences that predict future behavior and experiences. Learning, heredity, and environment together make the exhibition of certain behaviors more prevalent (Mumford & Stokes, 1992). Biodata items are designed to tap the developmental history of individuals in terms of typical interactions with the environment (Mumford & Stokes, 1992). There is some overlap between items tapped by biodata items and standard personality inventories. Zaccaro, White, Kilcullen, Parker, Williams, and O'Connor-Boes (1997) identified the personality relevant variables tapped by this ARI biodata instrument as follows: 1) dominance; 2) achievement; 3) energy level; 4) consideration; 5) stress tolerance; 6) dependability; 7) flexibility/adaptability; 8) agreeableness; 9) cooperation; 10) openness; 11) extroversion; 12) ego control; 13) emotional control; 14) conscientiousness; 15) locus of control.

### Development and Empirical Use

***The following review describes the development of the biodata instrument as a whole. We should note that not all facets of the instrument were intended to assess personality type variables.***

This instrument was developed using a variety of sample populations. Over 2000 first-line supervisors from a variety of occupations and grade levels, as well as Special Forces, Army War College participants, and Rangers were sampled. As a result, several versions were developed based on the Mumford, O'Connor, Clifton, Connelly, and Zaccaro (1993) model, containing different combinations of scales contingent on the population.

Civilian Supervisor Version. Rational scales were developed to measure 21 individual characteristics. A panel of psychologists reviewed the construct definitions, and each member generated 10-15 items related to past behaviors and life events for each construct. Next, these items were examined by the panel based on the following criteria: 1) construct relevance; 2) response variability; 3) relevance to Army civilian population; 4) readability; 5) non-intrusiveness; and 6) neutral social desirability. From the pool of items, 20-40 of the most representative ones for each construct were chosen and responses were weighted according to their relationship with the intended construct. A second panel of psychologists then reviewed this set of items, and a pilot test was conducted. Revisions were made based on the item analysis of the pilot data. The final version of the instrument contained 467 items.

Special Forces Version. A job analysis was conducted to determine the performance dimensions for SF. It identified 47 attributes relevant to successful performance in SF jobs, and 26 critical incident-based categories. SMEs rated attributes in terms of their importance to the job. The most highly rated attributes were: 1) teamwork and interpersonal skills; 2) adaptability; 3) physical endurance and fitness; 4) strong cognitive abilities; 5) strong leadership and communication skills; and 6) strong judgment and decision making skills. Based on the job analysis, a biographical questionnaire was developed to measure the SF traits.

The questionnaire consisted of 178 items ranging from social intelligence items to physical capability items. The questionnaire was completed by 1,357 soldiers participating in SF Selection and Assessment processes, as well as 293 SF officers. The items were then analyzed and scales were created by: 1) analyzing the internal reliabilities of different groups of items in terms of inter-item correlations, inter-total correlations, squared multiple correlations and the scale alphas when the item is removed (empirical); and 2) reading each item and determining the best scale for the item through content analysis (rational).

Along with the concurrent validation efforts, predictive validity of the questionnaire was tested with a SF Assessment Schools (SFAS) sample. The primary criteria were voluntary withdrawal and graduation.

**Psychometrics**

Civilian Supervisor Version. Convergent validities with related temperament scales were .60 and higher. The alphas for the 21 scales ranged from .65 to .85 (mean = .76). A

blocked regression analysis was used to evaluate the Leader Effectiveness Model. The first block contained cognition, self-confidence, and motivation, which significantly predicted ratings and performance records (multiple Rs equaled .21 and .35, respectively). The second block was composed of management skills and social skills, which led to a significant increase in the $R^2$ for performance records.

Special Forces Version. The alphas for this version are reported in the table below. The specific personality alphas from the SFAS and SF samples are:

| Scale | α (SFAS) | α(SF) |
|---|---|---|
| Aggression | .55 | .62 |
| Social Intelligence | .86 | .84 |
| Autonomy | .72 | .76 |
| Cultural Adaptability | .49 | .68 |
| Work Motivation | .62 | .64 |
| Anxiety | .65 | .72 |
| Openness/Cognitive Flexibility | .78 | .72 |
| Outdoors Enjoyment | .78 | .82 |
| Cooperation | .48 | .34 |
| Average | .66 | .68 |

**Generalizability**

Generalization to other military samples is highly likely given the fact that a wide range of specialties was encompassed in the development of the items. Generalizablity to non-military samples is probably feasible given the fact that the measure appears to be fairly representative of the general leadership domain. Nevertheless, civilian parallels for military experiences would be necessary to identify and incorporate.

**Face Validity/Ease of Use/Transparency**

This measure is a paper and pencil instrument that is easy to score. The face validity and transparency vary depending on the specific items, but, in general, the overall instrument is moderate on both criteria in our opinion.

## Myers Briggs Type Indicator-ARI

| | |
|---|---|
| Purpose | Determine psychological types to identify managerial attributes, behaviors and effectiveness |
| Population | Leaders in organizations, government, military, students |
| Acronym | MBTI |
| Scores | 1) extroversion/introversion; 2) sensing perception/intuitive perception; 3) thinking/feeling; 4) judgment/perception |
| Administration | Paper and pencil, individual |
| Price | 10 Form G, 10 *Introduction to Type,* Manual - $102.60<br>50 Form G, 50 *Introduction to Type,* Manual - $404.10 |
| Time | 20-30 minutes |
| Authors | Myers & McCaulley (1985; manual) |
| Publishers | Consulting Psychologist Press |
| Comment | This summary will describe the commercial version, and then how it has been used by ARI. |

**Theory**

This instrument is based on Jung's (1971) theory of psychological types, which proposes that the consciousness differentiates the use of the following four mental processes: 1) assessment of reality; 2) vision of the future; 3) logical decision making; and 4) value-oriented decision making, as well as the attitudes in which these are used.

Katherine Briggs and Isabel Myers operationalized Jung's (1971) "type theory" and developed the MBTI. The main purpose was to identify an individual's type to determine different patterns of interest. These interests are then assumed to effect performance in different situations, depending on the demands of the situation.

Psychological type theory proposed that people have four pairs of processes, but postulated that one of each pair is preferred over the other. The first pair deals with an individual's preferred mode of perception. Individuals are either sensing, which means they focus on facts and details, and tend to be practical, or intuitive, meaning they follow hunches and speculations and tend to be future-oriented. The second pair deals with the mode of judgment. Thinking individuals are objective and logical, whereas feeling individuals are subjective and humane, or empathetic. The third pair deals with an individual's attitudes that reflect their orientation of energy. One choice, extroversion, reflects a focus on people and things, and being sociable, whereas introversion reflects a focus on thoughts and concepts, and is inwardly-directed. The final pair deals with an individual's orientation toward the "outerworld." An individual prefers to be either judging, where they are organized, planned,

and settled, or perceiving, where they are curious, flexible, and spontaneous (Gardner & Martinko, 1996).

The four sets of preferences combine to form 16 different personality types. For a complete description of the dimensions and interactions see Myers & Myers (1980). Generally, every individual uses all eight processes, but type theory postulates that one of each pair is preferred over the other.

**Development and Empirical Use**

Briggs and Myers studied Jung's theory and operationalized type. They constructed items tapping the different types, and tested them on acquaintances for 20 years. The final version of the MBTI was constructed in 1941. After a long incubation period in the 1940s and 50's, Educational Testing Service published the MBTI in 1962 as a research tool.

It has been estimated that over three million people complete the MBTI each year (Myers & McCaulley, 1985). However, few empirically consistent relationships have been found between type and managerial effectiveness. For a review of approximately 50 empirical studies, see Gardner and Martinko (1996).

**Psychometrics**

The split-half reliabilities consistently exceeded .75 for continuous scales (Carlyn, 1977). The split-half reliabilities exceeded .60 for dichotomous scales for Phi coefficients (McCarley & Carskadon, 1983). The coefficient alphas for the four scales ranged from .67 to .79. Test-retest correlations ranged from .64 to .90 (Nunally, 1978). Scale intercorrelations for EI, SN, and TF are relatively independent, but the JP scale is often significantly correlated with the SN scale, and occasionally with the TF scale (Gardner & Martinko, 1996).

In order to assess validity, a factor analysis of Form G, the standard form was performed. This analysis yielded clear, simple factors matching those from the proposed theoretical background. Type distribution tables supply extensive evidence of the criterion-related validity across occupations, which is consistent with the theory (Gardner & Martinko, 1996). In addition, significant correlations of the MBTI scales with various interests, personality, academic, and observational measures have been found. However, the key structural assumptions of type theory and the MBTI's operationalization of them remain largely unsubstantiated (Podsakoff & Organ, 1986). This has led to concern about the MBTI's

factorial structure, as well as construct and criterion-related validities (Sipps & Alexander, 1987; McCrae & Costa, Jr., 1989).

**Generalizability**

Application and use of the instrument are broad. There has been a wide range of samples used, such as U.S. leaders, foreign businesses and industry, local, state and federal governments, participants in programs at the Center for Creative Leadership, consultants, and student samples (Gardner & Martinko, 1996).

**Face Validity/Ease of Use/Transparency**

The instrument uses a self-report, paper and pencil format, making administration easy. The instrument is a forced-choice questionnaire. There are several forms available (F, G, G-self scorable, J, and K), with form G being the standard and most widely used form. This form is composed of 126 items.

In terms of transparency, the instrument may be tapping impression management behaviors rather than basic psychological preferences (Gardner & Martinko, 1996), which would lead to problems with validity.

**ARI Use of the MBTI**

In terms of ARI research, the MBTI has been used to identify individuals possessing the trait of proclivity. The type that characterizes this personality is the NT (intuitive thinking) profile. Individuals of this type prefer intuition as the mode of perception. They gather information primarily by associating new information and ideas with previously acquired information. They dislike structure, details, and routine, and enjoy new problems and situations. They also exhibit the conceptual ability to perceive environments as wholes, and problems or events as parts of wholes. This individual's preferred mode of judgment is thinking. Here, the individual prefers to evaluate information and make decisions on the basis of logic. Individuals possessing this profile tend to take a rational, systematic approach to problem solving. They also order people, situations, and information in a structured framework, without consideration for the feelings of others. They prefer objective data and generally use logical, impersonal, and theoretical analyses to explore possibilities inherent in a problem (Zaccaro, 1996).

Two recent studies have attempted to identify the specific NT type in Army leaders. The MBTI was administered to Colonels and Lt. Colonels at the Army War College (Barber,

1990) and at the Industrial College of the Armed Forces (Knowlton & McGee, 1994). Knowlton and McGee found 32% of their sample were NT's. The average percent of NT's from the general military population (i.e., a mix of officers and enlisted soldiers) was 15% (Briggs-Myers & McCaulley, 1985). These percentages suggested a trend that the relative proportion of NT's among system-level military leaders may be higher than the general population, which is consistent with Jacobs and Jaques' SST (Zaccaro, 1996).

# Benchmark Instruments

## Hogan Personality Inventory

| | |
|---|---|
| Purpose | Assess individual's observable personality for personnel selection purposes |
| Population | College students, organizational employees |
| Acronym | HPI |
| Scores | 1) Adjustment; 2) Ambition; 3) Sociability; 4) Likeability; 5) Prudence; 6) Intellect |
| Administration | Paper and pencil or computer, individual |
| Price | Varies based on purpose of use |
| Time | 20 minutes |
| Authors | Robert Hogan (1986) |
| Publishers | National Computer Systems |

### Theory

The instrument is based on socioanalytic theory, which states that people are motivated to engage in social interaction (Hogan, 1986). Socioanalytic theory assumes that people are motivated by acceptance/recognition by peers, and seeking status and power relative to peers. Over time, people develop identities, and these self-images guide behavior. A person's self-presentational behaviors develop from the identities, and then guide social interactions. Social interactions in this context can be defined as the giving and withholding of acceptance and status. Based on sociological theory, individuals are predisposed to evaluate others in terms of the degree to which they will be an asset or a liability to their families or social groups. These decisions are based on behavior that is observed. Therefore, measurement of personality should be based on observable behavior. The HPI was constructed for this purpose, assessing six broad dimensions of personality (Hogan, 1986). The definitions of these dimensions and the specific facets included in each one are presented below.

The first dimension is adjustment. Adjustment measures the degree to which a person appears calm and self-accepting, and conversely, self-critical and overly self-reflective. The specific facets included in this dimension are: 1) empathy; 2) anxious; 3) guilt levels; 4) calmness; 5) even-tempered; 6) trusting; and 7) good attachment. The second dimension, ambition, is defined as the degree to which a person is socially self-confident, competitive, and energetic. The facets falling under this dimension are: 1) competitiveness; 2) self

confidence; 3) depression; 4) leadership; 5) identity; and 6) no social anxiety. The third dimension, sociability, measures the degree to which a person seems to need and/or enjoy interactions with others. This energy orientation includes liking parties and crowds, seeking experiences, and entertaining. The fourth dimension, likeability, measures the degree to which a person is seen as perceptive, tactful, and socially sensitive. The five facets under this dimension are: 1) being easy to live with; 2) sensitive; 3) caring; 4) liking people; and 5) showing hostility. The fifth dimension, prudence, measures the degree to which a person is conscientious, conforming, and dependable. The facets tapped in this dimension are: 1) morals; 2) mastery; 3) virtuosity; 4) autonomy; 5) spontaneousness; 6) impulse control; and 7) avoiding trouble. The last dimension in the HPI is intellect. This component is defined as the degree to which a person is perceived as bright, creative, and interested in intellectual matters. The six facets tapped by intellect are: 1) science; 2) curiosity; 3) thrill-seeking; 4) intellectual games; 5) generating ideas; and 6) culture (Hogan, 1986).

**Development and Empirical Use**

The original model for the HPI was the folk concepts of the CPI (Gough, 1957). These folk concepts tapped aspects of social behavior for the purpose of assessing and predicting social outcomes (Gough, 1957). After the assessment of the folk concepts, the Big Five taxonomy was also examined for information. Using these two sources and the socioanalytic theory as a guide, items were written. For example, items based on the Big Five were constructed by taking each of the major dimensions of the five-factor model, and asking what sorts of self-presentational behaviors might lead to high or low standings on that dimension (Hogan, 1986). The items were refined, assessed for internal consistency, and 225 items were pilot tested on 11,000 people employed in organizations across the country. Over 540 validity studies were conducted within various organizations. In addition, matched sets of data were gathered from other tests, inventories, and observer descriptions (Hogan, 1986).

Using all of the archival data, a factor analysis with orthogonal varimax rotation was conducted. The six primary scales were extracted based on the size of the eigen values, a scree test (Cattell, 1966), and an examination of the comprehensiveness for each dimension (Hogan, 1986).

**Psychometrics**

The internal consistency (coefficient alphas) ranged from .29 to .89, based on a sample of 960 employed males and females. Test-retest correlations ranged from .34 to .86, based on 150 male and female university students (Hogan, 1986). Reliabilities for the specific facets showed 34 of the 41 facets having alphas greater than .50. A total of 36 of the 41 facets also displayed test-retest reliabilities above .50 (Hogan, 1986). Norms are available from the publisher.

Construct validation evidence was presented for the HPI in three ways. The first was to correlate the primary scales of the instrument with other validated tests. Correlations with the following psychological measures have been calculated: 1) cognitive tests - ASVAB (US Department of Defense); 2) motives and interest - MBTI, Self-Directed Search (Holland, 1985); and 3) normal personality - Big Five Markers (Goldberg, 1992), Interpersonal Adjective Scale (Wiggins, 1991) as cited in Hogan, 1986), and MMPI-2 (Hathaway & McKinley, 1943). Second, HPI measures were correlated with peer ratings. Significant correlations between peer descriptions and the HPI scores allowed for the evaluation of the validity of the measure and the socioanalytic theory. A total of 128 college students completed the HPI, and also gave the HPI forms to two people whom they knew for at least two years. Findings showed significant relationships across scales, with adjustment consistently having the lowest correlations, and conscientiousness having the highest correlations (Hogan, 1986). The third method of proving construct validity included correlating HPI scores with relevant measures of organizational performance. Some sources of organizational performance included supervisor ratings, reported stress, training performance and leadership, and upward mobility (Hogan, 1986).

Moderate concurrent validity was also shown for a sample of nurses. Analyses yielded a significant correlation of .61 with a service orientation scale (Hogan, Hogan, & Busch, 1984). Additional studies may be found in the HPI manual (see pp. 66-67).

**Generalizability**

Samples have included college students, and a wide range of organizations and business employees. This leads us to the conclusion that the instrument has moderate to high generalizability.

**Face Validity/Ease of Use/Transparency**

The HPI is a paper and pencil measure with 206 items, which is scored remotely. This is a commercial instrument available for a fee, which increases the administrative burden somewhat. Based on our review of the items, face validity is low to moderate, depending on the specific questions. The authors argued that transparency and faking are moot issues because the goal is to sample a person's typical self-presentational style (Hogan, 1986).

## NEO-Personality Inventory

| | |
|---|---|
| Purpose | Comprehensive assessment of adult personality |
| Population | College students, business settings, clinical and vocational settings |
| Acronym | NEO - PI |
| Scores | 1) extroversion; 2) neuroticism; 3) agreeableness; 4) conscientiousness; and 5) openness. |
| Administration | Paper and pencil or computer, individual |
| Price | Comprehensive kit: Manual, 10 reusable questionnaires for self (S), 10 reusable questionnaires for others (R), 5 male/5 female of each, 25 hand scorable sheets, 25 form S & R profile sheets, 25 feedback sheets $129.00 |
| Time | 30 to 40 minutes |
| Authors | Paul T. Costa Jr. & Robert R. McCrae (1992) |
| Publishers | Psychological Assessment Resources |

### Theory

The NEO-PI was developed to operationalize the five-factor taxonomy. As mentioned in the literature review, the five-factor model is a representation of the structure of traits building on the taxonomies of Eysenck, Guilford, Cattell, Buss and Plomin, as well as others. The five factors account for the major dimensions of personality. The factors are defined by groups of intercorrelated traits, referred to as facets. The facet scales offer a more fine-grained analysis of the specific traits. Each of the five factors is represented by at least six facets. This insures coverage of a wide range of thoughts, feelings, and actions. It also permits internal replication of findings and identifies meaningful within-domain variation for individuals (Costa, Jr. & McCrae, 1992). In order to determine the specific facets under each dimension, the developers, Costa Jr. & McCrae, worked top down from the five-factor model to include the various facet measures.

The five factors are as follows: 1) extroversion; 2) neuroticism; 3) agreeableness; 4) conscientiousness; and 5) openness. The first dimension, extroversion/introversion, measures sociability (preferring large groups and gatherings), being gregarious, assertive, talkative, and active. The specific facets measured by this dimensions are: 1) activity/energetic; 2) assertiveness; 3) excitement seeking; 4) gregariousness; 5) positive emotions; and 6) warmth (Costa, Jr., McCrae, & Holland, 1984). The second dimension, emotional stability, measures the tendency to experience negative affects, such as fear, embarrassment, sadness, and anger. The specific facets encompassed by this dimension are: 1) angry hostility; 2) anxiety; 3)

depression; 4) discretion; 5) ego control; 6) emotional control; 7) impulsiveness; 8) self-consciousness; and 9) vulnerability (Costa Jr. & McCrae, 1992). The third dimension has generally been interpreted as agreeableness. This dimension primarily taps interpersonal tendencies such as being sympathetic, courteous, and flexible. The specific facets measured by this dimension are: 1) altruism; 2) caring; 3) cheerful; 4) compliance; 5) cooperative/competitive; 6) flexible; 7) good-natured; 8) modesty; 9) not jealous; 10) straightforwardness; 11) tender mindedness; and 12) trust (Costa Jr. & McCrae, 1990). The fourth dimension, conscientiousness, reflects dependability, carefulness, responsibility, organization, and planfulness. The specific facets measured by this dimension are: 1) achievement striving/oriented; 2) cautious; 3) competence; 4) deliberation/planful; 5) dutifulness; 6) orderly; 7) responsible; and 8) self-discipline (McCrae, Costa Jr., & Busch, 1986). The last dimension is openness to experience. Open individuals are: 1) curious about both their inner and outer worlds; 2) willing to entertain novel ideas; 3) engage in more divergent thinking; and 4) experience both positive and negative emotions stronger than closed individuals. The specific facets measured by this dimension are: 1) actions; 2) aesthetics/artistically sensitive; 3) curious; 4) fantasy; 5) feelings; 6) ideas/original; 7) independent; 8) intellectual; 9) imaginative; and 10) values (Costa, Jr. & McCrae, 1985).

**Development and Empirical Use**

The development of the scale was guided by both rational and factor analytic strategies. The five-factor taxonomy guided the constructs, then items designed to tap the constructs were written and administered to two longitudinal samples. The first sample consisted of 2000 primarily white male participants of the Veterans Administration's Normative Aging study in Boston. Peers of participants were asked to rate participants. The second sample was over 1800 male and female employees. The results were factor analyzed, and items were selected on the basis of their factor loadings. Items for the scales were also balanced in terms of positively and negatively keyed responses (Costa, Jr. & McCrae, 1992).

**Psychometrics**

The average internal reliability across many samples of the scales are as follows: 1) neuroticism - .92 (self) and .93 (other); 2) extroversion - .89 (self) and .90 (other); 3) openness - .87 (self) and .89 (other); 4) agreeableness - .86 (self) and .95 (other); and 5) conscientiousness - .90 (self) and .92 (other) (Costa, Jr. & McCrae, 1992).

A factor analysis showed a strong five-factor structure, indicating evidence of convergent and divergent validity (Costa, Jr., McCrae, & Dye, 1991). Two recent studies of the entire thirty-scale instrument lent increasing support to this finding. One study from the longitudinal archives of the BLSA correlated the NEO scales with scales from twelve different inventories (McCrae & Costa Jr., 1987). Of the 150 correlations, 66 were greater than .50. The second study correlated the NEO facet scales with alternative measures of similar constructs, such as the NEO anxiety facet with the anxiety scale in the State Trait Personality Inventory (Spielberger, Jacobs, Crane, Russell, Westberry, Barker, Johnson, Knight, & Marks, 1979), finding strong results. Similar findings occurred when the NEO scales were correlated with neuroticism and extroversion scales and second level facets from Eysenck's Personality Inventory (Esyenck & Esyenck, 1964). Other significant correlations (Costa, Jr. & McCrae, 1992) were found with the Personality Research Form (Jackson, 1984) and the Adjective Check List (Gough & Heilburn, 1983).

**Generalizability**

This instrument has been used for clinical applications, vocational counseling, educational research, psychological research, and business settings. Therefore, it should be highly generalizable.

**Face Validity/Ease of Use/Transparency**

For normal populations, studies have shown no marked distortion on social desirability. The instrument is available in a paper and pencil version, self and other versions, as well as a computerized version. There are two forms, the NEO-PI-R which is 240 items, and the short form, the NEO-FFI, that contains 60 items. It is fairly short and easy to score by computer. In terms of transparency, some items are more transparent than others. The face validity of the instrument is acceptable in our opinion.

## California Psychological Inventory

| | |
|---|---|
| Purpose | Assess variables to understand, classify and predict behavior |
| Population | Ages 14 and up, students, organizational, military, government, law enforcement, prison inmates, psychiatric groups |
| Acronym | CPI |
| Scores | 1) Dominance; 2) Capacity for Status; 3) Achievement via Conformity; 4) Achievement via Independence; 5) Communality; 6) Flexibility; 7) Femininity/Masculinity; 8) Good Impressions; 9) Empathy; 10) Independence; 11) Responsibility; 12) Intellectual Capacity; 13) Psychological Mindedness; 14) Tolerance; 15) Well-being; 16) Sociability; 17) Social Presence; 18) Self-Acceptance; 19) Socialization; and 20) Self Control |
| Administration | Paper and pencil, individual |
| Price | Profile preview kit, $12.25; Item booklets (reuseable), $45.90 for 25; Answer sheets mail-in, $76.50 for 10, scannable answer sheets, $11.50 for 25; Manual $55.00. |
| Time | 45 to 60 minutes |
| Authors | Harrison G. Gough (1957; 1988) |
| Publishers | Consulting Psychologists Press |

### Theory

The goal of this instrument was to assess everyday kinds of variables that can be considered folk concepts. These folk concepts arise from and are linked to social interactions (Gough, 1988). These folk concepts were identified based on modeling ordinary people. As a result, the following twenty scales were designed to measure a person's personality: 1) Dominance; 2) Capacity for Status; 3) Achievement via Conformity; 4) Achievement via Independence; 5) Communality; 6) Flexibility; 7) Femininity/Masculinity; 8) Good Impressions; 9) Empathy; 10) Independence; 11) Responsibility; 12) Intellectual Capacity; 13) Psychological Mindedness; 14) Tolerance; 15) Well-being; 16) Sociability; 17) Social Presence; 18) Self-Acceptance; 19) Socialization; and 20) Self Control.

This set of 20 folk concept scales is intended to be sufficient to predict a broad range of interpersonal behaviors. The 20 scales can be reduced to 4-5 major factors, with the principal themes of extroversion/introversion and adjustment by social conformity. In addition, the 20 scales can be combined into three higher-order dimensions labeled V1, V2, and V3. The V1 scale taps introvert/inwardly-oriented/reserved behavior. The V2 scale assesses conscientious. The V3 scale determines the reflective capability of the individual.

**Development and Empirical Use**

In the current 462-item version, 194 items were taken from the Minnesota Multiphasic Personality Inventory (MMPI) (Hathaway & McKinley, 1967). Other items were reworded from the MMPI or were entirely original (Gough, 1952).

Scales were developed in three ways. Sixty-five percent of the scales were constructed by empirical and qualitative techniques. Then, selection and keying of items was conducted in such a way as to maximize the relationship between the responses to the test and the outcome target to be forecast. Thirteen scales were developed by empirical methods, in which an item analysis was conducted against non-test criteria. Twenty percent of the scales were developed by an internal consistency technique. This method began with a set of items, which on judgment appeared to be relevant to the aim of the measurement. Then, by studying item correlations, items that were the least consistent with whatever was assessed were removed. Fifteen percent of the scales were developed by mixed means (Gough, 1952).

**Psychometrics**

The internal consistency ranged from .46 to .85 for the various scales. The test-retest correlation ranged between .43 - .78. Parallel forms of the instrument correlated between .46 - .83. The factors accounted for 66% of the variance. The instrument does include lie scales (Gough, 1988).

In terms of validity, the CPI has been correlated against Cattell's 16PF, the MMPI, the MBTI, and cognitive measures, such as WAIS and SATs, showing convergent and discriminate validity (Gough, 1952). Studies of predictive validities have been completed to determine probable academic achievement in high school (Repapi, Gough, Lanning, & Stefanis, 1983), college (Gough & Lanning, 1986), and performance as a police officer (Hogan, 1971), with significant results.

**Generalizability**

The instrument was designed to be used with ages 14 and up. Samples have included managers, military, students, engineers, architects, police, religious groups, prison inmates, and psychiatric groups. Therefore, its applicability is likely to be fairly diverse.

**Face Validity/Ease of Use/Transparency**

This is a paper and pencil or computerized instrument containing 462 items. It is self-scored, computer-scored, or can be scored by the publisher. There are no rigorous controls, therefore, the instrument is administered in informal sessions and through the mail, which produces quick results. This is a commercial instrument available for a fee. Based on our review of the items, face validity is low to moderate depending on specific questions.

## ARI Measures vs. Benchmarks

### Summary

As we began this section, we noted the wide diversity of theories and measures that are prevalent within the domain of personality. In general, and as evidenced by our benchmark measures, the psychology research community has adopted a five factor view of personality. While debate continues as to the precise number and composition of these factors, in general, it is fair to say that there is more consensus than disagreements concerning the basic structure of personality. The ARI researchers, however, have tended to employ measures of proclivity in an effort to operationalize features of SST. Unfortunately, they have not done so in a consistent manner. Furthermore, no study to date has attempted to fully assess the proclivity domain as articulated by SST. Consequently, it is difficult to draw firm conclusions about the role of proclivity either within or across investigations.

In general, most of the featured and benchmark instruments are fairly similar in format and administration. They are self-report, and either paper and pencil or computerized for individuals to complete. The biggest difference was found in the SOI, which takes considerably more time both to administer and to score. As noted above, the specific content of the measures varies significantly, even across the ARI measures. The measures are all attempting to tap some aspect of proclivity, but each taps different facets. None of the ARI measures seem superior in terms of comprehensiveness of measurement of the proclivity dimensions.

The development of the instruments is comparable across the biodata measures, SOI, and the benchmark personality measures, the NEO-PI, CPI, and HPI. Overall, they have strong, theoretical bases with appropriate empirical methods of instrument development. The MBTI, on the other hand, does not demonstrate as strong of a development process as compared to the other measures.

All of the benchmarks and the MBTI have a fairly long history of empirical use in many different types of settings. The ARI biodata and SOI measures have had more limited use. This could be due to many factors, one of which is the instruments' relative newness as compared to the other measures. Another reason for the limited use of the SOI is obviously attributable to its administration and scoring demands.

The ARI instruments and the benchmarks were all comparable in terms of reliability, showing moderate to strong evidence. The SOI is comparable to the benchmarks in terms of construct validity, and the HPI shows the strongest criterion-related validity. The instrument with the poorest validity evidence, both in terms of construct and criterion-related validity, is the MBTI.

All of the instruments tend to have somewhat low face validity, with the items comprising the biodata measures being viewed as the most face valid. In terms of ease of use, all instruments rank comparably, except for the SOI, whereas the ARI measures appear to be less transparent than the benchmarks.

**Recommendations**

Personality has gained a renewed place as a predictor of performance in applied settings in the past decade or so. Both the theoretical foundation and accumulating empirical evidence suggests that it will only gain in importance in understanding the effectiveness of leaders in the future. In order for research to advance, however, a clear measurement scheme must be articulated and applied across investigations. In particular, it is important to tie personality measures to leader effectiveness: 1) within different career stages; 2) in different task and operational environments; and 3) at different hierarchical levels. If the larger research base has informed us of anything, it is probably that there are few universally predictive personality attributes. The sole exception to this rule, however, is the conscientiousness dimension, which has consistently exhibited significant positive correlations with indices of job performance (Barrick & Mount, 1991). Accordingly, we believe measures of it should be incorporated in future ARI investigations concerned with the role of personality in leader effectiveness.

In terms of other recommendations for the future, we believe that some further foundation work is in order. We would characterize much of the work that has been done to date as attempts to link aspects of proclivity to leader effectiveness using whatever measures were available or would work within certain administrative constraints of a project. There is a need to back up, so to speak, and to articulate the precise structure thought to underlie the concept of proclivity, develop measures for each facet or subdimension, and then to empirically validate that structure using a varied sample of Army officers and sophisticated statistical techniques (e.g., confirmatory factor analyses). This would facilitate the

development of a single measure battery that could be used in future studies and thereby permit comparisons across studies. We would also suggest that during the course of such development, one or more measures of the Big 5, such as the benchmark instruments reviewed here, be administered. This would permit direct comparisons between the different approaches to studying personality, illustrate areas of overlap, and likely yield a more comprehensive assessment of personality that would be comparable to the larger research literature.

**Table 3. Personality Comparison**

| | ARI Featured Instruments | | | Benchmarks | | |
|---|---|---|---|---|---|---|
| ***Personality Facet*** | ***Subject-Object Interview*** | ***Biodata*** | ***Myers Briggs Type Indicator*** | ***Hogan Personality Inventory*** | ***NEO-Personality Inventory*** | ***California Personality Inventory*** |
| **Extroversion** | X | X | X | X | X | X |
| **Activity/energy** | | X | | X | X | |
| **Assertiveness** | | | | X | X | |
| **Excitement seeking** | | | | | X | |
| **Gregarious** | | | | X | X | X |
| **Positive emotions** | | | | | X | |
| **Warmth** | | | | | X | |
| **Neuroticism** | | X | | X | X | |
| **Anger hostility** | | | | | X | |
| **Anxiety** | | | | X | X | |
| **Depression** | | | | X | X | X |
| **Discretion** | | | | | X | |
| **ego control** | | X | | X | X | X |
| **Emotional control** | | X | | X | X | |
| **Impulsiveness** | | | | | X | |
| **self consciousness** | | | | | X | |
| **Vulnerability** | | | | | X | |
| **Conscientiousness** | | X | | X | X | |
| **Achievement striving/oriented** | X | | | | X | X |
| **Cautious** | | | | X | X | |

| | ARI Featured Instruments | | | Benchmarks | | |
|---|---|---|---|---|---|---|
| ***Personality Facet*** | ***Subject-Object Interview*** | ***Biodata*** | ***Myers-Briggs*** | ***Hogan Personality Inventory*** | ***NEO-Personality Inventory*** | ***California Personality Inventory*** |
| **competence** | | | | X | X | |
| **deliberation/** | | | | X | X | |
| **playful** | | | | | | |
| **dutifulness** | | | | | X | |
| **orderly** | | | | X | X | |
| **responsible** | | | | X | X | X |
| **self-disciplined** | | | | X | X | |
| **Agreeableness** | | X | | X | X | X |
| **altruism** | | | | | X | |
| **caring** | | | | | X | X |
| **cheerful** | | | | X | X | |
| **compliance** | | | | | X | |
| **cooperative/competitive** | X | X | X | X | X | X |
| **flexible** | | | | | X | |
| **good natured** | | | | X | X | |
| **modesty** | | | | | X | |
| **not jealous** | | | | | X | X |
| **straightforwardness** | | | | X | X | |
| **tender mindedness** | | | | | X | |
| **trust** | | | | | X | |

| | ARI Featured Instruments | | | Benchmarks | | |
|---|---|---|---|---|---|---|
| ***Personality Facet*** | ***Subject-Object Interview*** | ***Biodata*** | ***Myers-Briggs Type Indicator*** | ***Hogan Personality Inventory*** | ***NEO-Personality Inventory*** | ***California Personality Inventory*** |
| **Openness to Experience** | | X | | X | X | |
| **actions** | | | | X | X | |
| **aesthetics/artistically sensitive** | | | | X | X | |
| **curious** | | X | | X | X | |
| **fantasy** | | | | | X | |
| **feelings** | | | | X | X | |
| **ideas/original** | | | | | X | |
| **independent** | | | | X | X | |
| **intellectual** | | | | | X | X |
| **imaginative** | X | | X | X | X | |
| **values** | | | | X | X | |
| **Miscellaneous** | | | | | | X |
| **achievement via conformity** | | X | X | | | |
| **capacity for status** | | | | | | X |
| **communality** | | | | | | X |
| **dominance** | | | | | | X |
| **empathy** | | | | | | X |
| **flexibility/adaptability** | | X | | X | X | X |
| **good impressions** | | | | | | X |

| | ARI Featured Instruments | | | Benchmarks | | |
|---|---|---|---|---|---|---|
| ***Personality Facet*** | ***Subject-Object Interview*** | ***Biodata*** | ***Myers-Briggs Type Indicator*** | ***Hogan Personality Inventory*** | ***NEO-Personality Inventory*** | ***California Personality Inventory*** |
| **independence** | | | | | | X |
| **locus of control** | X | X | X | | | X |
| **responsibly** | | | X | | | X |
| **self acceptance** | | | | | | X |
| **self control** | | | | | | X |
| **sensing** | X | X | X | | | X |
| **sociability** | | | | | | X |
| **social presence** | | | | | | X |
| **socialization** | | | | | | X |
| **tolerance for ambiguity** | | | | | | X |
| **tolerance for failure** | | | | | | X |

# Section 3: General Knowledge Areas

The database of ARI leadership variables that we compiled over the course of this evaluation project yielded 75 variables that were categorized in the knowledge domain. Based on this representation as well as the nominations by the ARI Research scientists, we concluded that leader knowledge was an important area to feature in this report. In particular, ARI's research concerning general leader knowledge, problem solving, mental models, cognitive complexity, and tacit knowledge qualified for in-depth review. We should note, however, that for convenience we will be using the term "knowledge" fairly loosely to include variables that are sometimes considered to be cognitive abilities or skills. While distinctions between knowledges, and cognitive abilities and skills are often important in practice, this latitude provides a more simplistic organizational scheme for present purposes. Thus, the following seven ARI measures were chosen to be featured in this section: 1) general leader knowledge as tapped by biodata and 2) critical incidents; 3) problem solving tasks; 4) Constructed Response Exercises; 5) mental models; 6) the Career Path Appreciation (CPA); and 7) the Tacit Knowledge for Military Leadership Inventory (TKMLI).

**Overview**

Cognitive skills have been found to be necessary for effective military leadership. Research in mental abilities has a long history in the intelligence community. The most current work attempts to reintegrate intelligence as traditionally measured with a broader concept of intellect. The work we reviewed spanned a wide range of conceptions of knowledge, theoretical grounding, and assessment techniques. Consequently, a single generic overview is difficult to provide. In order to provide a common point of reference, however, we have outlined Fleishman & Quaintance's (1984) taxonomy of 17 general cognitive abilities. This framework, which is discussed more extensively in the next subsection, provides a well developed foundation against which to gauge measures of general cognitive abilities. However, more recently researchers have sought to develop more focused measures of knowledge such as tacit or practical knowledge for a given domain of work. Therefore, throughout this section we will provide brief foundation reviews of the theory behind different measurement approaches and how the particular assessment activities map back to them. We will revisit this general vs. focused knowledge measures theme in the summary and recommendations section at the end of the report.

We first consider ARI measures of general leader knowledge as assessed by background data (i.e., biodata) and critical incidents procedures. Following this, measures of higher-order concepts such as intellect will be presented, along with measurement options used at ARI. First,

problem-solving skills, as assessed in the problem solving tasks and Constructed Response Exercises will be considered. This will be followed by a description of mental models, a more recent topic in leadership research. Next, cognitive complexity, as measured by CPA, will be discussed. The ARI featured instrument section concludes with a review of the TKMLI.

Due to the wide range of variables covered by ARI research, it was necessary to locate many different benchmarks. General leader knowledge is benchmarked against the Watson-Glaser Critical Thinking Assessment, the Concept Mastery Test, and the Guilford Consequences. The problem-solving benchmark is the Leatherman Leadership Questionnaire. ARI mental model measurement will be benchmarked against a different measurement procedure called Pathfinder and illustrated in a recent study by Stout, Salas and Kraiger (1997). The CPA, which taps cognitive complexity, is benchmarked against the Low Fidelity Simulation instrument. The final featured ARI instrument, the TKLMI, will be compared to a tacit knowledge measure used in an academic context.

This section begins with the presentation of Table 4, which displays the various ARI instruments and benchmarks for this section on the eight criteria used for evaluation. Next, the various ARI instruments, followed by all of the benchmarks are presented. This section will conclude with an evaluation of the featured instruments, and suggestions for future research. Table 5 displays the Fleishman and Quaintance (1984) taxonomy (with additional dimensions included), and illustrates which ones each instrument addresses.

**Table 4**

Leader Knowledge (ARI Featured Instruments)

| Criteria | ARI Background Data Inventory | ARI Critical Incidents | Problem-solving Tasks | Constructed Response Exercise | Mental Models | Career Path Appreciation | Tacit Knowledge for Military Leadership Inventory |
|---|---|---|---|---|---|---|---|
| Theory | leader cognitive abilities | leader cognitive abilities | problem-solving in ill-defined, dynamic environments | problem-solving leadership skills | team, organization, and vision mental models; conceptual knowledge | conceptual complexity; stratified systems theory | tacit knowledge |
| Descriptive information | 11 scales that cover reasoning skills as well as basic cognitive abilities 398 items | 11 scales that cover reasoning skills as well as basic cognitive abilities 26 items | 5 scales: problem construction, information encoding, category search, category combination, and wisdom | 3 scales: solution construction, social judgment skills, and creative problem-solving | rated accuracy, breadth, depth, and organization of mental models | 3 tasks: phrase selection (cards), symbol sorting (cards), work history (interview) | assumes contingency approach |
| Development | moderate | moderate | moderate | strong | strong | strong | strong |
| Empirical use | moderate use | low use | limited use | widespread use | moderate use | limited use | limited use |
| Generalizability | moderate | low | low to moderate | high | high | moderate | low |
| Psychometrics | high reliability | N/A | moderate interrater reliability moderate to strong criterion-related validity | moderate interrater reliability strong criterion-related validity | high interrater reliability strong criterion-related validity | high reliability poor construct validity strong criterion-related validity | reliability – N/A moderate criterion-related validity validity in progress |

Table 4 (cont'd)

Leader Knowledge (ARI Featured Instruments)

| **Criteria** | **ARI Background Data Inventory** | **ARI Critical Incidents** | **Problem-solving Tasks** | **Constructed Response Exercise** | **Mental Models** | **Career Path Appreciation** | **Tacit Knowledge for Military Leadership Inventory** |
|---|---|---|---|---|---|---|---|
| Face validity | moderate | N/A | moderate | high | high | low | high |
| Ease of use | moderate | moderate | moderate | low | moderate | low | low |
| Transparency | moderate | dependent | low | N/A | low | low | moderate |

Table 4 (cont'd)

Leader Knowledge (Benchmarks)

| **Criteria** | **Watson-Glaser Critical Thinking Appraisal** | **Concept Mastery Test** | **Consequences** | **Low Fidelity Simulation (Motowidlo et al.)** | **Leatherman Leadership Questionnaire** | **PathFinder** | **Tacit Knowledge** |
|---|---|---|---|---|---|---|---|
| Theory | critical thinking skills | meta-cognitive processes and skills involving manipulation of abstract concepts | ideational fluency and originality as components of divergent thinking | past experiences and behavior predicts future behavior | problem solving | mental model structure | tacit knowledge |
| Descriptive information | 80 items scales: inference, recognition of assumptions, deduction, interpretation, evaluation of arguments | 2 types of items: synonyms-antonyms and analogies (verbal and number) | 10 items responses scored for obviousness and originality | multiple-choice format assumes contingency approach | knowledge-based measure of supervisory leadership | related-ness of pairs of concepts are rated, then converted into network; computer program | 12 work-related situations |
| Development | moderate | moderate | moderate | strong | weak | moderate | strong |
| Empirical use | widespread use | moderate use | moderate use | limited use | moderate | moderate | widespread (academics) |

Table 4 (cont'd)

Leader Knowledge (Benchmarks)

| **Criteria** | **Watson-Glaser Critical Thinking Appraisal** | **Concept Mastery Test** | **Consequences** | **Low Fidelity Simulation (Motowidlo et al.)** | **Leatherman Leadership Questionnaire** | **PathFinder** | **Tacit Knowledge** |
|---|---|---|---|---|---|---|---|
| Generaliz-ability | high | moderate | high | low | high | high | low |
| Psychometrics | moderate split-half reliability<br>moderate test-retest reliability<br>moderate construct validity<br>strong criterion-related validity | high reliability<br>high test-retest reliability<br>moderate construct validity<br>strong criterion-related validity<br>strong discriminant validity | high internal consistency reliability<br>strong construct validity | high reliability<br>strong construct validity<br>strong criterion-related validity | high internal consistency reliability<br>mixed criterion-related validity | moderate test-retest reliability<br>moderate construct validity<br>moderate criterion-related validity | mixed reliability<br>strong criterion-related validity |
| Face validity | high | moderate | low | high | high | low | high |
| Ease of use | high | high | moderate | low | low | Moderate | high |
| Transparency | low to moderate | low | low | moderate | moderate | low | low |

# ARI Research on General Knowledges

### General Knowledges: ARI Background Data Inventory

| | |
|---|---|
| Purpose | Validation study on cognitive abilities of leaders |
| Population | Civilian Supervisors, 1st, 2nd, and 3rd level in 6 work grades |
| Acronym | ARI-BDI |
| Scales | 1) Verbal comprehension; 2) Written comprehension; 3) Verbal expression; 4) Written expression; 5) Definition of the problem; 6) Fluency of ideas; 7) Originality; 8) Problem anticipation; 9) Deductive reasoning; 10) Inductive reasoning; 11) Information ordering |
| Time | 2 to 3 hours |
| Price | N/A |
| Author | Mumford, Zaccaro, Harding, Fleishman, & Reiter-Palmon, 1991 |
| Publisher | N/A |
| Comments | This section contains a part of a larger research project to validate the Knowledge, Skills, Ability and Personality (KSAP) model. The information below covers only knowledges, and specific cognitive abilities or skills. The personality and leader behavior sections of this report cover the other areas. |

**Theory**

Leadership in the U.S. Army is viewed as an open system where leaders are embedded in a social context. Based on this theory, many different elements are seen as tasks that the leader must address, such as subordinate motivation, coordinating needs, subsystem maintenance, and negotiation. Due to these situational influences, key leadership qualities include a number of interconnected characteristics, such as, personality and knowledges. The importance of personality has already been addressed in this report, particularly the notion of proclivity. Knowledges also influence leader effectiveness, especially in domains that are highly variable in terms of demand characteristics, or in situations in which novel approaches are needed to solve problems (Mumford et al., 1991). As a result, there is a premium placed on knowledges and skills, such as, intelligence, creativity, and crystallized cognitive skills (Jacobs & Jaques, 1989).

In order to organize the vast majority of variables that can encompass cognitive skills studied in ARI, Fleishman and Quaintance's (1984) taxonomy is used as a common framework. This taxonomy is illustrated in Table 5 and defined by the following components: 1) the ability category; 2) the ability; and 3) the definition of the ability.

Fleishman and Quaintance (1984) identified a total of seventeen abilities. The ARI-BDI addresses eleven of them.

The first ability category is labeled linguistic. ARI-BDI taps four different types of abilities in this category. They are the following: 1) verbal comprehension; 2) written comprehension; 3) verbal expression; and 4) written expression. Verbal and written comprehension are defined as the ability to understand language, either written or spoken, such as to hear a description of an event and understand what happened. Verbal and written expression are defined as using either verbal or written language to communicate information or ideas to other people. This includes vocabulary, knowledge of distinctions among words, and knowledge of grammar and the way words are ordered (Fleishman & Quaintance, 1984; Mumford et al., 1991).

The second ability category that ARI-BDI taps is creativity. Constructs highlighted in this category are: 1) the definition of the problem; and 2) fluency of ideas and originality. Problem definition involves the determination of what precisely is the problem, what its parts are, and how these parts are related to one another (Dillion, 1982). Fluency of ideas is the ability to produce a number of ideas about a given topic. This ability only concerns the *number* of ideas, not the quality. The third construct is originality which is defined as producing unusual or clever responses to a given topic or situation and/or to improve solutions in situations where standard operating procedures do not apply (Fleishman & Quaintance, 1984; Mumford et al., 1991).

The third ability category tapped by ARI-BDI is problem solving and reasoning. However, ARI-BDI uses different labels for these abilities. Problem anticipation, deductive reasoning, inductive reasoning, and time-sharing are under the dimension of general cognitive intelligence (Mumford et al., 1991). Information ordering is under the dimension of crystallized cognitive skills. However, their definitions remain the same as in Fleishman and Quaintance's taxonomy (1984). Problem anticipation is defined as recognizing or identifying the existence of problems; involving both the recognition of the problem as a whole, and the elements of the problem. This construct does not include the ability to solve the problem (Fleishman & Quaintance, 1984). Deductive reasoning is defined as applying general rules or regulations to specific cases, or proceeding from stated principles to logical conclusions (Fleishman & Quaintance, 1984). Inductive

reasoning is the skill of finding a rule or concept that fits the situation, such as determining a logical explanation for a series of unrelated events (Fleishman & Quaintance, 1984). The last skill, information ordering, involves applying rules to a situation for the purpose of putting the information in the best or most appropriate sequence. It also involves the application of previously specified rules and procedures to a given situation (Fleishman & Quaintance, 1984; Mumford et al., 1991).

**Development and Empirical Use**

A validation study was performed with 1037 men and 897 women who were freshman university students. The participants completed a 398-item background questionnaire (Owens & Schoenfeldt, 1979). From there, a self-evaluation leadership scale was constructed using 19 background data items.

To identify constructs related to leadership, a variation on rational clustering procedures was used. Items that yielded correlations greater than .10 and were significant at .01 level were used for cluster generation. Items were rationally assigned to clusters, which resulted in five clusters being established. One of the clusters was cognitive ability, with the subscales that are defined above in the theory section. The other four clusters were motivational characteristics, personality, social skills, and development (Mumford et al., 1991).

**Psychometrics**

The alpha coefficients obtained for the leadership scale were .80 for men and .82 for women in the university sample. The validation of the instrument was assessed by a blocked regression with item clusters entered in a stepwise fashion until all clusters were represented. Cognitive factors were entered first, and yielded multiple Rs of .41 for males and .44 for females. The strongest predictor from within that block was inductive reasoning (Mumford et al., 1991).

**Generalizablity**

The sample in the validation study was composed of university students. The survey has also be used with military leaders, therefore, generalizablity is high.

**Face Validity/Ease of Use/Transparency**

The items vary in terms of transparency and face validity. Overall, the instrument is moderate on both criteria, in our opinion. The measure is also easy to use, based on the paper and pencil format. However, it tends to be quite lengthy, with nearly 400 items.

## General Knowledges: ARI Critical Incidents

| | |
|---|---|
| Purpose | Validation study on cognitive abilities of leaders |
| Population | N = 4 |
| Acronym | ARI-CI |
| Scales | 1) Verbal comprehension; 2) Written comprehension; 3) Verbal expression; 4) Written expression; 5) Definition of the problem; 6) Fluency of ideas; 7) Originality; 8) Problem anticipation; 9) Deductive reasoning; 10) Inductive reasoning; 11) Information ordering |
| Time | 1 to 2 hours |
| Price | N/A |
| Author | Mumford, Zaccaro, Harding, Fleishman, & Rieter-Palmon (1991) |
| Publisher | N/A |
| Comments | This section contains a part of a larger research project to validate the Knowledge, Skills, Ability and Personality (KSAP) model. The information below covers only knowledges and specific cognitive abilities or skills. The personality and leader behavior sections of this report cover the other areas. |

### Theory

The same theoretical background described in the previous section was applied here.

### Development and Empirical Use

Twenty-six critical incidents, representing a diverse set of problems confronting mid- to upper-level management were selected from case studies in the general literature. The eleven cognitive dimensions were rated by four judges as to whether their possession would contribute to effective leader performance in the case study. A 1 to 5 likert scale with 5 being the highest was used for ratings (Mumford et al., 1991).

### Psychometrics

Eight of the eleven dimensions had means above 2.5. The following fell below that median range: 1) verbal comprehension (2.42); 2) written expression (2.21), and 3) information ordering (1.79) (Mumford et al., 1991).

### Generalizability

Since the sample only contained four individuals who were not identified in terms of age or occupation, generalizability is low.

**Face Validity/Ease of Use/Transparency**

This was essentially a content/construct validity judgement task. As such, issues concerning the face validity, ease of use, and transparency are rendered moot. These issues await administrations with a sample of targeted officers and ties with criterion measures.

## Problem-Solving Tasks

| | |
|---|---|
| Purpose | Assess leader problem solving skills |
| Population | Undergraduate students |
| Acronym | N/A |
| Scores | 1) Problem construction; 2) Information encoding; 3) Category search; 4) Category combination; and 5) Wisdom |
| Administration | computer, individual |
| Price | N/A |
| Time | estimated 90 minutes |
| Authors | Mumford, Baughman, Supinski, Costanza, & Threlfall (1993) derived measures from several sources |
| Publishers | ARI |

### Theory

An effective leader must have the ability to solve problems, not only in well-defined areas, but also in ill-defined, dynamic environments. A straightforward model for problem solving would begin with defining the problem situation (Mumford et al., 1993). Next, the leader must select information bearing on the problem situation, and concepts that will help to organize and understand the information. The leader must then combine and reorganize these concepts and relevant information to create a model for understanding the problem. This stage in problem solving will lead to the generation of initial solutions. Wisdom and perspective taking are then applied to assess others' reactions to the solution, and to identify any restrictions and revisions that may be necessary.

This problem-solving model stresses the importance of cognitive skills, such as problem construction, information encoding, category or concept search, and wisdom (Mumford et al., 1993). To identify the skills needed by a leader, the organizational leadership position must be examined. One useful model for examining the organizational leadership position is the systems theory (Pfeffer & Salancik, 1978). This socio-technical systems theory holds that organizations emerge because people achieve goals by working together. The organization represents a linked collection of subsystems, which operate together to produce services and to meet the goals of constituencies. In order to meet these goals, materials are taken from the environment

and transformed into useful products. The efficiency of the transformation is enhanced by specialization and role differentiation.

The performance requirements from an organizational standpoint are functional in nature. The job of the leader is to insure that all functions critical to both task accomplishment and group maintenance are adequately addressed (McGrath, 1976). The leader must generate and implement solutions to novel, ill-defined problems in a rapidly changing social context in order to be characterized as an effective leader (Mumford et al., 1993). Leaders must possess certain characteristics that allow them to locate and solve complex, ill-defined social problems. It is expected that intelligence, social skills, and dominance or achievement motives would consistently be related to leader performance (Mumford et al., 1993). Studies regarding individual characteristics have led to ambiguous findings, with the exception that leader performance is apparently dependent on basic cognitive capacities and social skills (Mumford et al., 1993; Connelly, Zaccaro, & Mumford, 1992; Mumford et al., 1991).

It has been argued that differential capacities, such as intelligence, are not directly responsible for the solution of ill-defined social problems leaders encounter. Instead, differential characteristics, such as social skills and cognitive capacity, operate by facilitating the development of and application of knowledge structures and problem solving skills (Snow & Lohman, 1984). These characteristics feed into the cognitions of both experts and novices. Those individuals with well-organized, more extensive knowledge structures are better able to identify, recall, and impart meaning to the information required for effective problem solving (Siegler & Richards, 1982).

However, in organizations, the existence of formal knowledge may not be ample for insuring adequate leader performance. Leaders also need to possess an informal understanding of the organizational system in which they will implement solutions. This informal knowledge allows the leader to identify viable strategies for applying knowledge, as well as appraising the results of the feedback (Mumford et al., 1993). This cognitive ability is known as tacit knowledge (Wagner & Sternberg, 1985), or knowledge acquired through experience on the job.

In addition to cognitive skills, a leader must possess social skills. Some of these skills include negotiation, empathy, and behavioral flexibility (Shiflett, Eisner, & Inn,

1981). Other skills involved in the acquisition and appraisal of social information may be social perceptiveness and/or wisdom (Zaccaro, Gilbert, Thor, & Mumford. 1991; Connelly et al., 1993).

Complex information processing skills, like expertise, tacit knowledge, and social skills can be expected to develop with experience as individuals work through different kinds of problem situations. The capacity to apply these knowledges and skills may emerge at a slow rate over a relatively long passage of time (Mumford et al., 1993). The rate of development will depend, in part, on the basic abilities, motives, and personality characteristics individuals bring to their problem solving experiences (Mumford et al., 1993).

Skill assessment in regard to problem solving and social appraisal skills can be accomplished in many different ways. One way is through open-ended responses to complex, realistic problems (Mumford & Teach, 1993). This approach is advantageous with regard to ecological validity because it assesses complex skills without overly structuring responses. However, developing the ratings or protocol scoring for the complex open-ended items is unusually costly and time-consuming. Typically, four or five judges must revise each subject's responses using benchmark rating scales. Further, the judges must typically be given at least one week of training before they can produce reliable ratings.

The five problem-solving skills assessed in this study were: 1) problem construction; 2) information encoding; 3) category search and specification; 4) category combination; and 5) wisdom. These skills are defined as follows:

1) problem construction – requires the identification and structuring of a problem; individual does not work with givens;
2) information encoding – ability to absorb information;
3) category search and specification – the ability to link information to existing concepts or schemas;
4) category combination – the ability to combine and synthesize diverse concepts;
5) wisdom – involves self-objectivity, self-reflection, judgment under uncertainty, system perceptiveness, sensitivity to fit, and social commitment.

**Development and Empirical Use**

Creative problem-solving and social appraisal skills were assessed by five tasks administered via computer. The first task, which was composed of four problem scenarios, was used to measure problem construction or problem finding skills. These scenarios were based on those developed by Baer (1988), and consisted of complex, ill-defined situations that may be structured in variety of ways. The response options for these scenarios were generated by four doctoral students who each composed four restatements of each problem. These problem restatements provided one each of the four possible combinations: 1) one high quality, high originality restatement; 2) one high quality, low originality restatement; 3) one low quality, high originality restatement; and 4) one low quality, low originality restatement. These restatements were presented to five additional doctoral students to be reviewed based on the following four types of information: 1) goals; 2) procedures; 3) key information; and 4) restrictions. Based on the consensus of three of five judges, responses were determined to mark a preference for a type of representational content.

Next, 30 doctoral students rated the restatements for quality and originality, as well as for the use of the four types of information. The interrater agreements for the quality and originality judgments were .92 and .89, respectively. The interrater agreement coefficients for goals, procedures, key information, and restrictions were .88, .82, .91, and .88, respectively.

Sixteen of the responses that were generated were chosen based on the following criteria: 1) four responses were chosen based on high and low quality, and high and low originality restatement ratings; and 2) responses were chosen that covered the four content dimensions (e.g., goals, procedures, key information, and restrictions), while varying on quality and originality. The scoring of these four scenarios was accomplished by the quantity of high quality and high originality restatements chosen, as well as the preference for structuring the problem in terms of goals, procedures, key information, or restrictions.

The second task, information encoding, was comprised of four problems. These problems were based on two business case studies and two political case studies (Athos & Gabarro, 1978; Janke, 1992 as cited in Mumford et al., 1993). Participants were

required to read six index cards as presented on the computer. Next, they were asked to type a paragraph solution to the problem. Three of the index cards for each of the four problems contained core facts based on the case studies. For two of the problems, the other three cards addressed additional information, such as, principles for organizing information, consistency information, and relatedness information. For the other two problems, goals, constraints, and the range of the problem situation were presented.

The scoring of this task was accomplished by the total time spent and proportion of time spent on each card from each category of information, as well as on the core facts. Four judges rated the quality and originality of solutions to the problems to provide criterion evidence of the effect of each style on performance.

The third task, category search, was composed of abstracts of four complex, ill-defined organizational scenarios based on ones used by Shorris (1981). The participants were required to answer the following questions: 1) why the situation occurred; 2) what major mistakes were made; and 3) what they would do. Four doctoral students reviewed the material presented in the scenarios, and then generated concepts or categories that would explain the problem situations. These concepts were generated with the following criteria in mind: 1) abstractness; 2) relatedness; 3) long-term outcomes; and 4) integration.

A total of 188 concepts was gathered from the students and was presented to an additional five students. These judges then rated each concept as to the dimension targeted. Two statements with high mean ratings and low standard deviations were then chosen. The eight statements that were generated became the response options. Participants, after reading the problem scenario, would select four of the concepts that they found helpful to understanding the problem situation. Dimensional weights were assigned for abstractness, relatedness, long-term goals, and integration based on the respondents' choice of useful concepts to the solution of the problem. These dimensional weights were based on the ratings by 30 judges as to how well the concepts fit the four dimensions.

The next task designed to evaluate problem-solving ability was category combination. This task included six category-exemplar generation problems that were based on Mobley, Doares, and Mumford (1992). Respondents were required to generate

a new category that would fit the four exemplars presented. They also needed to label the category and list more exemplars for the new category.

An expert scoring system was used for this task. Four judges rated labels, features, and exemplars for solution quality and originality. The interrater agreement coefficients ranged from .68 to .75. The labels, features, and exemplars obtained from respondents were compared to those from previous studies in order assign scores.

The final task that participants had to perform was a measure of wisdom. Ten of the less well-known Aesop fables were presented, and respondents identified the moral of the story. The scoring system was developed in a previous study that had five doctoral students rate proposed morals as compared to actual morals. These ratings were used in the current study to develop five response options for the ten fable problems. The five response options were different approximations of the actual moral of the story.

**Psychometrics**

The problem construction measures significantly predicted the criterion measure of problem performance, which was comprised of the following four measures: 1) advertising task quality; 2) advertising task originality; 3) problem-solving quality; and 4) problem-solving originality. The problem construction measures significantly predicted all four of the criterion measures. The information encoding measures also significantly predicted all of the criterion measures. The results of this study showed evidence that the category search measures significantly predicted three of the four criterion measures of problem performance (e.g., advertising task originality, problem-solving quality, and problem-solving originality). For category combination, there was evidence of this scale significantly predicting all four of the criterion measures. The final scale of wisdom yielded a significant prediction of only two out of the four criterion measures, (e.g., advertising task quality and advertising task originality).

In terms of incremental validity beyond basic abilities, all of the problem-solving skills produced significant gains in the prediction of the criterion measures.

**Generalizability**

The generalizability of these results may be questionable due to the ability level of college undergraduate students, as compared to other populations of lower general

ability levels. In addition, it is unclear whether the findings could be extrapolated to the field context.

**Face Validity/Ease of Use/Transparency**

Based on our review of examples of the five tasks, the problems appear to be moderately face valid. It may not be completely obvious to respondents that these problems are tapping problem-solving ability. In addition, it is a stretch for respondents to understand that these problems ultimately are meant to be indicative of leadership ability via problem-solving ability. These types of problems are difficult and time-consuming to develop; demanding the use of experts to generate response options and assign weights to responses. The problems are very easy to administer due to the use of computers. In our opinion, the nature of these problems is such that they are low in transparency.

We should, however, add two cautions about these measures. First, much of the development work was predicated on the judgment of graduate students. While we do believe that such a population is well equipped to make ratings and avoid traditional ratings errors such as halo, prototype biases, etc., they do not possess the extensive real worlds experience that incumbent SMEs provide. Therefore the "groundedness" of these measures is open to debate. Second, much of the criterion related validity evidence was garnered by correlating scores on these measures with other measures of knowledge. While such a strategy does offer evidence in terms of the construct validity of the measures, it does not yield information akin to concurrent or predictive validity designs. Accordingly, it is important to gather these measures from incumbent officers and correlate them with job-based criteria measures.

## Constructed Response Exercises

| | |
|---|---|
| Purpose | Assess problem solving leadership skills |
| Population | Army civilian leaders from lower, middle, and upper leadership levels |
| Acronym | N/A |
| Scores | 1) Solution construction; 2) Social judgment skills; and 3) Creative problem-solving |
| Administration | written essay, individual |
| Price | N/A |
| Time | 30 minutes |
| Authors | Zaccaro, White, Kilcullen, Parker, Williams, & O'Connor-Boes (1997) |
| Publishers | ARI |

### Theory

These measures begin the shift from generic general knowledge to that which is grounded in organizational situations. More specifically, grounded in the tenets of SST theory, three abilities were focused upon:

Creative problem solving is the ability to approach, define, and solve a problem in a novel yet realistic fashion (Zaccaro et al., 1997). Creative solutions to problems are those that attend to the problem's parameters yet go beyond role, typical responses.

Solution definition may be described as one's ability to structure complex, ill-defined problems while considering the particular solution constraints and situational constrictions that exist in the broader problem context (Zaccaro et al., 1997). Solution definition skills rely on the ability to interpret problem parameters correctly (e.g., budget constraints), thereby anticipating the characteristics of a likely solution.

Social judgment is an understanding of how multiple constituencies (e.g., individuals or customers) interact to influence problem interpretation and solution development (Zaccaro et al., 1997).

### Development and Empirical Use

The three scenarios used in this measure were adapted from previous study conducted by Zaccaro, Mumford, Marks, Connelly, Threlfall, Gilbert, and Fleishman (1996) to fit the context of Army civilian executives. Each of the scenarios measures one of the following skills: 1) solution construction; 2) social judgment; and 3) creative problem-solving. These scenarios contained complex, ill-defined problems with multiple components that needed to be addressed by the respondents.

Two of the scenarios used cues to elicit certain problem-solving skills during the response (e.g., solution construction and social judgment). This was expected to lead the respondents to use the targeted skills in solving the problem. Cuing was accomplished by asking three questions, which the respondents had to answer in their essay response. For the creative problem-solving exercise, no cues were provided.

According to the researchers, in order to score these exercises, it is necessary to extract skill application information from the participants' responses. Thus, scoring is dependent on expert ratings or judgments of the respondent essays. This requires raters to be trained carefully so that they are capable of differentiating the essays based on quality. In addition, the raters also need to recognize the application of the targeted problem skill as tapped by each essay.

The scoring protocols were also based on those from the Zaccaro et al. (1996) study, with revisions for this study made by experts. The first step to developing the scoring protocols for this application of the exercises was to have experts read the problem scenarios and indicate both high and low quality responses. For this study, the experts were upper-level civilian managers. These responses were then used to generate examples of strong and weak applications of the targeted skills for each of the measures. These examples also showed the effectiveness of the solutions. Once the scoring protocol was developed, the raters were trained on it.

The raters for this study were graduate students who were experts on the topics of leadership and cognitive psychology. Students were used as the experts in this study because Army civilian leaders were not available for scoring the exercises.

The sample for this study consisted of 543 Army civilian leaders from lower, middle, and upper leadership levels distributed across six government service grades.

**Psychometrics**

The interrater reliability for the solution construction skills measure was .68. The social judgment skills measure had an interrater reliability of .69. The third measure, creative problem-solving skills, had an interrater reliability of .70.

The solution construction skills measure yielded significant correlations with all five of the leader activity variables (e.g., planning, special organization-wide projects, boundary spanning. entrusted problem-solving responsibility, and networking/

mentoring). The social judgment skills measure did not significantly correlate with any of the leader activities. The creative problem-solving measure had significant correlations with planning, special organization-wide projects, and boundary spanning.

The entire problem-solving skills set in this study, which included two additional biodata scales, showed a significant incremental contribution to the prediction of leader advancement. However, the set did not add anything to the prediction of the other three criteria in this study (e.g., leadership job performance, administrative criteria, and senior leadership potential).

**Generalizability**

As a result of the range of leadership levels and service grades, the results should easily generalize to other samples. It may more specifically generalize to civilian leader populations within the Army.

**Face Validity/Ease of Use/Transparency**

According to the researchers, the revisions to the context of the problem scenarios increased the face validity. The constructed response format also results in the exercises being more realistic than when participants just have to recognize an answer. Respondents only have 10 minutes to complete each scenario, which means that responses can not be too long. However, the scoring of the scenarios is problematic. Experts have to be relied upon to recognize the targeted skills, as well as assess the quality of the essay responses. This entails training the raters on a scoring protocol. However, there will still be a great deal of subjectivity involved in scoring the responses. Therefore, this set of exercises is not the easiest measure of problem-solving to use.

Based on our review of the measure, transparency concerns are not really applicable. Because the exercises require responses to defined situations, there is no guesswork about what is being assessed. Naturally, as with any open-ended measure, one cannot be certain that participants are responding with what they "really believe" as opposed to what they "believe the right answer is likely to be." Nevertheless, as a measure of knowledge per se, this does not present a serious threat to validity

One caution we do have, however, concerns other extraneous influences on these scores. Because the responses are in the form of open-ended essays, clearly respondents' motivations to provide narrative responses and their writing abilities will influence the

quality of their responses. Since this measure is not intended to assess that ability, perhaps alternative administration techniques might be considered. For example, a pilot study that uses interview techniques in combination with the written protocol would help illuminate the extent to which scores are byproducts of written abilities.

## Mental Models - ARI

| | |
|---|---|
| Purpose | Assess team, organization and vision mental models of leaders |
| Population | 2nd level Lieutenant to colonel, undergraduate students |
| Acronym | N/A |
| Scales | 1) Accuracy; 2) Breadth; 3) Depth; and 4) Organization of mental model |
| Administration | Individual, paper and pencil |
| Time | 2 1/2 to 3 hours |
| Price | N/A |
| Authors | Zaccaro, Marks, O-Connor-Boes & Costanza (1995) |
| Publisher | ARI |
| Comments | ARI measures mental models through concept maps. |

### Theory

Mental models are defined as symbolic representations of conceptual knowledge that exist in long-term memory at varying levels of abstraction. They contain information about the relationships that exist among various components of a specific concept. The knowledge of these relationships is in large part responsible for the ability of humans to understand phenomena, to draw inferences/make predictions, and to decide what actions to take (Rouse & Morris, 1986; Johnson-Laird, 1983). The importance of mental models for effective organizational leadership in the military is based on the premise that such leadership often requires complex social problem solving in which leaders identify key issues relevant to organizational goal attainment, and generate solutions or approaches that address these issues (Jacobs & Jacques, 1987).

Types of Mental Models. Mental models are functional cognitive representations of complex systems and their operations (Hinsz, 1995; Holyoak, 1984). Mental models are also organized constructions of information pertaining to system functioning. These models specify cause and effect, and temporal or categorical associations among concepts and system elements. Fundamentally, mental models contain the constructs, elements, and variables that effectively describe system functioning. The three general types of mental models are as follows:

1) Declarative knowledge, which includes information about the concepts and elements in a domain, and about the relationships among them (Converse & Kahler, 1992);

2) Procedural knowledge, which reflects the information about the steps that must be taken to accomplish various activities, and the order in which these steps must be taken (Converse & Kahler, 1992); and
3) Strategic knowledge, which is defined as information that is the basis of problem solving. Some examples are: 1) action plans to meet specific goals; 2) knowledge of the context in which procedures should be implemented; 3) actions to be taken if a proposed solution fails; and 4) how to respond if necessary information is absent (Converse & Kahler, 1992).

Sources of Mental Models. Mental models are knowledge structures constructed from past experience that reflect the understanding generated from those experiences. This suggests that the quality of one's mental model of a conceptual domain will depend on the richness and breadth of experiences in that domain. The mental model will be more accurate and extensively developed if any of the following situations occur:

1) an individual has repeatedly experienced a particular content domain in depth;
2) an individual has experienced related, but separate domains and concepts to acquire information about the similarities and differences with the target concept; or
3) an individual has a fundamental intellectual capacity to abstract increasingly more complex and principal-based understandings regarding the conceptual domain (Zaccaro et al., 1995).

The relationships between individuals' mental models and their experiences are moderated by their intellectual capacities to extract principal-based abstractions from prior experience. The development of mental models is also moderated by the individual's predisposition to select certain kinds of experiences. Specifically, a predisposition that reflects a strong achievement orientation, openness to novelty and change, and adaptiveness in the face of adversity and challenge should result in more enriching and rewarding experiences that serve as the basis for well-adapted mental models. This predisposition is similar to the proclivity profile that is discussed in depth in the personality section (Mumford et al., 1993).

Expert vs. Novice. Experience has a great deal to do with the construction and use of mental models making comparisons between novices and experts a natural index of measurement fidelity. For example, expert knowledge tends to be highly integrated and

tightly organized. They tend to alternate between high-level and low-level analysis as needed for problem solving. Furthermore, experts have rich, high-level abstract knowledge, which they use to select problem-appropriate general principles and specific solution plans (Cantor & Kihlstrom, 1987). In addition, expert knowledge is highly differentiated, and they can recognize a vast number of different problem-instantiated patterns. Finally, experts possess detailed causal models that allow them to diagnose problems and understand how outcomes are affected by intended courses of action (Laskey et al., 1990).

Mental Models in ARI. A key characteristic of mental models that facilitates their use in dynamic and novel situations, such as those found in the military, is that they represent flexible constructions of reality. These constructions can be extended, refined, and revised with the addition of new elements, and the integration of anomalous or unexpected events (Carlsson & Gorman, 1992). Based on this characteristic, it is essential for leaders to have mental models that are specific enough to have applicability in a particular domain, while at the same time generalizing across organizational problems. Mental models are characterized both by their content and structure. A review of leader requirements has led to the conclusion that there are three specific mental models that are essential for leaders to possess (Zaccaro et al., 1995):

1) Team mental model - containing organized knowledge about the elements, characteristics, and dynamics that influence how individuals work interdependently to perform collective tasks;
2) Organizational mental models - containing organized knowledge about key components, events, and operations of the leader's organization and environment that bears possible relevance to his or her problem solving efforts; and
3) Vision mental models - representing organized cognitive representations of contextual entities that are used to evaluate the feasibility of particular solutions and the factors necessary to address when implementing a solution.

**Development and Empirical Use**

Objectively measuring mental models is fairly challenging. There is no single method that has been universally accepted. It is difficult because the existence and properties of mental models must be inferred from behavior (Hinsz, 1989). Rouse and

Morris (1986) define mental models as varying along two dimensions: 1) the nature of the model manipulation; and 2) the level of behavioral discretion. The nature of mental model manipulation refers to the awareness an individual has of his or her manipulation of the model. The level of behavioral discretion refers to the amount of choice an individual has in task completion. Current methods of measuring mental models often are intrusive and require subjective interpretation (Converse, Cannon-Bowers, & Salas, 1991).

Some examples of the measurement of mental models include empirical modeling, analytic modeling, and verbal/written reports. These techniques have been employed with some success, but each has some inherent limitations. Empirical modeling, or inferring model characteristics by observing people's observations and subsequent responses, may only be used on simple tasks where it can be assumed that the individual is correctly perceiving the information and is therefore, not restricted in the response (Rouse & Morris, 1986). Analytic modeling, which involves constructing a "likely" model of the task based on theoretical assumptions and then comparing it to empirical data, has the limitation that in complex tasks it is difficult to specify numerous model parameters simultaneously.

Rouse and Morris (1986) have measured mental models through verbalization protocols. These methods require participants to report in some manner the content and organization of their mental model. The verbalization procedure ranges from verbal protocols to think aloud methods to surveys and questionnaires in which individuals respond to items designed to elicit declarative and procedural knowledge. The patterns of responses are then analyzed to assess mental model content and structure. A potential problem with verbalizations may be that they change the task enough significantly and thereby change the manner in which it is executed. In addition, if the task is spatial or pictorial, it may create response distortions or bias (Rouse & Morris, 1986).

ARI. A common approach to the assessment of mental models is a "known groups strategy" where the responses of domain experts and novices on problem-solving exercises or on surveys that prompt the elicitation of declarative and procedural knowledge are compared (Rouse & Morris, 1986). The responses, when contrasted between experts and novices, should provide information regarding the accuracy,

breadth, depth, and organization of the respondent's mental model. This is the strategy used in the Zaccaro et al. (1995) work. Specifically, the measure presents scenarios that describe ill-defined problems in the context of a team or organization. Participants rate the importance of various action steps presented, select the action steps that are most important to the problem, and also provide pairwise ratings of items representing concepts in a particular leader mental model.

These responses are contrasted between experts and novices to provide information on accuracy, breadth, depth, and organization. Experts' mental models will be more accurate, have greater breadth, and have stronger linkages or more complex organization between concepts in the model (Chi, Glaser, & Rees, 1982).

This research endeavor developed the problem scenarios by relying on several sources of information. First, a review of the literature on team, organizational, and leadership vision was done. Second, interviews and surveys of experts from the military, academic, and business domains were completed. These procedures led to the specification of conceptual elements in each mental model, which were then converted into problem scenarios and action steps that could be taken (Zaccaro et al., 1995).

The result was three mental model measures. The team and organizational mental model measures were formatted to a problem scenario, with a set of appropriate and inappropriate action steps. The participants were asked to rate each step in terms of importance. Each measure contained a military scenario and a business scenario. Second, they were asked to pick the five most and least important action steps. Third, they were asked to rate how each of the ten action steps were related to one another (Zaccaro et al., 1995). The mental model measure, vision, presented a scenario requiring participants to construct a vision monograph for the Army. They needed to rate 78 items for inclusion in the monograph, and then select the 10 most important statements for the "vision core."

A total of 37 Army lieutenants, 37 Army majors, 27 Army colonel, and 50 undergraduate students were used to validate the measures (Zaccaro et al., 1995). They also completed measures of intelligence and creative thinking capacities. Participants completed all three mental model measures, and their responses were rated by a panel of leadership experts.

**Psychometrics**

The average interrater reliability across the ratings was .81. The average correlation between the ratings of military raters and nonmilitary raters was .51; and the correlation between nonmilitary raters was .52. These rating correlations indicate varying levels of military knowledge and experience, as expected.

The criterion related validity of the measures was demonstrated in the analysis of the responses from the problem-solving exercises. The regression analysis on the rated quality of the responses indicated significant contributions of each model (team $R^2 = .06$, $p<.05$; organization $R^2 = .10$, $p< .05$; vision $R^2 = .04$, $p<.05$) (Zaccaro et al., 1995). The three mental models as a set explained 34% to 38% of the variance in rated solution quality across the problem exercises (Zaccaro et al., 1995).

It was also found that military experts differed from novices and undergraduate students on approximately half of the scores from the four scenarios across the team and organizational measures, based on t-tests. The experts did not show greater breadth and complexity in their responses (Zaccaro et al., 1995).

**Generalizability**

The goal of this measurement development was to construct a generic measure of leader mental models. Therefore, the measures should be generalizable to many different contexts. The only limiting factor area, in our opinion, would be the specific military scenarios contained in the team and organizational measures which might limit applicability to Army settings.

**Face Validity/Ease of Use/Transparency**

The three mental model measures are time-consuming to develop, fairly time-consuming to complete, and difficult to score. Therefore, their ease of use tends to be low. The items on the measure, in terms of the action steps, do not appear to be transparent. Based on our review of the items, the measures also seem face valid due to the context relevant problem scenarios.

## Career Path Appreciation (CPA)

| | |
|---|---|
| Purpose | Assess the level of conceptual capacity |
| Population | Managers/Army leaders |
| Acronym | CPA |
| Scores | 1) Phrase selection; 2) Symbol sorting; and 3) Work history |
| Administration | individual, cards and interview |
| Price | N/A |
| Time | Several Hours |
| Authors | Stamp (1986) |
| Publishers | ARI |
| Comments | requires knowledgeable scorer; time consuming to administer |

### Theory

Conceptual capacity is not a behavior preference, but the breadth and complexity with which an individual organizes his or her experience. It is not a disposition to act, but a level of sophistication of an individual's organizing processes and an antecedent to action. Cognitive complexity can be categorized as a trait or ability. In order to understand cognitive ability, many different types of taxonomies have been proposed. In general, the taxonomy of cognition is proposed to be a four-part model, consisting of the following: 1) metacognition; 2) generic cognitive tasks; 3) higher-order cognitive processes; and 4) component cognitive skills (Markessini, 1991). Despite previous research, there is as yet no comprehensive system for organizing the domain of cognition. No general theory that effectively compares, contrasts, and integrates the various human cognitive abilities or "learning categories" into a plausible model of human cognition exists. The following discussion will briefly review some of the different taxonomies.

Fleishman (1975), as discussed earlier in this section, developed a taxonomy that is comprised of a list of seventeen cognitive abilities and seventeen physical abilities. The list of cognitive abilities is as follows: 1) linguistics (verbal comprehension and expression); 2) creativity (fluency of ideas and originality); 3) memory; 4) problem solving/reasoning (problem sensitivity, deductive and inductive reasoning); and 5) perceptual/information processing abilities.

A second taxonomy is Mumford's General KSAO Taxonomy (Mumford, Yoarkin-Levin, Korotkin, Wallis, & Marshall-Mies, 1986). This taxonomy is said to

provide a comprehensive and general summary description of the personal characteristics likely to influence effective performance in various leadership activities.

A third framework for cognition is Elliott Jaques' Model of Cognitive Functioning. This taxonomy is derived from the SST, in which cognitive functioning is based on cognitive power and discontinuous change in cognitive states. Cognitive power is defined as "the mental force a person can exercise in processing and organizing information and in constructing an operating reality" (Jaques, 1985, p. 107). Cognition in this framework, then, involves the combination of elements into meaningful patterns.

Sternberg ( 1988) also proposed a theory of cognition that identifies three types of intelligence: 1) social intelligence (i.e., "street smarts"); 2) analytic intelligence (measured by intelligence tests like the WAIS); and 3) creative intellect. This framework varies slightly from the others in that the types of cognition are not sequential or progressive, or hierarchical.

Building on the literature of past taxonomies, such as those cited above, a preliminary taxonomy of generic cognitive tasks and higher-order cognitive skills for effective executive leadership was developed (Markessini, 1991). This taxonomy is composed of the following variables (Jacobs & Jaques, 1990):

1) Mapping Ability - the ability to build into the leader's frame of reference enough cause and effect chains to enable inference to the overarching rules and principals that pertain to the organizational system at this level. The requirements for mapping ability increase by organizational level;
2) Problem Management/Solution – a generic skill that subsumes critical inquiry, self-knowledge, and communication. The executive approach is to "develop a workable course of action and then to manage the outcome over time so that it will be successful;"
3) Long Term Planning- the ability to develop effective and executable plans, particularly in innovative and nontraditional modes; and
4) Creative Thinking - time spent seeking to invent, design, and develop possible courses of action for handling situations.

Conceptual capacity is a description of the nature of the meaning-making process of the objective, real world to an individual. It consists of the following two key

variables: 1) the extent to which an individual can discriminate variables; and 2) the extent to which an individual can hold different variables simultaneously in his or her mind (Jaques, 1976). Conceptual ability is thought to develop through an invariant series of hierarchical, ordered stages or levels. Individual differences in conceptual capacity are thought to represent differences in developmental level (Jaques & Clement, 1991). The assessment of cognitive complexity deals with determining how individuals think. Therefore, individuals need to be assessed when engaged in a task that demands the demonstration of their conceptual capacity, such as in the CPA. SST suggests that the most fundamental individual difference variable that most often distinguishes successful strategic leaders from unsuccessful ones is the extent to which leaders' conceptual capacity meets or exceeds the conceptual demands inherent in their work (Lewis & Jacobs, 1992).

Specifically, in terms of the strata introduced in the SST, Streufert's early conceptualization of cognitive complexity can be used to explain its current definition. The tasks in the production domain are procedurally specific operations dealing with tangible things. The operations can involve linear pathways, and may require little in the way of abstraction. As there is movement in the organizational levels, the scope and scale of performance requirements are qualitatively different, and the complexity is greater. First, time frames are much longer. Second, there is the existence of multiple functions and subsystems. Third, managers at this level must deal with intangibles. Therefore, individuals must have a more complex cognitive map with which to pattern events, assign plausible causality, and develop strategies to influence outcomes. Finally, in the strategic domain, the complexity is even greater. The extended time frames required for the execution of long term acquisitions and developments preclude successful performance through abstract thinking and analytic skills alone. Individuals must also be concerned with broad political, economic, socio-cultural, and technological developments. Synthesis, similar to Streufert's concept of multidimensional integration, appears to work in this domain.

In terms of applying SST to the leadership domain, Jacobs and Jaques (1987) suggested three sets of leadership skills that are generic across organizational levels, but should vary in importance or use at the different levels. The first set of skills is

interpersonal, which are used to facilitate communication with a diverse set of external constituencies. The second set of skills is technical, which are directly related to the task at hand. The third set is conceptual skills, which include long-term planning, the ability to balance and integrate multiple business strategies, and skill in environmental analysis and interpretation. Leader effectiveness is a function, in part, of how well a frame of reference provided by a leader patterns the causal and other mechanisms in the environment (Jacobs & Jaques, 1987). The development of appropriate frames of reference requires effort. An individual's inclination to engage in reflective thinking and cognitive model building is included in the notion of proclivity discussed previously in the personality section.

In addition, metacognition is a skill that involves choosing and planning what to do, and monitoring what is being done. There are four main skill-related processes related to metacognition. The first is defining the nature of the problem to be solved. This includes awareness that a problem exists, identification and definition of the problem, and construction of its parameters. The second process is specifying the most appropriate solution paths. The third process is the implementation of the chosen solution, and the fourth is the evaluation of the solution and its consequences (Mumford et al., 1989).

Another characteristic related to behavioral complexity suggests that effective managers are not only cognitively complex, but are also able to perform a diverse set of roles and skills in the explicit behavioral realm. Effectiveness requires not only cognitive complexity within the individual, but also the ability to act out a wide array of roles in the interpersonal and organizational arena. Managers high in behavioral complexity will be able to perform many different roles, and will also be able to strike a good balance among the roles. The following are some common managerial roles (Hooijberg & Quinn, 1992):

1) innovator - creative, clever;
2) producer – task-oriented, work-focused;
3) director - decisive, directive;
4) coordinator - dependable, reliable;
5) monitor - technically expert, well-prepared.

The generic, cognitive tasks considered critical to and distinctive of effective functioning differ at varying levels. At the highest executive levels, the most crucial cognitive abilities are mapping ability, problem management/solution, long-term planning, and creative thinking (Nickerson, 1990).

CPA Assessment. The CPA technique primarily employs an interview methodology to assess an individual's current level of conceptual complexity. Based on the results, a maturation curve is constructed that predicts the individual's maximum attainable level of capacity and work level. The end result is an index of current and potential cognitive work capacity.

The first of three tasks in the CPA is the phrase selection task. For this task, participants are given nine sets of six cards with each one describing an approach to solving a problem or work assignment. Each set reflects six work levels proposed by SST (Stamp, 1986). Participants then pick the card that reflects their most and least comfortable approaches to work, and then explain their choices. The following are the six approaches: 1) work to a complete set of instructions; 2) work within a given framework; 3) work with connections when particular links are unclear; 4) work in abstracts and concepts; 5) work with a minimum of preconceptions; and 6) define the horizons of the work (Stamp, 1986).

The second task in the CPA is the symbol sorting task (Bruner, 1966). In this task, the participants are presented with four target cards, three with geometric symbols and the fourth one blank. They are then given a pack of symbol cards and asked to sort them under the four target cards by using self-developed sorting rules. Success on this task requires abstracting and conceptualizing the appropriate sorting rules.

The third part of the CPA is the work history interview where participants provide information regarding their prior, as well as current work positions and assignments.

The results from the three tasks are analyzed to place the participant in one of seven levels, each having categories of high, medium, and low with a range of scores from 1 to 21.

**Development and Empirical Use**

The CPA was initially tested on a multinational oil company with 84 respondents, a multinational engineering company with 35 respondents, a fertilizer company with 38 participants, and the management of a mining company in a developing country.

**Psychometrics**

Preliminary psychometrics suggested that the instrument is reliable. In a study where the CPA was given to two classes of colonels at the AWC, interrater reliability between assessors was .81. The Cronbach coefficient alpha for the responses across the nine sets of cards was .78, and .76 for the symbols section (Lewis, 1993). One potential concern with the instrument is construct validity. The work history interview was designed to assess an interviewee's degree of comfort in the level of work complexity required of prior positions. These prompts may reflect a number of qualities in addition to conceptual skills (e.g., mastery), and achievement motive (e.g., openness, tolerance of uncertainty), and flexibility (Zaccaro, 1996).

The construct and predictive validities were examined by comparing CPA scores to the following items: 1) Kegan's breadth of perspective concept; 2) instructor ratings of a student's strategic thinking skill; 3) general officer potential; and 4) peer popularity. Lewis (1995) found significant correlations with breadth of perspective, strategic thinking skill, and general officer potential. CPA scores were not correlated with peer popularity. These results suggested that the CPA may be tapping two constructs: 1) a construct reflecting a willingness or proclivity "to tolerate ambiguity and deal with complex environments" (McIntyre, Jordon, Mergen, Hamill, & Jacobs, 1993); and 2) a construct reflecting conceptual capacity.

Stamp (1988) provided evidence for predictive validity from a sample of 182 managers in four different organizations. Growth curves were calculated and compared to the actual level attained by managers 4 to 23 years later, with correlations ranging from .70 to .92.

**Generalizability**

Since this type of measure tends to be situation-specific, it has very low generalizability.

**Face Validity/Ease of Use/Transparency**

Overall, the instrument is very time-consuming, which limits its use. It also requires highly skilled individuals to administer and to score it. This fact further limits the use of the instrument. However, preliminary findings show that it is psychometrically sound, and may tap more than just conceptual capacity. McIntyre et al. (1993) suggested that the CPA might reflect two distinct constructs. One construct reflecting a person's level of conceptual capacity, and another tapping proclivity in the sense of being able to tolerate ambiguity in a complex environment. The CPA is conceptually multi-componential, reflecting more than one construct. However, there is a lack of clarity regarding the validity of each of the component constructs. In terms of ease of use, it is highly dependent on the administrator.

## Tacit Knowledge for Military Leadership Inventory

| | |
|---|---|
| Purpose | Assess a leader's tacit knowledge |
| Population | Battalion commanders, platoon leaders, company commanders |
| Acronym | TKMLI |
| Scores | 5-20 ratings on work-related situations |
| Administration | individual |
| Price | N/A |
| Time | Varies depending on number of questions |
| Authors | Horvath, Forsythe, Sweeney, McNally, Wattendorf, Williams, & Sternberg (1994) |
| Publishers | ARI |
| Comments | requires expert profile to score |

### Theory

Tacit knowledge describes that which is generally acquired on one's own through personal experience rather than through instruction. It is knowledge that people may not know they possess and/or may have difficulty articulating. Like much of expert knowledge, tacit knowledge guides behavior without being readily available to conscious awareness. Finally, tacit knowledge is action-oriented knowledge, with practical value to the individual. Unlike most disciplinary knowledge, it is knowledge that helps people pursue goals that they may personally value.

A second conceptualization of tacit knowledge treats it as a cognitive phenomenon, defining it in terms of the learning processes that produce it and the memory structures/systems that encode it. This is the explanatory model that distinguishes episodic and semantic memory. Episodic memory is defined as memory for specific, personally-experienced events; memory for the episodes that compose one's experience. Semantic memory is defined as memory for general, impersonal knowledge; memory for information that transcends specific episodes. According to the models of inductive learning, the transition from event knowledge to generalized knowledge involves mental processes that are sensitive to the covariance structure of the environment. These processes share features and/or structures across episodes, and construct abstraction or general representation of that shared structure.

The hallmark of practical intelligence is the acquisition and use of tacit knowledge. Tacit knowledge is practical know-how that usually is not openly expressed

or stated, and which must be acquired in the absence of direct instruction (Wagner, 1987). The scope of tacit knowledge refers to the range of situations to which tacit knowledge may be applied. This scope can be categorized in three ways. The first is the content of the situation, such as whether it primarily involves managing oneself, managing others, or managing one's task. The second is the context of the situation, in terms of whether it is local (short-range, self-contained) or global (long-range, "big picture") in nature. A third way is the orientation of one's focus, either idealistic or pragmatic (Wagner, 1987).

Managing oneself in the content domain refers to knowledge about self-motivational and self-organizational aspects of performance in work-related situations. Tacit knowledge about managing tasks refers to knowledge about how to perform specific work-related tasks well. The third type of content-based tacit knowledge, managing others, refers to knowledge about managing one's subordinates and one's interactions with others (Wagner, 1987).

A local context refers to a focus on short-term accomplishments of a specific task at hand. No consideration is given to one's reputation, career goals, etc. A global context refers to a focus on long-range objectives, and on how the present situation fits into the larger picture. Real world accomplishments require practical knowledge that can be applied in both local and global contexts (Wagner, 1987).

An idealistic orientation focuses on how good a solution is in isolation. The quality of some course of action is judged without regard as to how practical or impractical it might be. A pragmatic orientation refers to how workable a potential solution is. Effective performance requires knowledge relevant to both orientations (Wagner, 1987).

Academic intelligence refers to the abilities typically valued in schools. These abilities include reading or listening to formal, explicit instruction on the content and rules of a given discipline, as measured by conventional intelligence tests. Practical intelligence refers to abilities typically devalued in schools. These abilities involve observing, imitating, and applying the informal, unspoken strategies that lead to success in real world pursuits. Practical intelligence is the ability to learn about, rather than of, a

discipline, and it is poorly measured by conventional ability tests (Sternberg, 1985; Sternberg & Wagner, 1993).

There are three characteristic features of tacit knowledge: 1) procedural structure; 2) high usefulness; and 3) low environmental support for acquisition. Tacit knowledge can be described at three levels of abstraction. The lowest level is described as mentally represented knowledge structures. These knowledge structures take the form of complex, condition-action mappings. It is at this level of description that tacit knowledge has psychological reality and its consequences for intelligent behavior. It is necessary to infer tacit knowledge from subjects' behavior and articulated knowledge. It is at this level that items are used to elicit and record individuals' tacit knowledge.

At a higher, more abstract level of description, tacit knowledge items can be grouped together into categories of functionally related items. Category level description adds value to the identification of tacit knowledge by illuminating the broad functional significance of different aspects of tacit knowledge. Tacit knowledge is important for adapting to, selecting, and shaping one's external environment. Adapting to the environment means modifying one's behavior to meet the requirements of that environment. Tacit knowledge can play an important role in such adaptation. If the individual is unwilling or unable to adapt, and must instead find a new context in which to pursue success, a new environment is selected and tacit knowledge may be essential. Sometimes individuals neither adapt to a particular feature of their environment nor select another in which to pursue success. When this occurs, they may act to modify the environment rather than their own behavior.

Tacit knowledge has repeatedly been found to increase with experience in a domain. Even when the level of experience is held constant, tacit knowledge scores have been found to predict job performance according to a variety of criterion measures. Williams and Sternberg conceived of tacit knowledge for business management with the three domains of intrapersonal, interpersonal, and organizational. The intrapersonal domain encompasses four aspects of tacit knowledge. Challenge orientation refers to the propensity for choosing and enjoying situations that represent a challenge; situations that require breaking of new ground, and the learning of new areas and skills. Control orientation refers to the tendency to take charge of the situations and to place oneself in

control. Self-oriented personal effectiveness refers to the degree to which one is effective within the self. Context-oriented personal effectiveness refers to the degree to which one is effective in the context of compromising tasks and environment.

The interpersonal domain of tacit knowledge consists of knowledge about behaviors that relate to others. There are three categories: 1) influencing and controlling others; 2) supporting and cooperating with others; and 3) understanding others in terms of superiors, subordinates, and peers. The organizational domain of tacit knowledge consists of: 1) knowledge about behaviors relating to the organization; 2) optimizing the system by evaluating people and jobs in the system; and 3) matching people to jobs and tasks to create the most functional system. The second area is defining the organization as to the acts involved in articulating and locating challenges the system is best equipped to handle. It entails reviewing and choosing products and services that the organization will offer and excel at, and that the marketplace will receive positively. The third category refers to envisioning the future by analyzing the marketplace in general, and the strengths and weakness of the company in particular.

The structure of the tacit knowledge domain in military leadership consists of the same three dimensions, however the specifics under each vary slightly. For intrapersonal tacit knowledge, the leader must manage themselves in terms of: 1) organizing himself or herself; 2) managing time and priorities; 3) seeking challenges and control by taking initiative; and 4) taking responsibility and acting to increase one's discretion. In interpersonal tacit knowledge, the individual needs to: 1) influence and control others; 2) support and cooperate with others; and 3) learn from others. Finally, organizational tacit knowledge requires that the individuals solve organizational problems.

The following list integrates three different samples and results from our review of the ARI literature on the three tacit knowledge domains:

1) Intrapersonal Tacit Knowledge
   managing the self (b, c, p)
   seeking challenges and control (x)

2) Interpersonal Tacit Knowledge
   influencing and controlling others
   motivating subordinates (b, c, p)

directing and supervising subordinates (c)
influencing the boss (c, p)
developing subordinates (c)
communicating (p)
supporting and cooperating with others
taking care of soldiers (b ,c, p)
establishing trust (b, c, p)
cooperating with others (c)
learning from others (x)

3) Organizational tacit knowledge

solving organizational problems
communicating (c, p)
developing subordinates (b)
dealing with poor performers (b)
managing organizational change (b)
protecting the organization (b)

b = obtained from battalion commanders
c = obtained from company commanders
p = obtained from platoon leaders
x = obtained from literature review only (Horvath, Forsythe, Sweeney, McNally, Wattendorf, William, & Sternberg, 1994)

**Development and Empirical Use**

The empirical research in this area focuses on individual differences in the ability to acquire and use tacit knowledge, as well as on the consequences of those differences for performance in knowledge-intensive disciplines. Tacit knowledge can be effectively measured by employing work-related situations with between five and twenty response items. Each situation poses a problem, and the participant indicates how he or she would solve it by rating various responses. The set of ratings the person generates for all of the work-related situations is the measure of his or her tacit knowledge for that domain.

Tacit knowledge tests are knowledge-based tests built on a theory of human intelligence. They are intended to measure practical, experience-based knowledge, as

well as the underlying dispositions or abilities that support the acquisition and use of that knowledge. Tacit knowledge items are both indicators and exemplars of underlying tacit knowledge. These items can potentially shed light on the content of that knowledge, and the events or experiences through which it was acquired. Tacit knowledge tests are a hybrid of achievement tests and ability tests. Thus, they differ somewhat in construction and validation. There are no objectively right answers, and reference to an expert response profile is required.

Content validity for the items was assessed through interviews with the participants. They were oriented toward personal experiences and away from leadership theory and doctrine. The generalizability of tacit knowledge tests calls for generalization across roles within the organization, repeated administrations, and alternate forms of the test. By seeking to specify and measure the construct rather than merely pursue correlations with external criterion, it allows the test to be more generalizable. In the context of tacit knowledge tests, potential discriminate evidence would be with general intelligence, reading comprehension, and general job knowledge, and in the convergence of these scores with external indices of performance.

Tacit knowledge has been found to increase, on average, with job experience. However, it is not a direct function of job experience (Sternberg et al., 1993). The emphasis is not on the quantity of experience the person has, but on how well the person utilizes the experience to acquire and use tacit knowledge. Tacit knowledge almost never correlates significantly with IQ, and is not a proxy for measures of personality, cognitive style, or interpersonal orientation. The contribution of tacit knowledge to prediction of criteria indices was still significant after holding all other variables constant.

The dimensions of tacit knowledge for the **Battalion commander** are as follows:

1) communicating a vision - communicating goals by describing a future end state; including in that message issues of character, moral fortitude, and tough love;
2) establishing a climate for development - communicating a set of beliefs or attitudes that allows subordinate development; reinforcing the statements by providing a structure of activities that supports such a development;

3) managing the leader and the subordinate - managing oneself while simultaneously "managing by exception" the problems that occur within the organization; considering the actions the leader should take to establish subordinate trust in the culture/climate/ vision that has been communicated;

4) providing constancy - providing stability by reinforcing the desired end state at every opportunity; communicating and maintaining a uniform "commander's intent"; and

6) using influence tactics - providing structure that allows subordinates to achieve desired levels of performance; maintaining authority by employing the full range of influence tactics; establishing parameters (in the form of formal controls) that reinforce subordinates trust in core values.

The dimensions of tacit knowledge for **Company Commanders** are as follows:

1) caring for soldiers through task completion - knowing your job and making subordinate soldiers "do the right thing" (in terms of training readiness and task accomplishment);

2) prioritizing and solving problems - dealing with day to day problems; communicating priorities and providing guidance to solve problems;

3) proactive decision making - thinking ahead to anticipate problems; sharing information so that subordinates can assist in proactive problem solving;

4) assessing risk - determining the potential liabilities of an action; using team building to identify and potentially reduce hazardous situations; and

5) short term decision making - providing face-to-face directions to influence an action at a critical moment; making decisions that facilitate day-to-day operations.

The dimensions of tacit knowledge for **Platoon Leader** are as follows:

1) acquiring confidence in interpersonal skills - learning how to motivate subordinates; overcoming individual hesitancies toward motivating more experienced soldiers;

2) defining leadership style - understanding one's personal leadership style; knowing the type of influence to use in one-on-one situations;

3) taking a stand - confidently demonstrating concern for the unit's welfare with subordinates; being forthright when discussing the strengths and weaknesses of the unit; and

4) taking and fostering accountability - identifying problems (interpersonal or technical) within the unit and proactively seeking solutions to the problem; requiring the same actions of subordinates.

In order to develop this test, interviews were completed first to determine the content of the tacit knowledge items. Leadership knowledge was elicited in semi-structured interviews from active duty Army officers around the U.S. These respondents were drawn from three branches of the army: 1) combat arms; 2) combat support; and 3) combat service support from three different levels (e.g., platoon, company, and battalion leaders). The interviewers asked the participants to tell a story from which they had learned something about leadership that was not taught in class. After the interviews were conducted, tacit knowledge contained in the interview summaries was identified and coded by two researchers. The degree of interrater reliability was 73%. Each story was then annotated with a preliminary coding of the tacit knowledge. These summaries were then given to three senior military members with research experience for tacit knowledge content consensus. The items were then sorted into battalion commander, platoon leader, and company commander tacit knowledge areas.

Tacit knowledge items were analyzed with TRADOC data. The findings showed that experienced and novice leaders at each of the levels displayed the expected significant differences in terms of tacit knowledge. This suggested that the knowledge items in the tacit knowledge survey hold promise for development into tests that are fairly discriminating. Tacit knowledge was also analyzed against FORSCOM data, and a significant relationship between item ratings and leader effectiveness for a number of items at each level was found.

**Psychometrics**

Tacit knowledge predicted job performance moderately well, correlating .3 to .5 with performance measures, which compares favorably with those obtained for IQ measures (Sternberg et al., 1993).

A discriminant analysis using a TRADOC sample provided support that novice and experienced leaders responded differently to the tacit knowledge items on the instrument. The canonical correlation coefficients were $R = .73$, $p<.05$; $R = .72$, $p<.05$; $R = .55$, $p<.05$, for battalion, company, and platoon level data, respectively.

Content validity was considered fairly well during the development of the instrument. This was accomplished by interviewing Army officers and obtaining goodness ratings on tacit knowledge items. More construct and criterion-related validity studies are in progress. More conclusive evidence bearing on substantive and generalizability aspects of validity is needed. A scoring key is also currently in progress.

**Generalizability**

Tacit knowledge researchers suggest that score interpretations need to generalize across roles within the organization, repeated administrations, and alternate forms of the tests. They believe that generalization is concerned with test development in terms of content and structure of the items (Horvath et al., 1996). While we acknowledge the importance of this emphasis, it does not render the traditional concerns about generalizability moot. Tacit knowledge tests need to have a target population in mind just as any other form of test does. An issue here is whether a test, once constructed, would be useful for different jobs, in different settings, performed by different individuals, etc. These concerns are important in the development phase of a test as they drive how questions are framed, who constitutes SMEs, etc. That said, the three forms of tacit knowledge (i.e., Battalion and Company Commanders, and Platoon Leader) appear to be widely generalizable within those domains. The overlap across domains, however, appears to be very limited suggesting a natural boundary for generalizations.

**Face Validity/Ease of Use/Transparency**

Tacit knowledge measures, by their very nature, appear to be face valid to respondents. Less clear, however, are the scoring keys as referenced to experts' consensus ratings. The use of such a referent has an implicit assumption that there exists "a" best why or responding to a situation. It becomes difficult to develop a consensus regarding the appropriateness of one or a set of alternatives without making it fairly transparent. Moreover, the concept of "equifinality" – that there might be more than one way to be successful, is not acknowledged. The development of tacit knowledge measures is a time intensive effort, but once established, they are relatively easy to administer and score.

# Benchmark Instruments

### Watson-Glaser Critical Thinking Appraisal

| | |
|---|---|
| Purpose | Assess critical thinking skills |
| Population | Grades 9-12, adults |
| Acronym | N/A |
| Scales | 1) Inference; 2) Recognition of assumptions; 3) Deduction; 4) Interpretation; 5)Evaluation of arguments; 6) Total score |
| Time | 40 to 50 minutes |
| Price | $40 per 35 test booklets and manual; $10.50 per 35 Opt Scan sheets |
| Administration | Individual, paper and pencil, computer scored |
| Authors | Watson & Glaser (1964) |
| Publishers | Harcourt, Brace and World |

**Theory**

This instrument measures five subtests, which reflect the authors' views of critical thinking. They are: 1) inference; 2) recognition of assumption; 3) deduction; 4) interpretation; and 5) the evaluation of argument. These dimensions are tapped through reading. The exercises were developed to include problems, statements, arguments, and interpretation of data encountered on a daily basis at work, at school, or in literature (Watson & Glaser, 1964).

**Development and Empirical Use**

The current forms, A and B, are composed of 80 items per form. A total of 134 of these items were drawn from the previous versions of the instrument, the Ym and Zm.

The norms for high school students are based on a sample of 24 high school districts in 17 states, with attention to geographic region, size, socioeconomic status, sex and race. Similar samples were used for the development of college and business norms.

**Psychometrics**

The most recent forms A and B possess split-half reliability coefficients ranging from .69 to .83. The test-retest at a three-month interval is .73.

Validity was determined through construct and content analysis in the Watson & Glaser manual, although specific details were not given (Watson & Glaser, 1964). In a evaluation of the validity of the Watson-Glaser based on the ten essential validity standards from the Standards for Educational Psychology in 1974, Modjeski & Michael

(1983) found the instrument respectable. Twelve Ph.D. level psychologists determined that the instrument had high criterion-related validity in terms of development, however bias in the tests is possible.

**Generalizability**

Based on the wide range of use of this instrument, generalization is expected to be high.

**Face Validity/Ease of Use/Transparency**

Items were specifically written to have face validity (Watson & Glaser, 1964). The instrument is easy to administer due to its short length, paper and pencil format. Remote or computer scoring is available.

## Concept Mastery Test

| | |
|---|---|
| Purpose | Measures meta-cognitive processes and skills involving the manipulation of abstract concepts and ideas, as well as the complexity and interrelatedness of conceptual categories possessed by the individual. |
| Population | Advanced college students, adult |
| Acronym | N/A |
| Scales | N/A |
| Time | 35 - 45 minutes |
| Price | unknown |
| Administration | Individual, paper and pencil, |
| Authors | Terman & Olden (1959) |
| Publishers | Psychological Corporation |

### Theory

The Concepts Mastery test is a by-product of Terman's extensive studies from gifted children. It was developed to provide a good deal of information on a person's ability to deal with abstract concepts in a limited amount of time (Terman & Oden, 1959).

### Development and Empirical Use

This test is a high-level verbal test that contains two type of items. The first type of items is standard synonym-antonym items, which are constructed with rather unusual vocabulary. The second part of the test are items of analogy type, using number and verbal problems covering general knowledge and relationships between terms (Terman & Oden, 1959).

### Psychometrics

The correlation between the two parts of this test is .76 on a sample of the Stanford Gifted Study. Generally, reliability is found to lie between .86 and .94. Test-retest correlations for a twelve-year span are .90 (Terman & Oden, 1959).

The test distinguishes clearly between adults of different education levels, showing discriminate validity. It has also successfully shown predictive validity in university courses. The test has correlated moderately with the Owens-Bennett Test of Mechanical Comprehension, the Test for Productive Thinking, and the Test for S lecting Research Personnel. Every score from the above three tests had significant validities with the supervisor's creativity rating (Terman & Oden, 1959).

**Generalizability**

The test has been mainly used with advanced college populations, however it has also been used with adults who are being considered for research, executive, and other unusually demanding jobs.

**Face Validity/Ease of Use/Transparency**

The test is paper and pencil, with scanned scoring for ease of use. However, the items themselves may not seem face valid to respondents.

## Consequences

| | |
|---|---|
| Purpose | Assesses both ideational fluency and originality as components of divergent thinking skills |
| Population | Grades 9 to 16, adults |
| Acronym | N/A |
| Scales | 1) Fluency; 2) Originality |
| Time | 20 to 30 minutes |
| Price | N/A |
| Author | Guilford & Guilford (1980) |
| Publisher | Sheridan Supply Company |

### Theory

This test was developed to systematically explore the structure of the intellect and isolate what creative thinking is (Guilford & Guilford, 1980).

### Development and Empirical use

This instrument consists of ten items requiring the participant to list what the result may be if some unusual situation came to pass. Relevant, non-duplicated responses are classified as "obvious" or "remote." The frequency of "obvious responses" yields a score of fluency. The frequency of "remote" responses are originality scores (Guilford & Guilford, 1980).

### Psychometrics

Internal consistency reliability on the obvious score was .86 for a ninth grade sample. The remote score for the same sample was .67 (Fredericksen & Evans, 1974).

Construct validity has been shown by factor analysis. The obvious score has an average validity of .62 for the factor ideational fluency, on the basis of five samples of approximately 1,000 young adult males. A total of 29 to 38% of the score variance is attributable to this one factor (Guilford & Guilford, 1980).

### Generalizability

This instrument has been used on a wide range of college and adult samples, so generalizability is high.

**Face Validity/Ease of Use/Transparency**

Although the test is easy to administer, there have been some questions about scoring in terms the decision point of remote and obvious (Guilford & Guilford, 1980). One suggestion is the development of a scoring protocol to provide a standard scoring system. In our opinion, the items are neither face valid nor transparent.

## Leatherman Leadership Questionnaire

| | |
|---|---|
| Purpose | Aid in selecting supervisors, provide feedback on leadership knowledge |
| Population | Managers, supervisors, and prospective supervisors |
| Acronym | LLQ |
| Scores | 1) Assigning Work; 2) Career Counseling; 3) Coaching Employees; 4) Communication; 5) Managing Change; 6) Handling Employee Complaints; 7) Dealing With Employee Conflicts; 8) Counseling Employees; 9) Decision Making; 10) Delegating; 11) Discipline; 12) Handing Emotional Situations; 13) S tting Goals/Planning; 14) Grievances; 15) Conducting Meetings; 16) Feedback; 17) Negotiating; 18) Performance Appraisal; 19) Establishing Performance Standards; 20) Persuading; 21) Presentations; 22) Problem Solving; 23) Conducting Selection Interviews; 24) Team Building; 25) Conducting Termination Interviews; 26) Helping Employees Manage Time; 27) One On One Training. |
| Administration | Individual and group |
| Price | Set of 12 overhead transparencies, manual, 10 sets of booklets, answer sets and scoring service for $600 |
| Time | 5 hours for complete test, 2 1/2 hours per part |
| Authors | Richard W. Leatherman (1987) |
| Publisher | International Training Consultants, Inc. |

### Theory

This instrument was designed to be a knowledge-based measure of supervisory leadership for selection and feedback purposes. The theory states that there are 27 skills that a leader needs to be effective. These skills are the following: 1) assigning work; 2) career counseling; 3) coaching employees; 4) communication; 5) managing change; 6) handling employee complaints; 7) dealing with employee conflicts; 8) counseling employees; 9) decision making; 10) delegating; 11) discipline; 12) handing emotional situations; 13) setting goals/planning; 14) grievances; 15) conducting meetings; 16) feedback; 17) negotiating; 18) performance appraisal; 19) establishing performance standards; 20) persuading; 21) presentations; 22) problem solving; 23) conducting selection interviews; 24) team building; 25) conducting termination interviews; 26) helping employees manage time; 27) one on one training (Katkovsky, 1992).

### Development and Empirical Use

Supervisory tasks were identified through a literature review. Next, experts developed items, and constructed scales by placed these items into the dimensions.

The instrument yields two feedback reports; one benchmarks the organization on the 27 tasks versus other organizations. A second report is generated for each individual to present and compare his or her scores with other respondents in the organization, as well as with the international averages. The individual's strengths and needs are identified via this report. The instrument has 339 items in a multiple-choice format. The feedback report provides detailed information concerning each leadership task in terms of strengths and weaknesses, by comparisons against others in the organization and the population of previous participants (Katkovsky, 1992).

**Psychometrics**

The internal consistency reported for the LLQ based on Kuder-Richardson's formula 20 was .97. However, the correlations and reliabilities of the individual scales were not presented to allow assessment of the distinctiveness of each task. Given the high internal consistency, the measure may tap only one factor instead of the 27 different skills that were proposed (Katkovsky, 1992).

The content validity was established by agreement of six out of eight expert panel members on the importance of the tasks and assignment of items into scales. There is some concern for the construct validity of the scale. In a study of 229 participants from seven organizations, significant task differences were obtained across jobs. These differences suggest that there is not likely to be a single universal "best fit" profile of requisite skills across jobs. Concurrent criterion-related studies with the LLQ and assessment center scores show inconsistent results, with one study showing no significant relationships and a second study finding overall significant rhos for three different samples (Katkovsky, 1992).

**Generalizability**

The questionnaire taps supervisory content, so the instrument should generalize to any setting where leadership is being assessed.

**Face Validity/Ease of Use/Transparency**

The entire instrument takes approximately four to five hours to complete, limiting its use. The administration and scoring of the results are completed electronically. The items appear to be face valid, and vary in transparency in our opinion.

## Mental Models via Paired Comparisons & PathFinder

| | |
|---|---|
| Purpose | Assess the structure of mental models |
| Population | Varied, student, instructor, pilots, trainees |
| Acronym | PF |
| Scales | 1) Structure of mental model |
| Administration | Individual, paper and pencil |
| Time | 1/2 to 3 hours |
| Cost | N/A |
| Authors | Stout, Salas & Kraiger, 1997; R. W Schvaneveldt (1990) |
| Publisher | N/A |

### Theory

This approach to assessing mental models begins with a thorough analysis of the leadership task and an identification of the critical job facets or activities. Assuming one has a manageable number of such facets, similarly, relationship, or importance ratings are gathered for all potential facet pairs. This matrix of ratings is then analyzed using a network analysis algorithm (e.g., Path-Finder) to yield representations of cognitive structures. The items in the network are represented as nodes, and the associations between items are represented as links between nodes. Only those concepts that are closely related are connected by links in the PF algorithm. As a result, the PF represents complex conceptual relations in a simple fashion (Mohammed, 1995).

### Development and Empirical Use

Content scenarios are developed, similar to those presented in the ARI mental model write-up. Once the content scenarios are constructed, respondents assign a rating reflecting a judgment of relatedness or similarity to all possible pairs of N concepts on some scale. Proximity estimates are then analyzed by the PF algorithm (Schvaneveldt et al., 1985).

The output of PF is PFNET which is determined by the values of two parameters: 1) r (how the weight of each link is determined); and 2) q (limits the number of links allowed in paths). Links between concepts may be weighted to represent the strength of the relatedness of two concepts (Schvaneveldt, Durso, & Dearholt, 1989).

PF does have a standard, accepted procedure. When collecting paired comparison data, researchers use 30 or fewer concepts, and a 7 to 9 point rating scale (Schvaneveldt

et al., 1989).

**Psychometrics**

Goldsmith and Johnson (1990) found that repeated ratings of the same set of concept pairs correlated an average of .60. PF has also been found to predict free recall order and category/dimensional judgment time (Cooke, Durso, & Schvaneveldt, 1986; Cooke, 1992). There is also evidence that PF differentiates between experts and non-experts (Cook, 1992). In a study of the sampling Navy aviators, Stout et al. (1997) found that a structured training program had a significant impact on trainees' mental models which, in turn, related positively to their performance. Therefore, it is our opinion that the reliability, criterion-related and construct validity evidence of the PF is fairly supportive

**Generalizability**

This approach to assessing mental models is considered as "mixed" in terms of generalizability. On one hand, in order to yield grounded results, identifying the critical facets to be rated is a context specific effort. On the other hand, the assessment procedures and analytic techniques are generic once the dimensions or inquiry have been identified. Notably, we would suggest most applications of pair-comparisons for measuring mental models have <u>less</u> generalizability as compared to the methods employed in the ARI research (i.e., Zaccaro, et al., 1995).

**Face Validity/Ease of Use/Transparency**

Although somewhat time consuming to complete (and this depends primarily on the number of facets being rated), the measures are easy to administer. As for face validity, respondents' often report skepticism regarding the value of the information they are providing. This follows from the fact that their mental models are derived in an emergent fashion through the network analysis and may yield knowledge structures that the respondents were not even aware of.

## Low Fidelity Simulation

| | |
|---|---|
| Purpose | Sample behaviors that provide signs of underlying ability, temperament, and/or other traits presumed necessary for performance. |
| Population | Managers |
| Acronym | N/A |
| Scores | Dimensions based on specific job analyses |
| Administration | Paper and pencil, individual |
| Price | N/A |
| Time | Varies depending on number of questions |
| Authors | Motowidlo, Dunnette, and Carter (1990) |
| Publishers | N/A |

### Theory

This measure draws from the theory of behavioral consistency, in that past performance is the best indicator of future performance (Wernimont & Campbell, 1968). Motowidlo, Dunnette, and Carter (1990) argued that the low fidelity simulation can be more useful for predicting job performance than predisposition signs, such as, standard ability, personality and other measures. This approach is also grounded in a tacit knowledge framework and assesses the extent to which respondents can detect what the best course of action of a given situation is likely to be.

### Development and Empirical Use

Latham, Saari, Pursell, & Campion's (1980) work with situational interviews guided the development of the low fidelity simulation. There was an emphasis placed on critical incidents for the specific job. From these critical incidents, task descriptions could be formulated by SMEs. Then, the scoring process could be developed.

First, job analyses for managers in seven companies in the telecommunications industry were reviewed. Second, people representing all seven participating companies were interviewed in small groups to collect critical incidents of managerial effectiveness and ineffectiveness. Approximately 1,200 written critical incidents were collected, which were used to write brief descriptions of task situations. Third, SMEs were asked to write responses to how they would react to the situations effectively. On the basis of the responses, the researchers developed five to seven general strategies for each task situation. Next, a group of senior managers evaluated the effectiveness of the alternate

strategies and identified the best and worst alternatives. A total of 58 situational effectiveness questions remained after evaluation with an average intraclass correlation for the ratings of effectiveness from senior managers being .95.

**Psychometrics**

Three samples completed the simulation: 1) incumbents hired into management positions outside the company; 2) incumbents promoted to management positions from inside the company; and 3) applicants for management positions who were not incumbents from the sample.

Significant validity estimates of the externally hired incumbents were found: 1) .35 with ratings of interpersonal effectiveness; 2) .28 with ratings of problem-solving effectiveness; 3) .37 with the ratings of communication effectiveness, and 4) .30 with the ratings of overall effectiveness. For the internally promoted sample, significant correlations were found for ratings of problem solving effectiveness and communication effectiveness.

Scores were also correlated with other variables, such as those from assessment centers. For the applicant sample, GPA significantly correlated .30 with the simulation scores. Other significant correlations included: 1) oral fact finding (.30); 2) interpreting information (.41); and 3) writing fluency (.31). These results offer preliminary support that the simulation may tap important cognitive skills measured by aptitude tests or academic achievement scores.

In follow-up studies, Motowidlo & Tippins (1993) found predictive validity estimates of situational inventory scores with: 1) overall job performance ($\underline{r}$ = .31); 2) communication effectiveness ($\underline{r}$ = .33); 3) leadership ($\underline{r}$ = .28); 4) problem solving effectiveness ($\underline{r}$ = .20); and 5) interpersonal effectiveness ($\underline{r}$ = .15) in a telecommunications company with entry-level managers. A second study, using salespeople, administrative support, and technical support positions found significant correlations with performance activity, which provides some concurrent validity for the simulation (Motowidlo & Tippins, 1993)

**Generalizability**

The procedure of the low fidelity simulation can be applied to many different types of jobs and contexts. However, based on its situational specificity, there is not one standard instrument.

**Face Validity/Ease of Use/Transparency**

The instrument is face valid, since the situations are drawn specifically from critical incidents from the job. The simulation is easy to administer and score. However, development of the simulation is time-consuming and must be done for each job family it is to be used on. The instrument is also not transparent since the most effective strategy is not readily apparent from the choices as long as the instrument is properly developed.

## Tacit Knowledge

| | |
|---|---|
| Purpose | Assess tacit knowledge |
| Population | Students and faculty members in academia |
| Acronym | N/A |
| Scores | 12 work-related situations |
| Administration | paper and pencil, individual |
| Price | N/A |
| Time | estimated 1 ½ hours |
| Authors | Wagner, R. K. (1987) |
| Publishers | N/A |

### Theory

See general theory in TKLMI section.

### Development and Empirical Use

This tacit knowledge measure consisted of 12 work-related situations. Each situation was associated with between 9 to 11 response items. Four of the situations were meant to tap each of the three contents of tacit knowledge (managing self, tasks, and others). Half of the situations were constructed to tap tacit knowledge with a local context, with the other half tapping it with a global context.

### Psychometrics

Although tacit knowledge measures do not correlate significantly with measures of potentially confounding constructs, subscores within a domain (e.g., tacit knowledge of self, others, or task) do correlate moderately (.30) with one another. This correlation suggests a general factor underlying tacit knowledge within a domain that is different from the general factor measured by traditional tests of intelligence (Wagner, 1987).

Internal consistency reliabilities for the total tacit knowledge scale ranged from .74 to .90, with a median of .82 for the psychology faculty, graduate student, and undergraduate student samples. The reliabilities of the individual tacit knowledge subscales ranged from .48 to .90, with a median of .69. An expert-novice difference was found for tacit knowledge between the faculty, graduate students, and undergraduate students. The linear trend was significant and in the expected direction with faculty scoring highest, followed by graduate students, and then undergraduates.

Significant correlations between tacit knowledge scores and performance criteria for faculty were found for: 1) the number of citations; 2) performance appraisal ratings;

3) the number of publications; and 4) research (Wagner, 1987). Similar correlations were found for graduate students.

**Generalizability**

The specific measure is based on academic, work-related situations, making it applicable only to that particular context.

**Face Validity/Ease of Use/Transparency**

Once the work-related situations are developed, the instrument is easy to administer and score. In our opinion, the items are face valid due to their work-related nature. If developed correctly, the alternatives are not transparent.

## Summary

As we began this section we noted that a wide variety of measures fell under this broad heading of knowledge assessments. Having reviewed the vast array of ARI indices along with their benchmark analogues, this observation remains true. Nevertheless, some emerging themes are evident.

First, there is an issue of general vs. specific forms of knowledge. The Fleishman and Quintance (1984) taxonomy was offered as a framework of general types of ability against which to gauge the ARI and benchmark measures. A review of Table 5 illustrates that a majority of the dimensions listed are addressed by the ARI-BDI and ARI-CI, yet far fewer are tapped by the Tacit Knowledge or Mental Model assessments. Similarly, the benchmark measures tend to either assess a variety of general cognitive abilities, or hone in on a more limited number of requisite job specific knowledges. Naturally there is an implicit tradeoff here between measurement fidelity for any given application vs. generalizability and widespread use. Accordingly, it is important for researchers to articulate what type(s) of knowledge is are) important in their research context. We could easily envision applications where either, or both, general and specific knowledge assessment would prove valuable.

In terms of comparisons along the criteria factors, the ARI instruments were essentially parallel to the selected benchmarks. The development of the ARI instruments and benchmarks are comparable, with moderate to strong development. The LLQ is the exception, with a fairly weak instrumental development. In regards to actual use, the instruments range from limited (e.g., CPA, TKMLI) to widespread (ARI-BDI, most benchmark measures). When comparing the ARI instruments and benchmarks on reliability, all of the instruments are moderate to high. The tacit knowledge benchmark is the one instrument that has shown mixed reliabilities. Not all of the measures have construct validity evidence to report. Of those that do, the CPA has the poorest construct validity when compared to the benchmarks. The measures that had criterion-related validity reported were basically comparable, showing moderate to high validities. The LLQ is the one measure that displayed mixed results for criterion-related validity. The only measure that reported discriminant validity was the Concept Mastery Test, making it

difficult to compare to the other measures on this criterion.

As was the trend on the other criteria, the face validities of these measures varied between low and high. The CPA and the Consequences measure are the two with the lowest ratings on this criterion, which causes some concern. The majority of the other measures are comparable, with high ratings for face validity.

Ease of use represents an important decision parameter. The CPA is clearly the most time intensive and demanding to administer limiting its potential use. The tacit knowledge, mental models, and constructed response exercise all require a substantial investment of time initially during the development phase. Once established, the tacit knowledge and mental models assessments are relatively ease to administer and to score, whereas the constructed response exercise still demands substantial review and scoring by trained coders. We should note, however, that given the manner in which the ARI Tacit Knowledge and Mental Model assessments were developed, their generalizability to other army applications is likely to be better than would usually be available from such measures.

The ARI-BDI and ARI-CI assessments are both easy to administer and to score. It is important, however, that versions of these instruments move out of the development and refining stage and into the application stage. In other words, it is important to identify some "core" set of dimensions for these instruments that would remain intact and be administered in a variety of applicable circumstances. To the extent that different versions exist with each administration, it becomes difficult to draw any definitive conclusions.

**Recommendations**

Different research questions and applications will call for different strategies, but, in general, it makes sense to have a battery of general cognitive ability measures available for various uses. For example, test batteries such as the GATB or AFQT could be administered (or might even be available from personnel files, with appropriate confidentiality cautions respected) to personnel. Batteries such as these are readily available and would help to eliminate much of the current instrument development work. Moreover, we suspect that these general assessments would provide much of the generic knowledge indices currently supplied by the ARI-BDI and ARI-CI measures.

Alternatively, the ARI-BDI or ARI-CI might expand their coverage to better sample the knowledge domain, perhaps by using other methods for assessing other variables (e.g., personality). Either approach would also provide a more common framework to use as a comparison basis for different studies aimed at unpacking the importance of leaders' knowledges.

This would still leave a need, in many applications, to assess more specific forms of knowledge such as tacit or mental models. The approach adopted by ARI for these measures has been sound, in that, the researchers have sought to strike a balance between sensitivity to the knowledge requirements of individual assignments, yet maintain a limited range of generalizability. Such development strategies, combined with a comprehensive job analysis of leadership positions, would help to align specific knowledge assessments with the requirements of different positions.

Table 5

Leader Knowledge Comparison

| | ARI Featured Instruments | | | | | | |
|---|---|---|---|---|---|---|---|
| *Variable* | ARI Background Data Inventory | ARI Critical Incidents | Problem-solving Tasks | Constructed Response Exercise | Mental Models | Career Path Appreciation | Tacit Knowledge for Military Leadership Inventory |
| **Linguistic Ability** | | | | | | | |
| **Verbal Comprehension** | X | X | | | | X | |
| **Verbal Expression** | X | X | | | | X | |
| **Written Comprehension** | X | X | | | | | X |
| **Written Expression** | X | X | | | | | X |
| | | | | | | | |
| **Creativity** | | | | | | X | X |
| **Definition of Problem** | X | X | | | | | X |
| **Fluency of Ideas** | X | X | | | | | |
| **Originality** | X | X | | | | | |
| | | | | | | | |
| **Memory** | | | | | | | |
| **Memorization** | | | | | | | |
| | | | | | | | |
| **Problem Solving/Reasoning** | | | | X | | | X |
| **Problem Sensitivity/ Problem Anticipation** | X | X | | | | | |
| **Deductive Reasoning** | X | X | | | | X | |
| **Inductive Reasoning** | X | X | | | | X | |
| **Long Term Planning** | | | | | | | |
| **Conceptual Flexibility** | | | | | | | X |
| **Decision Making** | | | | | | X | X |

| | ARI Featured Instruments | | | | | | |
|---|---|---|---|---|---|---|---|
| *Variable* | ARI Background Data Inventory | ARI Critical Incidents | Problem-solving Tasks | Constructed Response Exercise | Mental Models | Career Path Appreciation | Tacit Knowledge for Military Leadership Inventory |
| Evaluation of Arguments | | | | | | | |
| Interpretation | | | | | | | |
| Recognition of Assumptions | | | | | | | |
| Inference | | | | | | | |
| Solution Construction | | | | X | | | |
| Social Judgment Skills | | | | X | | | |
| Problem Construction | | | X | | | | |
| Information Encoding | | | X | | | | |
| Category Search | | | X | | | | |
| Category Combination | | | X | | | | |
| Wisdom | | | X | | | | |
| | | | | | | | |
| Perception/Information Processing | | | | | | | |
| Information Ordering | X | X | | | | | |
| Category Flexibility | | | | | | | |
| Spatial Orientation | | | | | | | |
| Visualization | | | | | | | |
| Speed of Closure | | | | | | | |
| Flexibility of Closure | | | | | | | |
| Selective Attention | | | | | | | |
| Perceptual Speed | | | | | | | |
| Time Sharing | | | | | | | |

| | ARI Featured Instruments | | | | | | |
|---|---|---|---|---|---|---|---|
| *Variable* | ARI Background Data Inventory | ARI Critical Incidents | Problem-solving Tasks | Constructed Response Exercise | Mental Models | Career Path Appreciation | Tacit Knowledge for Military Leadership Inventory |
| **Technical Ability** | | | | | | X | X |
| **Cognitive Complexity** | | | | | X | X | X |
| **Tacit Knowledge** | | | | | | | |
| **Intrapersonal** | | | | | | X | X |
| **Interpersonal** | | | | | | | X |
| **Organizational** | | | | | | X | X |

| | Benchmarks | | | | | | |
|---|---|---|---|---|---|---|---|
| *Variable* | Watson-Glaser Critical Thinking Appraisal | Concept Mastery Test | Consequences | Low Fidelity Simulation (Motowidlo et al.) | Leatherman Leadership Questionnaire | PathFinder | Tacit Knowledge |
| **Linguistic Ability** | | | | | | | |
| **Verbal Comprehension** | | | | X | | | |
| **Verbal Expression** | | | | X | | | |
| **Written Comprehension** | | | | X | | | |
| **Written Expression** | | | | X | | | |
| **Creativity** | | | | | | | |
| **Definition of Problem** | | | | X | | | |
| **Fluency of Ideas** | | | X | | | | |
| **Originality** | | | X | | | | |
| **Memory** | | | | | | | |
| **Memorization** | | | | | | | |
| **Problem Solving/Reasoning** | | | | X | X | | |
| **Problem Sensitivity/ Problem Anticipation** | | | | | | | |
| **Deductive Reasoning** | X | | | X | | | |
| **Inductive Reasoning** | | | | X | | | |
| **Long Term Planning** | | | | | | | |
| **Conceptual Flexibility** | | | | X | | | |
| **Decision Making** | | | | X | | | |
| **Evaluation of Arguments** | X | | | | | | |

| | Benchmarks | | | | | | |
|---|---|---|---|---|---|---|---|
| **Variable** | Watson-Glaser Critical Thinking Appraisal | Concept Mastery Test | Consequences | Low Fidelity Simulation (Motowidlo et al.) | Leatherman Leadership Quesitonnaire | PathFinder | Tacit Knowledge |
| **Interpretation** | X | | | | | | |
| **Recognition of Assumptions** | X | | | | | | |
| **Inference** | X | | | | | | |
| **Solution Construction** | | | | | | | |
| **Social Judgment Skills** | | | | | | | |
| **Problem Construction** | | | | | | | |
| **Information Encoding** | | | | | | | |
| **Category Search** | | | | | | | |
| **Category Combination** | | | | | | | |
| **Wisdom** | | | | | | | |
| **Selective Attention** | | | | | | | |
| **Perceptual Speed** | | | | | | | |
| **Time Sharing** | | | | | | | |
| **Technical Ability** | | | | X | | X | |
| **Cognitive Complexity** | | X | | X | | | |
| **Tacit Knowledge** | | | | | | | X |
| **Intrapersonal** | | | | X | | | |
| **Interpersonal** | | | | X | | | |
| **Organizational** | | | | | | | |

# Section 4: Biodata

Based on our review of the last decade of ARI leadership research, and our discussions with the ARI Scientists, biodata was chosen as a featured area for this report. Whereas the other three sections represent substantive variables, biodata really describes a method of measurement. As noted earlier, biodata assessments tend to traverse several substantive areas including personality, knowledge, and previous performances. Consequently, we mentioned it briefly in Sections 2 and 3. In this section, however, we will consider biodata as a whole in terms of a measurement protocol and procedure.

Three biodata instruments used in ARI research will be presented as the featured measures. Two are for use with civilian supervisors and the other for use with Special Forces. Other variations exist, but have not been used as prominently or were not emphasized as much by the research scientists. We chose to feature these three instruments because more information was available in terms of scale definitions, development and empirical use, and psychometric information.

Two benchmark measures will be presented for this section. The first benchmark is one produced by the Life Insurance Marketing Research Association for life insurance field managers, called Assessment Inventory for Managers (AIM). The second benchmark is one from the external research community, the Biographical Questionnaire (BQ). This measure, unlike the others in this section, has been used more often in student settings.

This biodata section contains a brief literature review that is followed by the presentation of Table 6. The table displays a summary of the three ARI measures and two benchmarks used for comparison on the eight criteria used for evaluation. This table is discussed in more detail in the text. Next, our evaluation of the ARI measures when compared with the benchmarks is presented, followed by recommendations for future research. This section concludes with Table 7, which presents the variables tapped in each of the featured and benchmark measures.

**Literature Review**

Biodata consists of previous and current life events that have influenced the behavioral patterns, dispositions and values of the individual (Mael & Schwartz, 1991). It

describes things that have been done to a person (e.g., by teachers, parents, friends, employers, etc.) and experience that the individual has had. Biodata can be similar to temperament measures, personality measures, interest inventories, or cognitive ability indicators. However, there are certain characteristics that distinguish biodata from these other indices. Temperament measures focus on stable dispositional tendencies, not indicators of disposition as shapers of behavior. One difference between personality measures and biodata is that biodata focuses on prior behavior and experiences from specific situations. In addition, biodata items allow individuals to provide definite, specific, unique answers, whereas personality items may not (Gunter, Furnham, & Drakeley, 1993). Interest inventories tap an individual's willingness to enter into a specific situation, while biodata explores individuals' actual reactions to past situations. In terms of cognitive abilities, biodata measures expose a more practical intelligence than cognitive measures that present problem solving situations and assess the upper bound of cognitive ability (Mumford & Stokes, 1992). In short, biodata represents self-report measures of previous behavior rather than indicators of underlying latent traits thought to predict behavior. Stated differently, biodata represent previous samples of behavior rather than signs or predictors of future behavior (Wernimont & Campbell, 1968).

Biodata is based on the premise that past behavior will influence future behavior (Owens & Schoenfeldt, 1979). Therefore, if one wants to predict an individual's behavior, such as leadership ability, you would look at his or her past experiences. Prior learning, heredity, and environmental circumstances together help to determine an individual's behavior (Mumford & Stokes, 1992). Background data measures require respondents to retrospectively recall how they behaved in the past, over a specified time period and is thought to reveal individuals' characteristic ways of interacting with their environment.

**Table 6**
**ARI Biodata Measures and Benchmarks by Criteria**

| Criteria | ARI Civilian Supervisors Biodata | ARI Special Forces Biodata | Background Data Inventory | Assessment Inventory for Managers (LIMRA) | Owens' Biographical Questionnaire |
|---|---|---|---|---|---|
| Theory | past behavior/ experiences predicts future behavior | past behavior/ experiences predicts future behavior | past behavior/ experiences predicts future behavior | past behavior/ experiences determines future behavior | past behavior/ experiences determines future behavior |
| Descriptive information | 467 items with 21 scales cognitive, self-confidence, motivational, management skills, and social skills variables | 160 items with 17 scales range of variables from social intelligence to physical capability | 160 items with 10 scales motivational, personality, and problem-solving variables | 92 items 10 personal characteristics 5 cognitive abilities | 118 items demographic, experiential, and attitudinal variables |
| Development | strong | strong | strong | long history; revision in progress | moderate |
| Empirical use | widespread use in military | widespread use in military | widespread use in military | widespread use in Insurance Industry | widespread use with students |
| Generalizability | high (civilian army) | moderate (army) | high (civilian army) | moderate (Insurance positions) | moderate (students) |
| Psychometrics | mixed reliability strong construct validity strong criterion-related validity | mixed reliability strong construct validity strong criterion-related validity | mixed reliability moderate to strong criterion-related validity | mixed reliability moderate to strong criterion-related validity | moderate to high test-retest correlations |

| Criteria | ARI Civilian Supervisors Biodata | ARI Special Forces Biodata | Background Data Inventory | Assessment Inventory for Managers (LIMRA) | Owens' Biographical Questionnaire |
|---|---|---|---|---|---|
| Face validity | moderate | moderate | moderate | moderate | moderate |
| Ease of use | moderate | high | high | high | high |
| Transparency | moderate | moderate | moderate | moderate | moderate |

## ARI Research on Biodata

### ARI Civilian Supervisor and Special Forces Biodata

| | |
|---|---|
| Purpose | Predict leader effectiveness based on past behavior and experiences |
| Population | Civilian supervisors, Special Forces, Army War College students, Rangers |
| Acronym | N/A |
| Scores | **Civilian Supervisor version** - 21 scales: 1) Cognitive Ability; 2) Practical Intelligence; 3) Dominance; 4) Achievement; 5) Energy Level; 6) Self-Esteem; 7) Work Motivation; 8) Consideration 9) Self Monitoring; 10) Planning/Organizing; 11) Stress Tolerance; 12) Dependability; 13) Supervisory Skills; 14) Interpersonal Skills; 15) Social Maturity; 16) Communication Skills; 17) Defensiveness; 18) Need For Approval; 19) Need For Security; 20) Harm Avoidance; 21) Object Belief)<br>**Special Forces Version** - 17 scales: 1) Objective Belief; 2) Lie; 3) Swimming; 4) Aggression; 5) Social Intelligence; 6) Autonomy; 7) Cultural Adaptability; 8) Diverse Friends; 9) Physical Capabilities; 10) Organizational Identification; 11) Work Motivation; 12) High School Leader; 13) Anxiety; 14) Openness/Cognitive Flexibility; 15) Outdoors Enjoyment; 16) Mechanical Aptitude, 17) Team) |
| Administration | Paper and pencil, individual |
| Price | N/A |
| Time | Civilian Supervisor version: estimated 2 hours<br>Special Forces version: estimated 45 minutes |
| Authors | Kilcullen, White, Mumford, & O'Connor (1995) |
| Publishers | ARI |

**Theory**

Biodata can be described as past behavior and experiences that determine future behavior and experiences. Learning, heredity, and environment together make certain behaviors more prevalent (Mumford & Stokes, 1992). Biodata items are designed to tap the developmental history of individuals in terms of typical interactions with the environment (Mumford & Stokes, 1992). Some overlap between personality inventories and biodata is to be expected. However, biodata focuses more on prior behaviors within specific situations. In addition to personality attributes, other variables such as, interests, values, skills, aptitudes, and abilities may also be tapped. There are several purposes for which biodata has been used including: 1) classification of individuals into job families; 2) determining individual organizational action in terms of rewards, training, etc.; 3) reaching an understanding of organizational behavior and designing interventions; and 4) developing theory.

The following individual characteristics, grouped by five factors, were proposed to relate to leadership for the Civilian Supervisor version (Kilcullen, White, Mumford, & O'Connor, 1995):

**Cognition**

1) cognitive ability – the underlying and global capacity for reasoning, abstract thinking, and problem-solving;

**Management skills**

2) practical intelligence – displaying common sense and behaving intelligently in real-life situations;
3) planning/organizing – the ability to plan and organize resources in order to meet objectives;
4) supervisory skills – directing the work of others. This involves delegating and coordinating activities, monitoring the work, making decisions and assuming responsibility;
5) communication skills – the ability to communicate one's ideas;

**Self-confidence**

6) self-esteem- a sense of pride in past achievements and the feeling that one will be able to cope effectively with current and future life events;
7) stress tolerance – the ability to remain calm, even-tempered, maintain composure and think rationally under pressure. Also, the ability to cope with uncertainty or ambiguity;
8) defensiveness – the tendency to deny personal weaknesses;
9) need for approval – the desire to obtain acceptance from others;
10) need for security – the need to maintain stability and predictability in one's life;
11) harm avoidance – the desire to avoid exposure to peril;

**Motivation**

12) work motivation – the preference for work-related activities instead of social/leisure activities;
13) dominance – the desire to control, influence and direct the behavior of others;
14) achievement – the desire to set difficult goals and the ability to subsequently meet or exceed these goals;

15) energy level – the amount of activity and stamina displayed during the course of the day;

16) dependability – the ability to follow through on commitments, meet deadlines, and work accurately with few mistakes;

17) social maturity – the person with social maturity demonstrates honest, trustworthy, and law-abiding behavior. This person is impartial and unbiased in interacting with others;

**Social skills**

18) consideration – a behavioral dimension reflecting the degree to which a leader acts in friendly, supportive manner to subordinates;

19) self monitoring – reflects a concern for social appropriateness, a sensitivity to social/group demands, and the behavioral flexibility that allows the individual to respond effectively to situational demands;

20) interpersonal skills – the ability to establish effective working relationships with others;

21) object belief – the belief that others are merely tools to be used to further one's own objectives.

The following are definitions for the scales in the **Special Forces version**:

1) Object belief – self-focused; using others to get what you want;
2) Lie – choose response options which are socially desirable;
3) Swim – age you learned to swim and swimming ability;
4) Aggression – involvement in fights; publicly demonstrating aggressive tendencies;
5) Social intelligence – ability to read other people and understand others; social perceptiveness;
6) Autonomy – independence; desire to work alone;
7) Cultural adaptability/flexibility – work to understand and respect other cultures;
8) Diverse friends – have a variety of different types of friend with different backgrounds;
9) Physical capability – physical strength and endurance;

10) Organizational identification – global identification with groups and specific identification with Special Forces;
11) Work motivation – having high self-expectations and stretching your abilities;
12) High school leader – participation in student government; officer of student government;
13) Anxiety – tendency to over think situations and worry for unnecessary reasons;
14) Openness/cognitive flexibility – willingness to explore multiple paths to problem solutions;
15) Outdoor enjoyment – participation and enjoyment of outdoor activities such as fishing and hiking;
16) Mechanic aptitude – perform such tasks as car repairs and woodwork;
17) Team – preference to work with others and to play team sports;

A model was proposed that hypothesized that the three factors of cognition, self-confidence, and motivation affect the development of social skills and management skills, which then affect leader performance. Cognition and motivation were also proposed to have a direct influence on leader performance. This model of leader effectiveness was partially based on a leadership prediction model proposed by Mumford et al. (1993). This model proposed that individual characteristics and managerial and social skills influence leadership in the context of problem solving in an ill-defined, social domain. That research sought to examine whether a similar model would be useful in predicting on-the-job performance of Army civilian leaders (e.g., first-line supervisors).

**Development and Empirical Use**

Samples have included 2044 first line civilian supervisors from variety of occupations and grade levels, as well as Special Forces, Army War College participants, and Rangers. These different versions are all based on the same model. However, the versions contain slightly different combinations of scales. The Civilian Supervisor version and the Special Forces version were chosen as the featured versions because they had the most information available in terms of scale definitions, development and empirical use, and psychometric information.

Civilian Supervisor Version. Twenty-one rational scales were developed to measure 21 individual characteristics. A panel of psychologists reviewed construct definitions, and each member generated 10-15 items related to past behaviors and life events. Next, these items were examined by the panel based on the following criteria: 1) construct relevance; 2) response variability; 3) relevance to Army civilian population; 4) readability; 5) non-intrusiveness; and 6) neutral social desirability. From the pool of items, 20-40 of the best items for each construct were chosen, and responses were weighted as to the relationship between responses and the predictor construct. A second panel of psychologists then reviewed this set of items, and a pilot test was conducted. Revisions were made based on the item analysis of the pilot data. The final version of the instrument contained 467 items.

Special Forces Version. A job analysis was conducted to determine the performance dimensions for SF. The job analysis identified 47 attributes relevant to successful performance in SF jobs and 26 critical incident-based categories. SMEs rated attributes as important to this job. The most highly rated attributes were: 1) teamwork and interpersonal skills; 2) adaptability; 3) physical endurance and fitness; 4) strong cognitive abilities; 5) strong leadership and communication skills; and 6) strong judgment and decision making skills. Based on the job analysis, a biographical questionnaire was developed to measure the SF traits.

The questionnaire consisted of 160 items ranging from social intelligence items to physical capability items. It was completed by 1,357 soldiers participating in SF Selection and Assessment processes, as well as by 293 SF officers. The items were then analyzed and scales were created by:

1) analyzing the internal reliabilities of different groups of items in terms of inter-item correlations, inter-total correlations, squared multiple correlations and the scale alphas when the item is removed (empirical); and
2) reading each item and determining the best scale for the item through content analysis (rational).

**Psychometrics**

Civilian Supervisor Version. Convergent validities with related temperament scales were .60 and higher. The alphas for the 21 scales ranged from .65 to .85 (mean =

.76). A blocked regression analysis was used to evaluate the Leader Effectiveness Model. The first block contained cognition, self-confidence, and motivation, which significantly predicted ratings and performance records, multiple Rs equaled.21 and .35, respectively. The second block was composed of management skills and social skills, which led to a significant increase in the $R^2$ for performance records.

Special Forces Version. The alphas for this version are reported in the table below.

| Scale | α (SFAS) | α(SF) |
|---|---|---|
| Objective Belief | .38 | .32 |
| Lie | .50 | .45 |
| Swimming | .82 | .80 |
| Aggression | .55 | .62 |
| Social Intelligence | .86 | .84 |
| Autonomy | .72 | .76 |
| Cultural Adaptabiity | .49 | .68 |
| Diverse Friends | .62 | .58 |
| Physical Capabilities | .82 | .83 |
| Organizational Identification | .75 | .70 |
| Work Motivation | .62 | .64 |
| High School Leader | .87 | .88 |
| Anxiety | .65 | .72 |
| Openness/Cognitive Flexibility | .78 | .72 |
| Outdoors Enjoyment | .78 | .82 |
| Mechanical Aptitude | .72 | .74 |
| Team | .48 | .34 |
| Mean | .67 | .67 |

**Generalizability**

Generalizability for the Civilian Supervisor version to other civilian military samples is to be high given the fact that a wide range of occupations was included in the sample. Generalizability to non-military samples is probably feasible given the fact that the measure appears to be fairly representative of a general leadership domain. The Special Forces version may be less generalizable due to its application to such a distinct population as the Special Forces

**Face Validity/Ease of Use/ Transparency**

The multiple-choice format used in both versions of the biodata instruments makes them easy to use, administer, and score. Based on our review of the items contained in both versions, the instruments appear moderately face valid. The two instruments also appear to us to be moderately transparent.

## Background Data Inventory

| | |
|---|---|
| Purpose | Predict leader effectiveness based on past behavior and experiences |
| Population | Civilian supervisors: 1st, 2nd, and 3rd level in 6 work grades |
| Acronym | BDI |
| Scores | 1) Achievement; 2) Need For Dominance; 3) Openness; 4) Tolerance For Ambiguity; 5) Consideration; 6) Tolerance For Stress; 7) Social Understanding; 8) Behavioral Appropriateness |
| Administration | Paper and pencil, individual |
| Price | N/A |
| Time | estimated 45 minutes |
| Authors | Zaccaro, White, Kilcullen, Parker, Williams, & O'Connor-Boes (1997) |
| Publishers | ARI |

### Theory

The underlying theory for this measure views leadership as complex social problem solving as described by Mumford et al. (1993). Cognitive, motivational, and personality variables were proposed to facilitate the leader's solution of complex problems. A combination of instruments was used to assess these various components of the model. The biodata instrument included several motivational, personality, and problem-solving variables. Achievement and need for dominance were included as the motivational variables. The personality variables included openness, tolerance for ambiguity, consideration, and tolerance for stress. Finally, two problem-solving skills were included, social understanding and behavioral appropriateness.

The following are the definitions of the variables used in the BDI (Zaccaro et al., 1997):

1) Achievement - tendency to strive energetically for success in one's work;
2) Need for dominance - tendency to seek interpersonal influence and control over others;
3) Openness - willingness to consider novel approaches to solving problems and a preference for learning about new ideas;
4) Tolerance for ambiguity - preference for work environments in which problems and potential solutions are unstructured and ill-defined;
5) Consideration - the tendency to be helpful to others;
6) Tolerance for stress - reactivity and emotional stability under physical or emotional stress;

7) Social understanding - the ability to accurately distinguish the different and sometimes conflicting goals of the multiple constituencies that must be considered when developing problem solutions. An awareness of the needs, goals, and demands of other social entities;

8) Behavioral appropriateness- the ability to behave flexibly across multiple organizational situations. Competence in interacting with others in social situations.

**Development and Empirical Use**

A panel of experts selected the items for inclusion in this biodata measure based on the relevance to the construct and demonstrated psychometric quality. The researchers defined psychometric quality as: 1) response variability; 2) relevance to the Army civilian population; 3) readability; 4) non-intrusiveness; and 5) neutral social desirability. The items were then formed into scales to tap each construct. The resulting instrument contained 160 items with 10 scales.

**Psychometrics**

The alphas for the ARI-BDI scales are as follows:

| | |
|---|---|
| Achievement | .59 |
| Need for dominance | |
| Emergent leadership | .77 |
| Team orientation | .75 |
| Bluntness | .56 |
| Personality | |
| Openness | .82 |
| Tolerance for ambiguity | .78 |
| Consideration | .77 |
| Tolerance for stress | .87 |
| Problem-solving skills | |
| Social understanding | .86 |
| Behavioral appropriateness | .64 |

The scales were correlated with the following leader characteristics: 1) planning; 2) special organization-wide projects; 3) boundary spanning; 4) entrusted problem-solving responsibility, and 5) networking/mentoring. All of the biodata scales correlated significantly with these 5 characteristics. When the scales were grouped by motivation,

personality, and problem-solving, each of the sets also correlated significantly with the 5 characteristics.

Hierarchical regression analyses were conducted with the scales entered in sets of motivation, personality, and problem-solving. The following four criteria were entered into the analyses: 1) advancement; 2) leadership job performance; 3) administrative criteria; and 4) senior leadership potential. Each of the three sets of variables added incrementally to the prediction of leader advancement. The results for the other three criteria were mixed, but most were not significant. The set of motivation variables did add incrementally to the prediction of administrative criteria and senior leadership potential. However, leadership job performance was not predicted by the three sets of variables.

**Generalizability**

Based on our review of this research, generalizability would be high within the context of civilian supervisors in the army. This follows from the fact that the three supervisory levels and six service grades were represented in the sample. The measure may generalize more easily to Army civilian populations than external organizations. However, the constructs appear to be applicable to leadership in general.

**Face Validity/Ease of Use/ Transparency**

Based on our interpretation of the items, the measure appears to have moderate face validity. It is easy to use, administer, and score due to the multiple-choice items. The items appear to be moderately transparent.

## Benchmark Instruments

### Assessment Inventory for Managers (AIM)

| | |
|---|---|
| Purpose | Predict manager performance |
| Population | Field managers in life insurance industry |
| Acronym | AIM |
| Scores | **10 personal characteristics**: 1) Achievement Orientation; 2) Adaptability; 3) Relationship Orientation; 4) Commitment; 5) Interpersonal Orientation; 6) Integrity; 7) Leadership; 8) Creativity; 9) Other Orientation; 10) Energy<br>**5 cognitive abilities**: 1) Time Sharing; 2) Originality; 3) Selective Attention; 4) Memory; And 5) Idea Generation |
| Administration | Paper and pencil, individual |
| Price | N/A |
| Time | estimated 35 minutes |
| Authors | Life Insurance Marketing and Research Association (1991) |
| Publishers | N/A |

**Theory**

A list of personality characteristics was generated based on focus groups with company mangers, the personality assessment literature, and the review of other managerial selection tests. SMEs familiar with the field manager position were provided with the lists and their definitions, and checked characteristics they felt were required to perform a field manger's job. A group of industry researchers and company representatives then reviewed, revised, and consolidated the personal characteristics.

The following is the list of **personal characteristic variables** (Baratta & McManus, 1991):

1) achievement orientation - motivated by doing well or attaining goals. Individual has the drive to stay with a position or plan of action until the desired objective is attained;
2) adaptability - dealing with change, opposition, disappointment, or rejection in a composed manner; flexibility;
3) relationship orientation - a desire to be liked and regarded well by other people, to be part of a group;
4) commitment - establishing and maintaining loyalty to the company and to the company's goals;

5) interpersonal orientation - interacting with others with understanding and relating to others' needs; showing respect for others in a sincere manner while being sensitive to individual differences;

6) integrity - conducting business in a honest, fair, and lawful manner. This includes adhering to policies and procedures, avoiding conflicts of interest, communicating in a straightforward manner, accepting responsibility for own actions, and crediting others when warranted;

7) leadership - using appropriate interpersonal styles to guide and motivate people toward task accomplishment through example, encouragement, guidance, and feedback;

8) creativity - integrating abilities, knowledge, and new ideas and putting them into practice;

9) other orientation - getting a sense of accomplishment through the success of others; willingness to work with others and help others to success;

10) energy - establishing and maintaining a high activity level.

The following is the list of **cognitive ability variables**:

1) time sharing – ability to shift back and forth between two or more sources of information while remaining focused on the problem at hand;

2) originality – ability to come up with creative solutions to problems or to develop new procedures to situations where standard operating procedures do not apply;

3) selective attention – ability to concentrate and not be distracted;

4) memory – ability to remember relevant sets of information such as names, numbers, procedures, and presentations;

5) idea generation – ability to produce a number of ideas about a given topic.

**Development and Empirical Use**

The biodata questionnaire contains multiple-choice items that ask individuals to report their prior behavior, experiences, or feelings in certain situations. Items were selected to tap 10 personal characteristics and five cognitive abilities. A set of 137 items was reviewed by researchers who were unaware of their intended dimensions. Items were eliminated if 50% of the reviewers did not agree on which construct the item tapped. At the end of this process, 100 items remained.

**Psychometrics**

A pilot study was conducted to evaluate the 100 biodata items and 32 social desirability items. A sample of 1,218 managers and sales representatives was mailed the survey, and 272 were returned. The items were scored rationally and 11 items were dropped, 13 revised, and three were added. The resulting version of the biodata measure included 92 items, along with the Marlowe-Crowne Social Desirability Scale. Alphas for the various dimensions ranged from .18 to .67 (mean = .36).

**Generalizability**

This measure was specifically developed for LIMRA field managers. Therefore, it may have limited generalizability to other types of jobs outside of this context. Nevertheless, there should be some generalizability to other civilian occupations.

**Face Validity/Ease of Use/ Transparency**

It is a paper and pencil instrument, and easy to administer. The social desirability scale is used along with the biodata measure, which will help to illuminate potential faking. The instrument is easily scored for a biodata measure due to the multiple-choice format. Based on our review of the instrument items, it appears to be moderately face valid and moderately transparent. The company developed this instrument to be used within the Insurance industry.

## Biographical Questionnaire (BQ)

| | |
|---|---|
| Purpose | Predict future behavior |
| Population | Students |
| Acronym | BQ |
| Scores | **13 factors for males**: 1) Warmth Of Parental Relationship; 2) Intellectualism; 3) Academic Achievement; 4) Social Introversion; 5) Scientific Interest; 6) Socioeconomic Status; 7) Aggressiveness/ Independence; 8) Parental Control Vs. Freedom; 9) Positive Academic Attitude; 10) Sibling Friction; 11) Religious Activity; 12) Athletic Interest; And 13) Social Desirability<br>**15 factors for females**: 1) Warmth Of Maternal Relationship; 2) Social Leadership; 3) Academic Achievement; 4) Parental Control Vs. Freedom; 5) Cultural-Literary Interests; 6) Scientific Interest; 7) Socioeconomic Status; 8) Expression Of Negative Emotions; 9) Athletic Participation; 10) Feelings Of Social Inadequacy; 11) Adjustment; 12) Popularity With Opposite Sex; 13) Positive Academic Attitude; 14) Warmth Of Paternal Relationship; And 15) Social Maturity. |
| Administration | Paper and pencil, individual |
| Price | N/A |
| Time | estimated 25 minutes |
| Authors | Owens (1968) |
| Publishers | N/A |

### Theory

Owens (1968; 1971) based his biodata research on the developmental-integrative (D-I) model. He proposed that in order to discover the laws of human behavior, it is necessary to explain the behavior of more than a narrow band of individuals. He suggested that it would be possible to identify subgroups of subjects to which a law applied. A way in which to group individuals involved their patterns of prior experience, which could be collected via biodata. Two hypothetical categories are part of this model: 1) inputs to the individual; and 2) prior experiences of the individual. A basic tenet of mental measurement, that the best predictor of an individual's future behavior is his or her past behavior, is revised to explain groups in the D-I model. The D-I model refers to subgroupings of individuals based on similarities in patterns of their prior experience.

The items on the BQ are demographic, experiential, and attitudinal variables, that were proposed to relate to personality structure, personal adjustment, or success in social, educational, or occupational pursuits.

**Development and Empirical Use**

Item topics were developed by expanding on the outlines implied under input variables and prior behaviors. Two thousand items were developed as a result. Rational screening was used, and the number of items was reduced to 659. These items were administered to 1700 male university freshmen. The items were divided into five subsets, and factor analyzed in an overlapping manner five times. The sixth factor analysis contained all high-loading items from the previous analyses, as well as some additional items that had previously been demonstrated to be valid or seemed to tap a major developmental hypothesis. These analyses yielded nine factors. One, a "difficulty" factor was dropped, which reduced the factors to eight. Next, the data were re-analyzed including items that loaded above .30. Redundant and ambiguous items were removed, leaving 389 items.

This set of items was administered to 1037 male and 897 female university freshmen. Items were eliminated for the following reasons: 1) poor response distributions; 2) tapping unlikely activities; and 3) redundant items. Factor analyses were performed separately for males and females, and ultimately resulted in 13 factors for males and 15 factors for females. These items were used to develop the 118-item short form, named the University of Georgia Biographical Questionnaire. This form was then administered to four successive years of freshman at a university, and the data was factor analyzed.

The following 13 factors were found for the male version of the questionnaire: 1) warmth of parental relationship; 2) intellectualism; 3) academic achievement; 4) social introversion; 5) scientific interest; 6) socioeconomic status; 7) aggressiveness/ independence; 8) parental control vs. freedom; 9) positive academic attitude; 10) sibling friction; 11) religious activity; 12) athletic interest; and 13) social desirability. The following 15 factors were found for the female version: 1) warmth of maternal relationship; 2) social leadership; 3) academic achievement; 4) parental control vs. freedom; 5) cultural-literary interests; 6) scientific interest; 7) socioeconomic status; 8) expression of negative emotions; 9) athletic participation; 10) feelings of social inadequacy; 11) adjustment; 12) popularity with opposite sex; 13) positive academic attitude; 14) warmth of paternal relationship; and 15) social maturity.

**Psychometrics**

The original principal components analysis, conducted separately for men and women, involved 275 items. The principal components analysis conducted by Eberhardt and Muchinsky (1982) involved the 118 items that appeared in the final form of the measure. A total of 13 components were extracted for men and 15 for women.

Test-retest correlations ranged from .49 to .91 (mean=.78) for males, and from .50 to .88 (mean=.76) for females (Shaffer, Saunders, & Owens, 1986).

**Generalizability**

Generalizability beyond students is difficult to establish in the case of this measure, because these factors may not all be relevant to predicting leader behavior in organizations.

**Face Validity/Ease of Use/Transparency**

Based on our review of this measure, some items may appear to be face valid as they request retrospective information on actual previous behavior. However, respondents may not see a direct correspondence between behavior in high school and their leadership effectiveness. The instrument is easy to use, administer, and score due to the multiple-choice format. We expect that the items are moderately transparent to participants.

## ARI Measures vs. Benchmarks

**Summary**

In terms of the theory behind the ARI biodata instruments and the benchmarks, they are all based on the same concept of past behavior predicting future behavior. The differences between the instruments lie more in terms of the specific models of leader effectiveness they are based on and the dimensions that they include.

In terms of direct comparisons, four of the instruments reviewed are similar in format, with approximately the same number of items. The ARI Civilian Supervisor Biodata instrument is substantially longer than the other instruments, having over 400 items. All are paper and pencil with multiple-choice formats. In terms of content, Table 7 illustrates the differences in emphases. The ARI-Civilian and SF versions address two facets of Cognition each, although they are not consistent. In contrast, the LIMRA benchmark measure addresses six facets. Self-confidence, motivation, and social skills are well represented in the Civilian measure, but to a much lesser extent by the other instruments. All measures include some scales related to management skills and personality, but the diversity of the specific dimensions selected for inclusion is striking. Finally, the SF version shares more with Owen's biographical questionnaire in terms of addressing physical abilities, a lie or social desireability check, and outside interests, as compared to the other instruments.

The development of the instruments, both internal and external to ARI, is comparable. Most efforts began with a thorough job or task analysis, followed by an item generation and reduction phase. Usually both SME judgments and empirical methods were used together. All of the instruments have had widespread empirical use, within the restriction of specific contexts. The two ARI biodata versions and the BDI have been restricted to use within the Army. The AIM was developed for use in the life insurance industry, and specifically within one company. Owens' BQ has been studied the most in student populations.

The ARI Civilian Supervisor version and the BDI have high generalizability due to the coverage of different supervisory levels. Therefore, within the Army, these instruments should be applicable to many other samples. The Special Forces version may be more moderately generalizable due to the unique characteristics of Special Forces.

The AIM is moderately generalizable within the context of insurance positions, but these would not have as wide a range as the Army positions.

Based on our review of the instruments, all of them displayed a moderate amount of face validity. In terms of ease of use, the instruments were all comparable with the exception of the ARI Civilian Supervisor instrument. This one was ranked as moderately easy to use due to the length of the instrument, which would make it more time-consuming. For the final criteria of transparency, all of the instruments received the rating of moderate, based on our evaluation of the items.

**Recommendations**

Biodata represents a bit of a paradox, as it simultaneously appears to be "everything" and "nothing." Attempting to classify what biodata is proves to be very difficult. As so eloquently stated by Owens (1976; p. 623), "It is entirely appropriate wish to allocate biodata to some position within the network of variables which constitutes the measurement domain. The task, however, is not singular but plural, since biodata is not one measure of one dimension but multiple measures of multiple dimensions. Thus, one must first decide the essential dimensions and then decide how each relates to some key variables in the domain (emphasis in original)".

Following Owen's advice, we recommend that future biodata efforts adopt a more a priori framework. The prototypical procedure followed to date has been to generate a lengthy list of potential items, to reduce them using rationale and empirical methods, and to derive a new set of dimensions for each application. What is needed, we suggest, is a more theory guided approach where specific underlying dimensions are articulated initially, items written to address those specific dimensions, and then confirmatory analyses be conducted to determine how well those dimensions were assessed. Moreover, we believe that a "core set" of leadership effectiveness related dimensions likely exists that could be generalizable, at least across Army classifications. In other words, we believe that a core set of dimensions could be constructed and included in virtually all leader effectiveness studies where biodata predictors are warranted. Naturally, these could be supplemented with additional scales to the extent justified by the research design, criteria addressed, sample population, etc. However, there should definitely be some (relatively large) degree of carry-over across studies.

We should also comment on a fairly technical, yet important, analytic issue related to biodata. Traditional methods of data reduction and reliability assessment, such as exploratory and confirmatory factor analyses, internal consistency estimates, etc., presume that a latent unobservable variable exists that gives rise to certain essentially parallel indicators. In other words, the underlying dimension causes how one responds to a given set of items. This logic makes perfect sense when considering, for example, traditional knowledge, personality, and attitudinal variables. One's mechanical aptitude, extroversion, or organizational commitment would lead one to respond in certain ways on testing devices. However, the logic of biodata is often that one's personal characteristics or experiences lead to or create some underlying theme that may relate to future activities. Here survey responses describe causes not effects of the underlying dimension. In these cases, different statistical techniques are warranted such as grouping items on the basis of cluster analysis, or applying cause indicator methods of confirmatory factor analyses, and different reliability models are warranted (cf., Bollen, 1989; Nunnally & Bernstein, 1994).

We suspect that the difficulties associated with consistently identifying biodata dimensions and the somewhat low reliabilities reported may be attributable, at least in part, to mixing items that are thought to be causes vs. effects. Again, this underscores the importance of a priori specification of what the targeted dimensions are and how they will be manifest in the assessment device. That foundation, then, drives the analytic tools to be applied.

In conclusion, biodata represents a powerful assessment technique that can provide information across a number of substantive areas. With greater a priori specification of targeted dimensions we would hope that a "core set" of subscales could be established and the need for others identified.

Table 7
Biodata Comparison

| | ARI Featured Instruments | | | Benchmarks | |
|---|---|---|---|---|---|
| *Variable* | *ARI Civilian Biodata* | *ARI Special Forces Biodata* | *Background Data Inventory* | *Assessment Inventory for Managers (LIMRA)* | *Owens' Biographical Questionnaire* |
| **Cognition** | | | | | |
| **Cognitive Ability** | X | | | | X |
| **Practical Intelligence** | X | | | | |
| **Creativity** | | | | X | |
| **Cognitive Flexibility** | | X | | | |
| **Mechanical Aptitude** | | X | | | |
| **Time Sharing** | | | | X | |
| **Originality** | | | | X | |
| **Selective Attention** | | | | X | |
| **Memory** | | | | X | |
| **Idea Generation** | | | | X | |
| **Self Confidence** | | | | | |
| **Self esteem** | X | | | | |
| **Defensiveness** | X | | | | |
| **Harm Avoidance** | X | | | | |
| **Dominance** | X | | X | | |
| **Assertiveness/ Independence** | | | | | X |
| **Positive Academic Attitude** | | | | | X |

| | ARI Featured Instruments | | | Benchmarks | |
|---|---|---|---|---|---|
| *Variable* | *ARI Civilian Biodata* | *ARI Special Forces Biodata* | *Background Data Inventory* | *Assessment Inventory for Managers (LIMRA)* | *Owens' Biographical Questionnaire* |
| **Motivation** | | | | | |
| **Achievement** | X | | X | X | |
| **Energy Level** | X | | | X | |
| **Stress Tolerance** | X | | X | | |
| **Work Motivation** | X | X | | | |
| **Emergent Leadership** | | | X | | |
| **Bluntness** | | | X | | |
| **Social Skills** | | | | | |
| **Communication Skills** | X | | | X | |
| **Dependability** | X | | | | |
| **Interpersonal Skills** | X | | | X | |
| **Object Belief** | X | | | | |
| **Self Monitoring/ Behavioral Appropriateness** | X | | X | | |
| **Social Maturity** | X | | | | |
| **Social Intelligence** | | X | | | |
| **Social Understanding** | | | X | | |
| **Management Skills** | | | | X | |
| **Adaptability** | X | | | X | |
| **Consideration** | X | | X | | |
| **Leadership/High School Leader** | | X | | X | X |

| | ARI Featured Instruments | | | Benchmarks | |
|---|---|---|---|---|---|
| *Variable* | *ARI Civilian Biodata* | *ARI Special Forces Biodata* | *Background Data Inventory* | *Assessment Inventory for Managers (LIMRA)* | *Owens' Biographical Questionnaire* |
| **Planning/Organizing** | X | | | | |
| **Supervisory Skills** | X | | | | |
| **Interests** | | | | | |
| **Cultural-Literary Interest** | | | | | X |
| **Mechanical Interest** | | | | | X |
| **Outdoors Enjoyment** | | X | | | |
| **Religious Activity** | | | | | X |
| **Scientific Interest** | | | | | X |
| **Relationships** | | | | | |
| **Object Belief** | | X | | | |
| **Other Orientation** | | | | X | |
| **Parental Control** | | | | | X |
| **Relationship Oriented** | | | | X | |
| **Sibling Friction** | | | | | X |
| **Warmth of Parental Relationship** | | | | | X |
| **Diversity Issues** | | | | | |
| **Diverse Friends** | | X | | | |
| **Cultural Adaptability** | | X | | | |

| | ARI Featured Instruments | | | Benchmarks | |
|---|---|---|---|---|---|
| *Variable* | *ARI Civilian Biodata* | *ARI Special Forces Biodata* | *Background Data Inventory* | *Assessment Inventory for Managers (LIMRA)* | *Owens' Biographical Questionnaire* |
| **Personality** | | | | | |
| **Aggression** | | X | | | |
| **Anxiety** | | X | | | |
| **Autonomy** | | X | | | |
| **Social Extraversion** | | | | | X |
| **Need for Approval** | X | | | X | |
| **Need for Security** | X | | | | |
| **Team Orientation** | | X | X | | |
| **Openness** | | | X | | |
| **Tolerance for Ambiguity** | | | | | |
| **Physical Ability** | | | | | |
| **Athleticism** | | X | | | X |
| **Physical Capability** | | X | | | |
| **Other** | | | | | |
| **Organizational Commitment** | | | | X | |
| **Organizational Identification** | | X | | | |
| **Lie** | | X | | | |
| **Social Desirability** | | | | | X |
| **Integrity** | | | | X | |
| **Socioeconomic Status** | | | | | X |

# Section 5: Leader Behavior

Based on a review of the ARI leadership projects and discussions with ARI research scientists, leader behavior was an area that has received considerable attention. A scan of the ARI leadership database showed that over 34% of the variables were categorized as relating to leader behavior. Leader behaviors are important to ARI because they are seen as contributing to organizational effectiveness. Leader behaviors are assessed by three different means in the Army context. The first measure featured is the Multifactor Leadership Questionnaire (MLQ). This measure was designed to aid leadership development by identifying the types of leadership styles used and which ones work best in certain contexts. This measure has been employed on a wide range of participants occupying various anks and leadership positions. The second featured measure is the Cadet Performance Report (CPR). The CPR measures leader behavior in cadet performance at the USMA for developmental purposes. This measure is completed by peers and superiors for a target cadet. The third and most comprehensive assessment tool presented is AZIMUTH/SLDI. This assessment instrument is a 360-degree tool that taps leaders' knowledges and behaviors. The assessment is completed by peers, self, subordinates, and superiors. It has been used on various military and civilian officers.

Benchmarks in the mainstream literature and commercial world were compared to ARI's tools. Based on the wide range of behavioral variables that are tapped by the three ARI instruments, there are a great number of benchmarks outlined below. This was necessary to ensure comprehensive coverage on all the identified behaviors assessed in ARI. Along with the many different benchmarks, there is also a wider range of procedures used to tap leader behavior. The MLQ is a self-report instrument, whereas the CPR and AZIMUTH/SLDI are, to some degree, 360-degree systems. A 360-degree system is one in which a variety of sources, such as the self, peers, and supervisors, complete ratings on an individual. A brief literature review below will outline this method more thoroughly.

As with the other sections, a literature review of leader behavior is presented followed by the evaluation of the ARI and benchmark instruments, as illustrated by Table 8. At the conclusion of the instruments' review, a critique of the ARI measures as

compared to the benchmarks is presented, followed by Table 9 which highlights the overlap in all of the measures on specific leader behavior variables.

**Literature Review**

Formal attempts to define the domain of leadership behavior have a long history. The first work on leader behavior revolved around initiating structure and consideration, based on the influence of the Ohio State Leadership Studies. These were two behavior categories that containing a wide variety of specific types of behavior. Initiating structure is broadly defined as the degree to which a leader defines and structures his or her role and the roles of followers to attain goals (Stodgill, 1963). Consideration is defined as the degree to which a leader acts friendly and supportive, showing concern for followers, and looking out for their welfare (Stodgill, 1963). As leadership research has developed, most researchers realized that it was necessary to examine more specific types of behaviors beyond consideration and initiating structure. As a result there have been a plethora of taxonomies attempting to organize leader behaviors (e.g. Stodgill, Goode, & Day, 1965; Mintzberg, 1973; Oldham, 1976; Farr, 1982; Van Fleet & Yukl, 1986).

A major problem in the research and assessment of leader behavior has been the identification of behavioral categories that are relevant and meaningful. Differing behavioral taxonomies and the content of behavior descriptions assessed in measures have resulted in many behaviors that are thought to apply to leaders. Behavioral categories are derived from observed behavior in order to organize perceptions of the world and make them meaningful. However, these categories are really abstractions with no absolute set of correct behavior categories. Therefore, the categories tapped must be based on some specific expectations or focus (Yukl, 1994).

While it is true that the specific behaviors in taxonomies may vary widely, a review of sixty-five classification systems (Fleishman, Mumford, Zaccaro, Levin, Korotkin, & Hein, 1991) concluded that there are three common trends. In almost all classification systems, there are dimensions that focus on the facilitation of group social interaction and objective task accomplishment, management or administrative functions, and information acquisition and utilization (Fleishman et al., 1991). Along with this discovery comes a new approach to studying leader behavior. While most past research on leader effectiveness has examined behaviors individually, there is now a recognition

that patterns of specific behaviors may identify leader effectiveness more clearly (Yukl, 1994). Descriptive studies of leadership have found complex interactions of specific behaviors (Kaplan, 1986). A leader's skill in selecting and using these specific patterns is what leads to effectiveness. Behavior taxonomies are helpful descriptive aids, but the really important information in studying leader behavior occurs with the interaction between the specific behaviors (Yukl, 1994).

**360 Degree Feedback**

Feedback from multiple sources representing different organizational levels, or 360-degree feedback, has become a popular tool of organizations, especially in the areas of assessing leader behavior. While performance appraisals tend to be evaluative in nature and linked to organizational consequences, 360-degree feedback has more of a developmental focus. Another benefit of 360 feedback is that leader behaviors can be examined for consistency, and the reliability of the information gathered from various sources can be ascertained (London & Beatty, 1993). Gathering information from multiple sources at different levels will result in a more complete picture of a leader's behavior. Raters will have different perspectives due to their levels in the organization, which may lead to differences in weightings of leadership factors. Raters may be exposed to certain behaviors in varying degrees, so that information from various sources may be more detailed and complete than ratings from a supervisor alone.

Although there are many benefits to using 360-degree feedback, there are also many considerations that have to be addressed. The first deals with administrative issues. Who should be raters needs to be determined, along with what dimensions of behavior they should rate. Once the measures are completed by the various sources, the integration of the responses must be determined. The differential weighting of the sources needs to be determined, such as, determining the relative impact of supervisor vs. peer ratings.

There are also some rater bias concerns with 360-degree feedback systems. Research has shown that self-ratings may only have moderate correlations with ratings from other sources (Harris & Schaubroeck, 1988). This meta-analysis of ratings showed that correlations between peer and supervisor ratings was relatively high (rho = .62), while self-supervisor and self-peer ratings correlated .35 and .36, respectively. These

differences have been explained by the presence of rater biases, as well as by organizational level. However, this potential problem may not be as important in developmental situations as in instances where the feedback is being used for purely evaluative purposes.

In summary, there are four important considerations to maintain when reviewing assessments of leaders' behaviors. First, one must specify the purpose of an evaluation. Are the ratings to be used strictly for developmental purposes, or might they be used for compensation purposes or perhaps as predictors of future behaviors? Second, the content of the measures must be considered. What exactly are the relevant behavioral dimensions to be assessed? Third, the number and sources of input must be considered. For example, most 360 feedback systems include supervisor, peer, and subordinate ratings. However, self-ratings, those from adjacent departments or units, or "customers" (whether they are internal or external to the organization) are but a few other potentially valuable perspectives. Finally, the process of the system, or how it is used, is important. Some process decisions are rather mechanical, such as how much weight to assign to different scores or sources. Other process decisions are more dynamic, such as how does one sample peers to provide ratings, how is information fed back to leaders, how does one deal with discrepancies across sources, what developmental systems are in place to address shortcoming that are identified, etc. At the end of this section we will revisit these four considerations.

Table 8
Leader Behavior (Featured ARI Instruments)

| Criteria | Multifactor Leadership Questionnaire | Cadet Performance Report | Azimuth/Strategic Leader Development Inventory | Leader Practice Inventory |
|---|---|---|---|---|
| Theory | transactional/ transformational leadership | twelve dimensions of performance | Stratified Systems Theory | fundamental pattern of leadership behavior |
| Descriptive Information | identify leadership style<br>78 items<br>5 point scale<br>multiple forms | evaluate cadet performance at USMA<br>Peer, instructor, and self evaluation<br>12 dimensions, one item each<br>5 point scale | 360 degree evaluation and feedback for self development<br>98 items<br>6 point scale | feedback and self development<br>30 statements<br>5 point scale<br>self and other rating forms |
| Development | strong | Weak | Strong | strong |
| Empirical Use | moderate use | moderate use (cadets) | Limited use | limited use |
| Psychometrics | moderate to high reliability<br>strong factor structure | low to moderate construct validity | low to moderate reliability<br>validity in progress | high reliability<br>high construct validity |
| Generalizability | high | moderate - military use only | Moderate - military/civilian leaders | high - no military application |
| Face validity | moderate | high | Moderate | moderate |
| Ease of Use | high | moderate | Low | low |
| Transparency | moderate | high | Moderate | low |

Table 8
Leader Behavior (Benchmarks)

| Criteria | Leader Behavior Description Questionnaire | *Benchmarks* | California Leadership Index | Profiler | Prospector |
|---|---|---|---|---|---|
| Theory | initiations structure and consideration | handling challenges of job/leading people/respecting others | Leadership, Energy, Affability, Dependability, Resilience | managing performance and effectiveness | engage in opportunities to learn and create a context for learning. |
| Descriptive Information | provide subordinate description of leader behavior | 360 degree to assess leader skill and development<br>154 items<br>5 point scale<br>1 form self, 11 forms other | 360 degree to develop profile comparing self and others ratings<br>160 item adjective checklist<br>6 point scale<br>1 form self, 3-5 forms other | 360 degree to provide feedback on competencies to be successful manger<br>159 items<br>5 point scale | 360 degree to assess leader ability to learn from experience<br>48items<br>7 point scale<br>1 form self, 11 forms other |
| Development | strong | moderate | strong | strong | moderate |
| Empirical Use | Widespread use | moderate | moderate | moderate | moderate |
| Psychometrics | moderate to high reliability<br>moderate to low construct validity | high reliability<br>moderate to high construct and criterion-related validity | high reliability<br>high content validity<br>high construct validity | low to moderate reliability<br>validity in progress | high reliability<br>high content validity<br>high criterion-related validity |
| Generalizability | high | moderate - management context | high | moderate - management context | moderate - management context |
| Face validity | high | high | low | low | high |

| Criteria | Leader Behavior Description Questionnaire | *Benchmarks* | California Leadership Index | Profiler | Prospector |
|---|---|---|---|---|---|
| Ease of Use | high | low | low to moderate | high | low |
| Transparency | moderate to high | high | moderate | moderate | high |

## ARI Research on Leader Behavior

### Multi-factor Leadership Questionnaire

| | |
|---|---|
| Purpose | Identified transactional/transformational leadership behaviors. |
| Population | Business, military, government, educational institutions |
| Acronym | MLQ |
| Scores | Transactional: 1) Contingent reward; 2) Management by exception-active; 3) Management by exception-passive; 4) Laissez-faire<br>Transformational: 1) Charisma; 2) Inspiration; 3) Intellectual stimulation; 4) Individualized consideration |
| Administration | Paper and pencil, individual |
| Price | $120 for 1 measure, scoring, and feedback report |
| Time | MLQ 5X short form, 15 to 30 minutes |
| Authors | Bass (1996) |
| Publisher | Mind Garden Inc. |

**Theory**

The MLQ is based on the constructs of transactional and transformational leadership. Transactional leadership is defined as rewarding or disciplining one's followers based on the level of their performance (Bass, 1996). It focuses on the exchange between a leader and follower that is based on conditions as specified by the leader. There are essentially three components of transactional leadership. They are: 1) contingent reward; 2) management-by-exception-active; and 3) management-by-exception-passive. Contingent reward is tapped by nine items, and refers to rewarding followers after obtaining their agreement on a task and once the task is accomplished. The leader assigns or gets agreement on what needs to be done, and rewards others or promises to in exchange for satisfactorily completed assignments. Management-by-exception-active (MBE-A) is measured by seven items and refers to the style where the leader is actively tracking mistakes in the follower's assignments and taking corrective action when necessary. The leader actively monitors discrepancies from standards, mistakes, and errors in the followers' actions and takes corrective action. Management-by-exception-passive (MBE-P) is measured by seven items, and refers to a leader only taking action once a mistake has been made. In other words, the leader waits passively until errors are made and then takes corrective action. Laissez-faire leadership, measured

by eight items, is described as the absence of leadership and represents a nontransaction situation (Bass 1996).

Transformational leadership expands on transactional leadership by identifying the leader style as one that motivates followers to move beyond their performance expectations. This type of leader is one who employs charisma, inspiration, intellectual stimulation, and/or individualized consideration in order to attain superior results (Bass, 1996). Charismatic behavior results in followers admiring, respecting, and trusting them. The followers identify with the leader and attempt to emulate them. The leader becomes the role model: 1) by engaging in behaviors such as considering the needs of others over his or her personal needs; 2) can be counted on to do the right thing; and 3) demonstrates high standards of ethical and moral conduct. A charismatic leader also: 1) takes risks that are shared by followers; 2) is consistent in his or her behavior; and 3) lives by high ethical and moral standards. This component is measured by ten items.

The second component, inspirational motivation, is defined as the behavior a leader engages in to motivate and inspire followers by challenging them and providing meaning in their work. The leader gets the followers enthused and optimistic, and involves them in envisioning attractive future states. The leader also clearly communicates expectations that followers strive to meet, and demonstrates commitment to goals and a shared vision. This dimension is measured by ten items.

The third component, intellectual stimulation, is when the leader stimulates their followers' effort to be innovative and creative. This is accomplished by questioning assumptions, re-framing problems, and approaching old situations in new ways. Followers' ideas are not criticized and they are challenged to try new approaches. Leaders encourage followers to thoroughly think and rethink solutions to problems. The leader challenges followers to be creative and innovative, even if the generated ideas are not similar to the leader's own ideas. This dimension is tapped by nine items.

The final component of transformational leadership is defined as when transformation leaders act as coaches or mentors by paying attention to each individual's needs for achievement and growth. A two-way communication channel is encouraged and the leader listens effectively. The leader also delegates tasks as a means of

developing followers, with monitoring to ensure additional support or direction are available if needed. This dimension is measured by ten items (Bass, 1996).

**Development and Empirical Use**

The MLQ was developed by gathering accounts of leaders that met the transforming leader criteria. These accounts were turned into 141 behavioral statements, which were then assessed by eleven judges, resulting in 73 items reflecting transactional or transformational leadership. Principal component factor analyses were completed on the frequency 196 U.S. Army colonels said each of the items described one of their immediate superiors. Numerous subsequent factor analyses, LISREL, and Partial Least Squares analyses supported the components that emerged (Bass, 1985; Avolio & Howell, 1993; Avolio et al., 1995). Further behavioral examples of leadership types were gathered using the diaries of VMI cadets. These cadets reported behavioral examples of leadership types from leader observations during a given set of days. These logs were scored in terms of the components from the factor analysis and correlated with independently obtained MLQ results (Atwater, Avolio, & Bass, 1991). Interviews with executives about leadership they had observed produced other behavioral examples of transformational leadership that matched the MLQ (Yokochi, 1989).

Replication for the purpose of assessing Bass' transactional and transformational leadership theory was conducted by Bycio, Hackett, and Allen (1995). They obtained a sample from registered nurses belonging to a nursing association. The outcome variables were performance, satisfaction, intent to leave, and organizational commitment. The confirmatory factor analysis was somewhat supportive of the Bass's five-factor model. However, the two-factor Active-Passive model may be a better fit with the data.

**Psychometrics**

In a military setting, the coefficient alphas for the scales ranged from .71 to .91 (mean = .86) for followers; from .75 to .88 (mean = .84) for upper classmen; and from .53 to .86 (mean = .77) for focal cadets. A principle components analysis with varimax rotation was performed using the follower MLQ data, and 11 factors emerged with eigen values of 1.0 or above. The eleven factors are as follows: 1) inspiration; 2) management-by-exception-passive; 3) management-by-exception-active; 4) charismatic behavior; 5) individualized consideration; 6) intellectual stimulation; 7) laissez-faire; 8) passive versus

active management-by-exception; 9) transformational leadership; and 10) two uninterpretable factors. The first factor, inspiration, accounted for the majority of the variance (35.8%). The second factor, management-by-exception-passive accounted for 8.4%, and the third factor of management-by-exception-active for 4.5%. Together the eleven factors accounted for 61% of the variance.

A sample of 1053 followers from a single organization rated their leaders using the components of the 70 item MLQ Form 5. The results are presented below. In addition, similar reliabilities have been obtained for more recent Form 5X for 2080 respondents from 12 different organizations (Bass, 1996b).

Transformational

Charismatic (Idealized Influence) ($\alpha$ = .89)

Inspirational Motivation ($\alpha$ = .76)

Intellectual Stimulation ($\alpha$ = .86)

Individual Consideration ($\alpha$ = .89)

Transactional

Contingent Reward. ($\alpha$ = .89)

Management by Exception

Active ($\alpha$ = .74)

Passive ($\alpha$ = .73)

Laissez-faire Leadership ($\alpha$ = .79)

**Generalizability**

In terms of generalizability, the instrument has been used in numerous studies in the contexts of business, industry, military, government, educational institutions, and non-profit organizations (Bass, 1996b).

**Face Validity/Ease of Use/Transparency**

It is a paper and pencil measure, and is easy to administer. Items appear to have face validity, and the measure is easy to complete by participants.

## Cadet Performance Report

| | |
|---|---|
| Purpose | Used to evaluate cadet performance at USMA |
| Population | Cadets at USMA |
| Acronym | CPR |
| Scores | 1) Duty motivation; 2) Military bearing; 3) Teamwork; 4) Influencing others; 5) Consideration for others; 6) Professional ethics; 7) Planning and organizing; 8) Delegating; 9) Supervising; 10) Developing subordinate; 11) Decision making; 12) Oral and written communication; 13) Global rating |
| Administration | Paper and pencil, self, peers, supervisors |
| Price | N/A |
| Time | 10 minutes |
| Authors | Schwager & Evans (1996) |
| Publisher | ARI |

### Theory

The CPR was designed to provide a common benchmark of Army Cadets' training performance that could be tracked over time. It stems from an analysis by the Office of Institutional Research Analysis regarding USMA cadets' performance in a variety of leadership roles (Schwager & Evans, 1996). The measure consists of 12 dimensions: 1) duty motivation; 2) military bearing; 3) teamwork; 4) influencing others; 5) consideration for others; 6) professional ethics; 7) planning and organizing; 8) delegating; 9) supervising; 10) developing subordinates; 11) decision making; and 12) oral and written communication. These dimensions are similar to those in two other Army classification systems: 1) Center for Army Leadership (CAL) competencies; and 2) Leadership Assessment Program (LAP) taxonomy.

### Development and Empirical Use

The CPR was originally developed from a job analysis by USMA's Office of Institutional Research, and used as a tool for observing and rating cadet performance (Schwager & Evans, 1996). The content validity of the instrument has been established (OIR, 1989), but construct validity is currently in progress. Construct validity is being assessed by employing the CPR as a measure of leadership behavior for a program of longitudinal leadership development research.

**Psychometrics**

An inductive approach to construct validation was used. The first step was a comparison of the twelve dimensions to other leadership performance measures (e.g., CPR global score and leadership grade). The second step involved examining the interrelationships among the dimensions in order to comprehend the conceptual structure of the instrument. The final step was to examine how different raters (e.g., peers, supervisors, and subordinates) used the various dimensions (Schwager & Evans, 1996).

Each of the twelve dimensions was correlated with the global CPR rating. The following are the mean correlations across all types of raters: 1) duty motivation (r = .58); 2) military bearing (r = .48); 3) teamwork (r = .35); 4) influencing others (r = .35); 5) consideration for others (r = .29); 6) professional ethics (r = .26); 7) planning and organizing (r = .26); 8) delegating (r = .18); 9) supervising (r = .21); 10) developing subordinates (r = .32); 11) decision making (r = .21); and 12) oral and written communication (r = .22) (Schwager & Evans, 1996).

The 12 dimensions were found to be interrelated, with four broader factors emerging from the principal components analyses. The following four components were hypothesized from the analyses: 1) cognition; 2) formal interpersonal; 3) informal interpersonal; and 4) self-management. For the cognition factor, the three dimensions of planning and organizing (.75), decision-making (.73), and oral and written communication (.69) loaded the highest. Delegating (.73), supervising (.73), and developing subordinates (.64) loaded highly on the formal interpersonal factor. The informal interpersonal factor was composed of teamwork (.56), influencing others (.54), consideration for others (.71), and professional ethics (.53). The self-management factor contained the dimensions of duty motivation (.81) and military bearing (.80) (Schwager & Evans, 1996). The results also indicated that different raters placed more emphasis on different dimensions.

**Generalizability**

The CPR is similar to other Army classification systems (CAL and LAP), and relates to leadership behavior in general, as well as to military leaders (Schwager & Evans, 1996). Therefore, this instrument may generalize to populations other than the military academies.

**Face Validity/Ease of Use/Transparency**

The instrument is brief with only one item per dimension, which makes it easy to use. However, this may lead to questions about the comprehensiveness of assessment for the dimensions. One item may be tapping the dimensions at a very general level. The items appear to be face valid and fairly transparent due to the direct nature of the items.

## Leader AZIMUTH/Strategic Leader Development Inventory

| | |
|---|---|
| Purpose | 360-degree evaluation and feedback process that was designed to be used by Army officers as a means of guiding their leadership self-development plans. |
| Population | Military and Civilian Leaders, students at Combined Arms and Services Staff School (CAS3) classes and Command and General Staff Officer Course |
| Acronym | AZIMUTH/SLDI |
| Scores | 1) Communication/influence; 2) Political skills; 3) Problem solving skills; 4) Planning/organizing skills; 5) Ethics; 6) Team-focused supervision; 7) Mission-focused supervision; 8) Compulsive behavior; 9) Self-centeredness; 10) Social maturity; 11) Interpersonal supervision; 12) Tactical and technical knowledge |
| Administration | Paper and pencil, individual, subordinate, supervisors, peers |
| Price | N/A |
| Time | 10-15 minutes |
| Authors | ARI, Army War College (AWC), and the Industrial College of the Armed Forces (ICAF); Keene, Halpin, & Spiegel (1996) |
| Publisher | ARI |

### Theory

AZIMUTH was derived from a previous instrument, the SLDI, which was part of a joint project between ARI, AWC, and ICAF institutes under the direction of T. Owen Jacobs. The theoretical basis for the SLDI is SST, which puts forth the premise that leadership positions at different levels in hierarchical organizations demand different skill sets to be effective. The SLDI was developed primarily to assess the abilities of and needs for development of strategic leaders. That instrument was based on personal interviews with over one hundred general officers, and on information provided by Army War College students.

The factors addressed by the SLDI are as follows: 1) strong work ethic; 2) political sensibility; 3) conceptual flexibility/complex understanding; 4) long-term perspective; 5) arrogant/self-serving/unethical; 6) team performance facilitation/rigid/micro-manages; 7) professional maturity/personal objectivity/explosive/abusive; 8) empowering subordinates; and 9) quick study/perceptive/technical competence (Stewart, Kilcullen, & Hopkins, 1994). This instrument has four different forms to be used by peers, self, subordinates, and

supervisors. These sources caused some practical problems in integrating the ratings to provide feedback. This was especially difficult due to the fact that scores on the various factors for the sources differed in terms of derivation. Thus, comparisons among sources were made more difficult as the factors were not necessarily equivalent. Additional problems relating to applicability of the instrument to leadership levels other than strategic and a lack of coverage of the Army leadership competencies led to the development of the Azimuth (Keene et al., 1996).

**Development and Empirical Use**

The start of the AZIMUTH development was based on data collected at CAS3 at Forth Leavenworth. Approximately 3000 junior officers who attended this nine week course were administered the SLDI in 1994. A factor analysis revealed that the factor structure for junior officers differed from that obtained from the original senior officer sample. Based on this analysis, weak items were removed and replaced with new items to bolster the factor structure. The instrument was also revised to require only one form for all sources, thus solving problems with integration of feedback. This version also purports to apply to all leadership levels, not just strategic leadership (Keene et al., 1996).

The elements which make up AZIMUTH are as follows: 1) communication/influence; 2) political skills; 3) problem solving skills; 4) planning/organizational skills; 5) ethics; 6) team-focused supervision; 7) mission-focused supervision; 8) compulsive behavior; 9) self-centeredness; 10) social maturity; 11) interpersonal supervision; and 12) tactical and technical knowledge (Keene et al., 1996). The instrument is comprised of 98 items, with a six-point scale from A (extremely poor description) to F (extremely good description), as well as a seventh "not applicable, cannot assess" category.

Respondents are instructed to examine items that contain either desirable or undesirable qualities in a leader. Then, they are asked to consider the ratee on these items in comparison to familiar colleagues at the approximate age and position of the ratee. The feedback element of the instrument provides information from the four sources on each of the twelve elements. Graphic feedback illustrates comparison group scores on the elements (Keene, et al., 1996).

**Psychometrics**

Due to the early developmental stage of AZIMUTH, reliability and validity information is still pending. The alpha coefficients for the twelve elements ranged from .26 to .54 (mean alpha = .44) for a sample of approximately 545 CAS3 students. The Spearman-Brown coefficients fell between .48 and .78 (mean Spearman = .70) (Keene et al., 1996). Clearly these fall short of traditional minimal conventions of .60-.70 (Nunnally & Berstein, 1996).

Currently, a second version of AZIMUTH is being developed with the following elements: 1) communicating; 2) decision making; 3) motivating; 4) developing; 5) building; 6) learning; 7) planning; 8) executing; 9) assessing; 10) respect; 11) selfless service; 12) integrity; 13) technical and tactical skills; 14) conceptual skills; 15) critical thinking; 16) metacognition; and 17) epistemic beliefs/other. Research will be focused on establishing a stable factor structure with Army captains and majors, further examination of reliability and validity, and constructing training for individuals with areas in need of improvement (Keene et al., 1996).

**Generalizability**

In terms of generalizability, the instrument has been administered to several CAS3 classes, to some students at the Command and General Staff Officer Course, and to military and civilian leaders at a Training and Doctrine Command installation. It is likely to have greater generalizability than did the SLDI, but given its content, the instrument's boundaries are likely to remain within the Army.

**Face Validity/Ease of Use/Transparency**

Administrative work is more detailed and complex with this instrument due to the 360-degree format. More paper work and time are needed to interpret results for each individual. The instrument appears to have desired face validity, with users regarding it as "user friendly" (Keene et al., 1996).

## Benchmark Instruments

### Leader Practice Inventory

| | |
|---|---|
| Purpose | Feedback for self-development on various sets of leadership behaviors |
| Population | Students, domestic/international managers |
| Acronym | LPI |
| Scores | 1) Challenging the process; 2) Inspiring a shared vision; 3) Enabling others to act; 4) Modeling the way; 5) Encouraging the heart. |
| Administration | Paper and pencil, individual, peers, subordinates, supervisors |
| Price | N/A |
| Time | 20 to 40 minutes |
| Authors | Posner & Kouzes (1990) |
| Publishers | Center for Creative Leadership |

**Theory**

The measure is based on a fundamental pattern of leadership behavior, that covers the five leadership practices (Posner & Kouzes, 1990): 1) Challenging the process, (a) search for opportunities, (b) experiment and take risks; 2) Inspiring a shared vision, (a) envision the future, (b) enlist the support of others; 3) Enabling others to act, (a) foster collaboration, (b) strengthen others; 4) Modeling the way, (a) set the example, (b) plan small wins; and 5) Encouraging the heart, (a) recognize contributions, (b) celebrate accomplishments.

**Development and Empirical Use**

The scale began with qualitative development on what leaders do. Managers were asked to describe personal best experiences as a leader (approximately 1000 case studies). The result was a 12-page document with 37 open-ended questions. Next, 38 in-depth interviews lasting 45-60 minutes were conducted with various managers. The case studies were content analyzed, showing more than 80% of the behaviors and strategies described in respondents' personal best case studies as overlapping with the categories listed above (Posner & Kouzes, 1990).

After development, 120 MBA students with approximately half having supervisory experience originally completed the measure. Then, an item by item

discussion to replace and revise difficult and inconsistent items with the MBA sample, HRM, OB, and Psychology professionals was undertaken (Posner & Kouzes, 1990).

The measure was then administered to 2100 managers and executives. A factor analysis was conducted and internal reliabilities computed. The factor analysis extracted five factors with eigen values greater than 1.0, accounting for 59.9% of variance (Posner & Kouzes, 1990).

**Psychometrics**

Internal reliabilities ranged from .79 to .90, with reliabilities ranging from .70-.84 on the LPI-Self to .81-.91 on the LPI-other. Test-retest reliabilities from an MBA sample averaged nearly .94. Social desirability responses using the Marlowe-Crowne Personal Reaction Inventory resulted in no significant correlations (Posner & Kouzes, 1990).

Criterion-related validity evidence was available from two sources. First, stepwise regression analysis found a highly significant regression equation explaining nearly 55% of variance on the leadership practice model of subordinates assessment of their leader's effectiveness. Second, discriminant analysis, a classification technique, was used. The discriminant function correctly classified 92.62 % of the known cases. When the middle of the sample (e.g., managers with moderate effectiveness scores) were included, the discriminant functions were able to classify 71.13% of the cases ($p < .001$) (Posner & Kouzes, 1990).

For both feedback (self-development) and research purposes, the LPI (other version) appears to provide relatively reliable and valid assessments of respondent behavior. More than one half of the variance of subordinates' evaluations of their managers' effectiveness can be explained by their perceptions of the manager's behavior along the conceptual framework of the LPI (Posner & Kouzes, 1990).

An additional study was completed with 36,000 mangers and subordinates to reexamine the psychometric properties of the instrument. The findings were very similar to those listed above, with the same factor structure emerging (Posner & Kouzes, 1996).

**Generalizability**

The measure has been used with students, managers/executives, and foreign managers, so generalizability seems broad.

**Face Validity/Ease of Use/Transparency**

The measure is paper and pencil, and can be self-scored. Tests for social desirability response bias found no significant results, so transparency is minimal. A review of sample items showed moderate face validity in our opinion.

## Leader Behavior Description Questionnaire (Form XII)

| | |
|---|---|
| Purpose | Provide subordinate descriptions of leader behavior |
| Population | Students, government, military, organization |
| Acronym | LBDQ-XII |
| Scales | 1) Initiating Structure; 2) Consideration |
| Administration | Paper and pencil, subordinates |
| Price | N/A |
| Time | 30 to 45 minutes |
| Authors | R. M. Stodgill (1963) |
| Publisher | Bureau of Business Research, The Ohio State University |

### Theory

The LBDQ and LBDQ-XII measures are the result of leadership studies conducted at Ohio State University. The major objective of the Ohio State Leadership Studies was to identify effective leadership behaviors (Yukl, 1994). These studies found that leadership could be described by two constructs. The first construct is initiating structure, which is defined as production-oriented or task-focused behavior. The second factor, consideration, is the degree of concern a leader has for his or her followers. There are twelve dimensions tapping these two dimensions, which are: 1) representation; 2) demand reconciliation; 3) tolerance of uncertainty; 4) persuasiveness; 5) initiating structure; 6) tolerance of freedom; 7) role assumption; 8) consideration; 9) production emphasis; 10) predictive accuracy; 11) integration; and 12) superior orientation.

### Development and Empirical Use

During the late 1940s through the mid-1950s, Ohio State University was the site of a major research program that focused on leader behavior. Researchers conducted surveys and observational studies in order to determine what behaviors leaders perform. The initial task was to develop questionnaires for subordinate use in describing the behaviors of managers and leaders. Nine initial dimensions of leader behavior were set forth tentatively. These dimensions were as follows: 1) integration; 2) communication; 3) production emphasis; 4) representation; 5) fraternization; 6) organization; 7) evaluation; 8) initiation; and 9) domination. These dimensions were used as a framework to gather leader behavior items. Over 1800 items were developed to described the

various facets of leader behavior. The source of these items was information gathered by individuals in leadership positions and work group members from many organizations.

These items were then examined and categorized into one of the nine dimensions. Then, the items were discussed as to: 1) content overlap; 2) independence from items in other dimensions; 3) content range; 4) general evaluative tone; and 5) other criteria. This discussion resulted in about 200 items being selected, with that number being reduced to 150. The items were then subcategorized within the nine dimensions in order to examine the content emphasis. This led to a revision of the dimensions in order to correspond to the actual item content. The nine revised dimensions were: 1) initiation; 2) membership; 3) representation; 4) integration; 5) organization; 6) domination; 7) up communication; 8) down communication; 9) recognition; and 10) production.

The LBDQ was administered to 357 individuals, of which some described a leader of a group to which they belonged and others described themselves as leaders. From this data, an item analysis was conducted. In order for an item to be considered useful, the responses needed to be attractive enough to be used in at least some of the leader descriptions.

**Psychometrics**

For an Army officer sample, the Kuder-Richardson internal reliabilities ranged from .58 - .85 (Stodgill, 1963). An administrative officer sample had IRs from .66 - .87. The IRs for the corporation president sample ranged from .54-.84. Schriesheim and Stogdill's (1975) study of university employees obtained Kuder-Richardson internal reliability coefficients of .90 and .78 for consideration and initiating structure. They also conducted a factor analysis with varimax rotation and obtained four primary factors, with consideration and initiating structure being the first two. Consideration and initiating structure sub-scales have been found to correlate with role ambiguity, role conflict, and job satisfaction. However, there is some concern about the correlation between consideration and initiating structure (Schriesheim & Stogdill, 1975).

For a student sample, alphas of .83 and .74 were found for initiating structure when students were told to fake good and fake bad (Schriesheim, Kinicki, & Schriesheim, 1979). The alphas for consideration as a result of these two sets of instructions were .84 and .74 (Schriesheim et al., 1979). Another student sample

completed the measure, and seven of the ten initiating structure items and all consideration items were found to be descriptive of socially desirable leader behaviors. Therefore, these scales, especially consideration, may result in lenient descriptions of leader behavior.

A hierarchical factor analysis was conducted with a field sample (e.g., maintenance employees and white-collar employees) in order to establish any response biases. For the first sample, the factor loadings revealed that one of the consideration items did not load above .30 on the apex factor. The initiating structure items faired worse, with seven of the ten items loading below .30. For the second sample, two consideration items did not reach the .30 level, while four structure items loaded below .30. Thus, consideration is more of a problem in terms of leniency (Stodgill, 1963).

The consideration and initiating structure scales were evaluated for leniency effects on the relationships with various dependent variables (Schriesheim et al., 1979). Zero-order correlations were calculated for 8 dependent variables: 1) satisfaction in general; 2) satisfaction with supervision; 3) group productivity; 4) group cohesiveness; 5) group drive; 6) on-the-job anxiety; 7) role clarity; 8) and role conflict. These correlations were compared with partial correlations that controlled for leniency. The difference between the two types of correlations revealed the decrease in explained variance as caused by controlling for leniency. The results showed that the differences for initiating structure are inconsequential. However, the results showed that leniency has a great deal of influence on correlations with the 8 dependent variables. All of the dependent variables, with the exception of role clarity, suffered at least a 45% decrease in explained variance (Schriesheim et al., 1979).

**Generalizability**

In terms of generalizability, development and early use had a broad range of samples, such as army officers, administrative officers, corporation presidents, students, engineering managers, maintenance employees, and white-collar employees of a large heavy equipment manufacturer. However, it is not clear whether more recent samples have been as broad.

**Face Validity/Ease of Use/Transparency**

Items appear to have face validity and the measure is easy to use, although the full instrument is lengthy. There appears to be a concern about leniency, especially for the Consideration scale (Schriesheim et al., 1979). This should be carefully considered due to the inflation leniency causes with relationships to dependent variables.

## *Benchmarks*

| | |
|---|---|
| Purpose | 360-degree evaluation to assess leadership skills and enhance the development process. |
| Population | Managers of various levels |
| Acronym | N/A |
| Scores | 1) Handling the challenges of the job; 2) Leading people; 3) Respecting self and others; 4) Problems that can stall a career; 5) Handling challenging job assignments |
| Administration | Paper and pencil, individual, peers, subordinates, and supervisors |
| Price | Set of 12 surveys (1 self and 11 others) $245; $215 per set if organization does administration and collection of instruments; $180 per set for remote scoring. Two day workshop for certification for Benchmarks, $1,000.00 |
| Time | 30-40 minutes |
| Authors | Center for Creative Leadership (1996) |
| Publishers | Center for Creative Leadership |

**Theory**

*Benchmarks* is an outgrowth of a continuing study that focuses on key events that have impacted the careers of high potential managers. The assumption is that critical leadership lessons can and must be learned by challenging experiences and learning. These experiences teach important lessons. The lessons learned are not random, but flow from specific experiences. *Benchmarks* uses these experiences to provide feedback on three areas: 1) leader skills; 2) problems that can stall a career; and 3) handling challenging job assignments. The first area taps specific leadership skills from different sources. These different perspectives help managers understand how they and others see their leadership skills. The following skills and perspectives are addressed in this section:

A. Handling the challenges of the job
   1) Resourcefulness
   2) Doing whatever it takes
   3) Being a quick study
   4) Decisiveness

B. Leading People
   5) Leading employees
   6) Setting a developmental climate
   7) Confronting problem employees
   8) Work team orientation
   9) Hiring talented staff

C. Respecting self and others
   10) Building and mending relationships

11) Compassion and sensitivity

12) Straightforwardness and composure

13) Balance between personal life and work

14) Self-awareness

15) Putting people at ease

16) Acting with flexibility

The second section measures problems that can stall a career. There are six problem areas that are addressed in this section that may lead managers to derailment. These areas are: 1) problems with interpersonal relationships; 2) difficulty in molding staff; 3) difficulty in making strategic transitions; 4) lack of follow through; 5) over-dependence; and 6) strategic differences with management. The third section assesses how individuals handle a variety of challenging job assignments (CCL, 1996).

**Development and Empirical Use**

Seventy-nine in-depth interviews with successful male executives in three Fortune 100 companies were conducted. Following that, seventy-six interviews were conducted with women executives from the same companies (Morrison, White, & Velsor, 1992). Interviews were content analyzed, with 16 categories of critical development events emerging. Next, 112 high performance executives responded to the key interview questions via an open-ended questionnaire to confirm the 16 categories identified. The 16 categories were further classified as job assignments, events involving people, hardships, and miscellaneous events (Velsor & Leslie, 1991).

Once the categories were identified, items were constructed. Other researchers and human resource professional reviewed the initial items. Revisions were made from the feedback of the reviews, and the remaining items were pre-tested with a group of executives and human resource professionals. Items were deleted, refined, or added based on this step with the final pool consisting of 274 items; 210 in Section 1; 46 in Section 2; and 18 in Section 3 (Velsor & Leslie, 1991).

Items covering the same concept were clustered into scales in each section by eigen values derived from a factor analysis based on a sample of 336 managers. Further item analysis and evaluation of conceptual overlap in the items led to refinement of the groupings.

**Psychometrics**

The scale reliability for Section 1 is .88; the average test-retest for self-rating is .72; the average test-retest for others is .85; and the interrater agreement is .58. The scale reliability for Section 2 is .83; the average test-retest for self-rating is .55; the average test-retest for others is .72; and the interrater agreement is .43. The scale alphas were calculated using data from the scale construction process. For self test-retest and interrater agreement analyses, the sample consisted of 75 managers from different organizations with two or more co-workers. For test-retest of the others scale, a sample of 33 managers were rated by a coworker (CCL, 1996).

Concurrrent validity of the *Benchmarks* measure was determined by correlational analysis with:

1) overall assessment by the boss on the manager's promotability, using a six-point scale;
2) independent rating by corporate management committee on level of satisfactory performance by one organization;
3) performance evaluation rating two years after initial Benchmarks administration; and
4) subsequent movement of manager within organization during 24 to 30 months after initial Benchmarks ratings (Velsor & Leslie, 1991).

*Benchmarks* ratings are also correlated with scores from the MBTI, Kirton Adaptation-Innovation Inventory and Shipley Institute of Living Scale, providing some evidence of construct validity (Velsor & Leslie, 1991).

**Generalizability**

Generalizability is strong in domestic and international management settings. There has been no use of the measure in non-management contexts.

**Face Validity/Ease of Use/Transparency**

This measure is a commercial product, and is available in many languages and used in many countries. The instrument is paper and pencil, but *Benchmarks* can only be administered by a certified professional facilitator. Remote scoring of the measure is available. It appears to be easy to complete, but may be more difficult to administer based on the use of many sources. Our evaluation of the sample items showed them to be somewhat transparent, but very face valid. Normative data is available.

## Campbell Leadership Index

| | |
|---|---|
| Purpose | 360-degree measure to determine leader perception an agreement between the leader and his/her subordinates |
| Population | Students, military, organization, leader trainers |
| Acronym | CLI |
| Scores | 1) Leadership; 2) Energy; 3) Affability; 4) Dependability; 5) Resilience |
| Administration | Paper and pencil, individual and 3-5 observers |
| Price | Set of 6 surveys and feedback reports $160; Manual $50.00 |
| Time | 45 to 60 minutes |
| Author | David Campbell (1991) |
| Publisher | National Computer Systems |

### Theory

For the purpose of this measure, the developer defined leadership as actions that focus resources to create desirable opportunities. Actions of leadership include a wide range of behaviors such as planning, organizing, managing, or any behavior that leads to a high probability of a desirable organizational outcome. Resources on which leaders must focus include people, money, time, and more nebulous assets, such as public opinion, unique talents, etc. Desirable opportunities created by effective leaders include profits, education, and in general any increase in truth, beauty, and happiness (Campbell, 1991).

The definition itself is not sufficiently detailed to provide much guidance about which personal characteristics should be assessed by a leadership index. Therefore, assumptions were made about the seven crucial tasks that a leader must face to be successful. These seven tasks are: 1) vision; 2) management; 3) empowerment; 4) politics; 5) feedback; 6) entrepreneurship; and 7) personal style (Campbell, 1991). The first task, vision, is the clarification of the overall goals of the organization. The second task, management, is the ability to focus resources on the organizational goals, and then to monitor and manage the use of these resources. The third task, empowerment, is defined as the ability to select and develop subordinates committed to the organization's goals. Politics, the fourth task, is the ability to forge coalitions with peers, superiors, and important outside decision makers. The fifth task, feedback, is the ability to listen carefully to organizational members, clients, customers, voters, students, alumni, and

other relevant groups, and then react appropriately. The sixth task, entrepreneurship, is the ability to find future opportunities and to create desirable change. The last task, personal style, is defined as setting an overall organizational tone of competence, integrity, and optimism by personal example (Campbell, 1991).

Campbell (1991) argued that no leader is strong on all of these tasks, and therefore, they need to be aware of their personal shortcoming so that they can work on the weak areas. The instrument is based on the seven tasks, and identifying the strengths and weaknesses of leaders on each task (Campbell, 1991).

**Development and Empirical Use**

The development of CLI was guided by the experience of Campbell and four sources of input which were: 1) informed discussions; 2) literature review; 3) case studies; 4) interviews; 5) anecdotes and biographies; and 6) personal opinion.

The first step in instrument development was to generate a pool of adjectives. Next, the adjectives were defined, revised, and their number were reduced based on redundancy and social acceptability. There were 300 adjectives, which dropped to 160 based on redundancy and time considerations. After pilot tests with several thousand respondents in standardized testing, a total of 100 adjectives remained. Scales were developed in a statistical/intuitive way, with consideration of trying to develop a list of scoring scales that was particularly related to the factors underlying successful leadership. This consideration guided the selection of adjectives to be used in the Index. Statistical data were used when available to guide decisions, but a fair amount of reasoned judgment was also used to determine what factors should be included in the profile. As a result, five major scales or orientations were developed with 22 measures of specific leadership characteristics. The four orientations are **L**eadership, **E**nergy, **A**ffability, **DE**pendability, and **R**esilience (Campbell, 1991).

**Psychometrics**

The alpha coefficients ranged from .56 to .90; the interrater reliability ranged from .68 to .82; the test-retest (fraternity student sample) was .91 for self ratings and .89 for other ratings; for the women business forum sample, self ratings had test-retest correlations of .87, and .85 for other ratings (Campbell, 1991).

The content validity of the adjectives within each scale was statistically interrelated, and the cluster of each adjective focused on the topic represented by the scale's title. The concurrent validity was built into the scoring system since ratings were available from at least three observers on an extensive checklist of adjectives. Of course, this did not offer criterion-related validity evidence, but more in terms of construct validity support. To address construct validity, mean profiles from a variety of samples were calculated and plotted. For discriminant validity, self and observer ratings showed reasonable agreement. In addition, the scales (orientations) while not completely unrelated did provide unique information about the person being assessed. Preliminary norms are available (Campbell, 1991).

**Generalizability**

Generalizability of this instrument is broad. It has been used on samples of university students, military academy cadets, mangers, senior executives, fire chiefs, leader seminar participants, and trainers.

**Face Validity/Ease of Use/Transparency**

This is a commercial product available for a fee. The format is a paper and pencil, self-report adjective checklist that makes completion easy. However, due to the 360-degree nature of the measure, more administrative work is necessary in terms of distributing the reports and compiling the results. The sample items reviewed seem face valid, but somewhat transparent.

## PROFILER

| | |
|---|---|
| Purpose | 360-degree instrument to provide feedback on competencies required to be a successful manager. |
| Populations | Mid-level managers |
| Acronym | N/A |
| Scores | 1) Thinking skills; 2) Administrative skills; 3) Leadership; 4) Interpersonal skills; 5) Communication; 6) Motivation; 7) Self-management; 8) Organizational knowledge. |
| Administration | Paper and pencil, individual, superiors, peers, subordinates |
| Price | One set of 1 self and 10 other questionnaires, scoring, and feedback report, $275.00 |
| Time | 35 to 50 minutes |
| Authors | Holt & Hazucha, (1991) |
| Publishers | Personnel Decisions Inc. |

### Theory

PROFILER was developed from a model of managerial performance and effectiveness described by Campbell, Dunnette, Lawler, and Weick (1970), as well as from assessment center research. It focuses on specific job-related skills, rather than on managerial style or other abstract concepts that are difficult to translate into the job behaviors. The first factor in the eight-factor model is thinking skills, which includes how well leaders: 1) gather information systematically; 2) consider a broad range of issues or factors; 3) seek input from others; and 4) use accurate logic in analyses. The second factor, administrative skills, measures how well the leader establishes plans and manages execution. The third dimension is leadership, which measures the facets of: 1) providing direction; 2) leading courageously; 3) influencing others; 4) fostering teamwork; 5) motivating others; and 6) coaching/developing. The next dimension, interpersonal skills, measures how well the leader: 1) builds relationships; 2) displays organizational savvy; and 3) manages disagreement. Communication, the next dimension, taps speaking effectively, fostering open communication, and listening to others. The next factor is a motivational dimension that measures drive for results and showing work commitment. The sixth dimension, self-management, is comprised of: 1) acting with integrity; 2) demonstrating adaptability, and 3) developing oneself. The last factor, organizational knowledge, measures technical/functional expertise, knowledge of the business, and overall performance (Holt & Hazucha, 1991).

**Development and Empirical Use**

In the early 1980's, the precursor to the PROFILER, the Management Skills Profile (MSP), was developed for the purpose of differentiating between effective and ineffective managers. A content-related approach to the development and validation of the MSP began with a literature review to identify dimensions. Items were then written and rewritten, with repeated piloting. A factor analysis of the 19 scale scores for 1096 managers resulted in the 4 factors of: 1) cognitive skills; 2) human relations skills; 3) administrative skills; and 4) leadership skills. In 1991, another factor analysis of the 19 scale scores for over 14,000 managers yielded the 3 factors of: 1) administrative management; 2) empowering leadership; and 3) individual contributor skills. Separate analyses for the self, supervisor, peer, and subordinate responses have yielded 2-factor structures consistently (Holt & Hazucha, 1991).

In 1990, the focus changed to participative management and teamwork. Therefore, job analysis questionnaires and group interviews were conducted to build more dimensions into the MSP. New individual scales and an overall performance dimension were added to tap these issues. The result was the PROFILER instrument (Holt & Hazucha, 1991).

**Psychometrics**

The corrected item-scale correlations for the PROFILER ranged from .32-.81 (supervisor); .29-.80 (subordinate); .37-.81 (peer); and .17-.78 (self). The test-retest correlations ranged from .50s to .60s, over 12-24 months for the non-self ratings. The test-retest reliabilities for self-ratings were between .36 - .66. Cronbach's alpha for the 19 scales fell between .70-.91, with an average internal consistency of .83. Interrater reliability for peers and subordinates separately ranged from .60-.80. For supervisors, the interrater reliability was .30-.52; for subordinates, .28-.48; for peers, .21-.37; and the average was .28-.47 (Holt & Hazucha, 1991).

The instrument was developed with a content validity approach. Therefore, the instrument has high content validity. Predictive and concurrent validity are in progress.

**Generalizability**

The PROFILER has been mainly used on mid-level managers, but can span broader management levels.

**Face Validity/Ease of Use/Transparency**

The instrument appears to be face valid based on a review of sample items. The items appear to tap behaviors, abilities, and knowledge, making the instrument face valid. The instrument is easy to administer and complete. Once completed, the instrument is returned to PDI for scoring. PDI provides a feedback report with how the individual was rated by the different sources, comparison with norms, and highlights of strengths and development needs. PROFILER certification is required in order to administer the instrument.

## Prospector

| | |
|---|---|
| Purpose | 360-degree measure that assesses an individual's ability to learn from experience. The measure enables leaders to gain insight into their strengths and development needs. |
| Populations | Domestic/international managers, military |
| Acronym | N/A |
| Score | 1) engages in the opportunity to learn; 2) how well the leader creates a context for learning. |
| Administration | Paper and pencil, individual and 11 supervisors, peers, and co-workers. |
| Price | 1 set of 12 surveys, 1 feedback report, 1 learning guide, $195; If organization administers and collects data, $175. |
| Authors | Center for Creative Leadership (1996) |
| Publishers | Center for Creative Leadership |

### Theory

The Prospector measures 11 different dimensions of learning, based on two main components. The first component taps the way the leader engages in opportunity to learn. Specifically, it measures how the leader: 1) seeks opportunities to learn; 2) seeks and uses feedback; 3) learns from mistakes; and 4) is open to criticism. The second component considers how well the leader creates a context for learning for those around him or her and includes: 1) how committed the leader is to making a difference; 2) how insightful the leader is in terms of viewing things from new angles; and 3) having the courage to take risks. It also assesses whether leaders can: 1) bring out the best in people; 2) act with integrity 3) adapt to cultural differences; and 4) seek broad business knowledge (McCall, Spreitzer, & Mahoney, 1996).

### Development and Empirical Use

First, a comprehensive review of the literature on executive development was conducted. In addition, interviews were conducted with experienced corporate executives who had been involved in identifying people with potential to successfully handle international assignments. The individuals interviewed were actively involved in early identification of executive potential. In addition, a sample of non-U.S. executives working in the U.S. for multinational firms was also interviewed. Content analysis on the

data from the interviews suggested that the ability to learn from experience was manifested in the following situations:

1) Individuals seeking out more experiences that provide learning opportunities;
2) Once in the opportunities, some individuals create an environment and act on the environment in ways to produce useful information and feedback;
3) Some individuals are more receptive to information on their performance and incorporate more of that information into future behavior (McCall et al., 1996).

Second, behavioral examples from the interviews were used to create a pool of 200 items, addressing 19 learning dimensions. These items were pre-tested on managers attending an international business education and research program. Questions were refined to a 116-item survey, tapping the 11 dimensions listed above (McCall et al., 1996).

**Psychometrics**

The scale alphas ranged from .76 to .89, based on a sample of 838 managers. Items were reviewed by practicing international managers during the pre-test, and by international HR professionals on two separate occasions in order to address content validity. In addition, a concurrent validity study was conducted with the following criterion measures of (McCall et al., 1996):

1) executive potential - discriminant analysis was conducted with 73% successful identification;
2) current performance - findings showed that those who scored high on the Prospector were also high performing employees, with the same findings for competency measures;
3) on the job learning - learning content knowledge and learning behavioral skills were correlated with the Prospector dimensions;
4) international criteria - predicted success in an expatriate executive assignment and predicted success dealing with international issues, but not as an expatriate correlated with adopts to cultural differences; and
5) derailment potential - each dimension was negatively correlated with the inability to make the transition to a senior management perspective.

Another validation study was conducted with 53 alumni of leadership development programs from CCL. The participants completed the Prospector and a 10-item survey, the Learning Research Questionnaire. The correlation between the two measures provided by bosses was .87. The correlation between the boss' Prospector score and peer ratings on the Learning Research Questionnaire was .35, suggesting differences in rater perspectives (McCall, 1996).

**Generalizability**

Generalizability of this instrument is broad in a management context, including international contexts. It has also been used in some military contexts.

**Face Validity/Ease of Use/Transparency**

The Prospector is a commercial instrument, and costs about $200 per set of measures. The vendor claims that it can only be administered by certified professionals. The items appear to be fairly transparent, which could lead to social desirability problems. At least five raters must complete the form, which creates more administrative work. Normative data are available.

## ARI Measures vs. Benchmarks

**Summary**

In terms what they do and how well they do it, the ARI measures of leader behavior are comparable to the benchmarks. The overall purpose of all instruments is purportedly to assess behaviors that are believed to be related to successful leadership. There is a wide range in the format and administration of the instruments, but most are paper and pencil ratings that are usually completed by subordinates (e.g., MLQ, LBDQ-XII) or employed in a 360 fashion (e.g., AZIMUTH, CPR).

The psychometric properties of the ARI instruments and the benchmarks do vary. The MLQ, LBDQ-XII, LPI, *Benchmarks*, CPI, and Prospector have the strongest reliabilities. The PROFILER and AZIMUTH possess low to moderate reliabilities. In terms of validity, it is our opinion that the CPI and Prospector possess the strongest validity evidence. Following them, *Benchmarks* and the LPI showed strong construct validity. The LBDQ-XII displayed low to moderate construct validity, followed by the MLQ and CPR. These two measures show the weakest evidence of validity. The AZIMUTH validity evidence is still in progress.

The ARI instruments, CPR and AZIMUTH, have been used exclusively in a military context, whereas the MLQ has been used in a variety of contexts. The benchmarks have been mostly administered in management settings, but generalizability to military settings is feasible. In terms of face validity, they are all fairly equal and the same can be said for the transparency of the instruments. There are varying levels of ease of use for the instruments. The MLQ, LBDQ-XII, and CPR are the easiest to complete. The AZIMUTH and the 360-degree *Benchmarks* rank low in terms of ease of use due to administration and scoring difficulties.

**Recommendations**

Whereas the ARI instruments are essentially comparable to those available in the private sector, all, in our opinion, lack a clear focus. A quick review of Table 9 reveals that some instruments focus on leader behaviors (e.g., MLQ, LBDQ-XII), others largely on personality type dimensions (e.g., AZIMUTH, Campbell Leadership Index), and most include a variety of skill assessments. This "mixed-bag" limits the extent to which these indices can be unequivocally employed as predictors or criteria in any given study. It also

presents difficulties when it comes to establishing clear frames of reference for raters and targeted feedback for ratees. In short, there is a need to refocus ratings of leader behaviors on behaviors per se, not on leader attributes.

We submit that it would be advantageous to develop a 360 rating system for Army leaders. The purpose for such ratings, however, would need to be articulated clearly. Raters' motivation and how leaders respond to feedback are driven largely by the purpose(s) of any evaluation. Second, the content of the ratings must be grounded. Based on a thorough analysis of Army leader positions, in the context of Army Doctrine, specific behavioral dimensions should be identified. As discussed elsewhere, once these target dimensions are articulated, items can be designed to assess them and tested using confirmatory analytic techniques. Revisions of SLDI in to AZIMUTH have progressed in this fashion, but need to be more formalized and developed. We would anticipate that some core set of dimensions would be applicable across Army leadership positions, whereas others might be applicable to only a limited range of levels or specialty areas. However, we would expect that the "unique" dimensions should be the exception not the rule. Of course, a thorough job analysis would make this clear.

Third, the sources for such ratings should be clarified, both in terms of who is best positioned to provide information of what type(s), and how the different perspectives will be integrated. We would fully anticipate that the most appropriate sources and combination rule might differ depending on the purpose(s) of the assessment and the target population of leaders. Finally, the process of assessment should be addressed. This includes two sub-issues. First, there are administrative concerns dealing with timeliness, sampling of raters, and an abundance of data. We believe that a fairly generic software program could easily be developed that would help to manage data and to provide customized feedback reports. These exist in the civilian sector and could be easily adapted. What we envision here is that such a program would contain "core rating dimensions" for use with all, or nearly all, leadership positions. Menus could be available for adding supplemental dimensions as warranted. Then, for any particular job class and application, a sample of raters could be generated. In an ideal situation, these forms could be administered and retrieved electronically through e-mail systems. While we offer these as simply ideas, systems such as this are common place in current large-scale

organizations.

The second process issue concerns the feedback and use of such a system. How information is feed back to leaders, what remedial and developmental opportunities are available to them, and what, if any, consequences the process has for them have clear implications for how leaders will react to the process. Answers to these questions and more are what will enable ARI to truly embed leader behavior assessments in to ongoing Army training and assessment programs and to gather data systematically while minimizing the administrative burden of doing so.

Table 9

Leader Behavior Comparison

| | ARI Featured Products | | | Benchmarks | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Criteria** | ***Multifactor Leadership Questionnaire*** | ***Cadet Performance Report*** | ***Azimuth/ Strategic Leader Development Inventory*** | ***Leader Practice Inventory*** | ***Leader Behavior Description Questionnaire*** | ***Benchmarks*** | ***Campbell Leadership Index*** | ***Profiler*** | ***Prospector*** |
| **Behavior** | | | | | | | | | |
| **Consideration** | X | | | | X | | | | |
| **demand reconciliation** | | | | | X | | | | |
| **Initiating structure** | X | | | X | X | | | | |
| **Integrity** | X | | | | X | | | | |
| **role assumption** | X | | | | X | | | | |
| **take risks** | | | | X | | | | | |
| **Cognitive Skills** | | | | | | | | | |
| **conceptual flexibility/ understanding** | | | | | | | X | | |
| **long term perspective** | | | X | | | X | X | | |
| **predictive accuracy** | | | | | X | | | | |

| | ARI Featured Products | | | | | Benchmarks | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Criteria** | ***Multifactor Leadership Questionnaire*** | ***Cadet Performance Report*** | ***Azimuth/ Strategic Leader Development Inventory*** | ***Leader Practice Inventory*** | ***Leader Behavior Description Questionnaire*** | ***Bench-marks*** | ***Campbell Leadership Index*** | ***Profiler*** | ***Prospector*** |
| **technical competence** | | | X | | | X | | | |
| **Interpersonal skills** | | | | | | | | | |
| **Vision** | X | | | X | | | | | |
| **Persuasiveness** | X | | | | X | | | | |
| **Representation** | | | | X | X | | | | |
| **respect** | | X | | | | X | X | | X |
| **Management Skills** | | | | | | | | | |
| **Empowering** | X | X | X | X | | X | X | | |
| **experience** | | | | | | X | X | X | |
| **leading production** | | | | | X | | | | |
| **emphasis** | | | | | | | | | |
| **recognize accomplishment** | X | | | X | | | | | |
| **Developmental climate** | | X | | | | X | | X | X |
| **set example** | X | | | X | | | | | |

| | ARI Featured Products | | | | Benchmarks | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Criteria** | ***Multifactor Leadership Questionnaire*** | ***Cadet Performance Report*** | ***Azimuth/ Strategic Leader Development Inventory*** | ***Leader Practice Inventory*** | ***Leader Behavior Description Questionnaire*** | ***Bench-marks*** | ***Campbell Leadership Index*** | ***Profiler*** | ***Prospector*** |
| **team facilitation** | | X | X | | | X | | | |
| **Personality** | | | | | | | | | |
| **arrogant/ self serving** | | | X | | | | | | X |
| **Dependability** | | X | X | | | X | X | | |
| **Political Sensibility** | | | X | | | | | | |
| **Professional maturity** | | X | X | | | | X | | X |
| **Resilience** | | | | | | | | X | X |
| **superior orientation** | X | | | | X | | | | |
| **Tolerance** | X | | | | X | | | | |
| **work ethic** | | X | X | | | X | X | | |

# Section 6: General Summary and Recommendations

This document has chronicled the development and use of a vast array of leader assessment measures. Moreover, the number of measures reviewed here are but a subset of the ones that have been used by ARI research scientists over the past 10 years. In this section we will attempt to identify some common themes running throughout the body of work that we reviewed. In addition, we offer some recommendations for future research. We caution the reader to appreciate, however, that the following comments must be tempered in terms of the objectives and goals for any assessment effort. We had begun this project with the hopes of classifying clearly the intended purpose(s) of each assessment device we reviewed. Unfortunately, such clarity did not exist. Some measures were used for predicting leader effectiveness, some as indices of leader effectiveness, some as both, yet others as neither. Therefore, our following comments are framed more in terms of reactions and recommendations regarding the utility of assessment procedures and measurement tools in general rather than with an appreciation for the intended purposes of each.

Theory

In terms of the theoretical background driving the ARI work, it is fair to say that a wide spectrum of theories has been utilized. However, Stratified Systems Theory (SST) is, perhaps, the most widely cited and used. As outlined earlier, SST suggests that different leader knowledges and personal orientations (i.e., proclivity) are important as individuals progress through their careers and organizational hierarchies. This suggests that measures of different types of leader knowledge and personal characteristics must be articulated, defined, and assessed. It also suggests that criteria indices of leader effectiveness must exist in order to test the validity of the theory. This places a premium on the kinds of measures included in this review.

Existing Measures

Several promising ARI measurement strategies do exist. In terms of personality assessments, specific facets of the SST proclivity theme have been identified and assessed (e.g., SOI, Biodata). However, it is also fair to say that the proclivity construct has not yet been fully articulated and thoroughly assessed by the efforts and measures that we reviewed. Moreover, the commercial benchmark measures that we reviewed have long track record of successfully

assessing facets of the Big 5 personality framework. We would strongly encourage the incorporation of these types of assessments in efforts designed to examine the role that personality plays in leader effectiveness.

ARI assessment of leaders' knowledge shows some promise. Recall that we differentiated between general types of cognitive abilities such as (problem solving and information processing) and more specific types of knowledge such as tacit or mental models. In terms of the general cognitive abilities, the ARI biodata measures yield several useful indices. As compared to the Fleishman and Quaintance (1984) taxonomy, the biodata indices still lack coverage of 35% of the areas. Accordingly, targeted development of additional subscales would be warranted if a complete sampling of the ability taxonomy is desired. Alternatively, commercial analogues exist that have proven histories of assessing these abilities that should be considered.

As for assessments of more focused types of knowledge, both the ARI tacit knowledge and mental model measures that have been developed show promise. These types of assessments require a substantial investment in the development stage because of two concerns. First, as compared to more generic approaches, these types of knowledges are more embedded in the specific job requirement and organizational settings. In others words, they are grounded more specifically in job conditions and therefore require development efforts that delve more deeply into job nuances. Second, there are no objective right or wrong answers to these types of assessments; so they require either reference against an "ideal response profile" derived from a consensus of experts, or must be evaluated individually by experts. Here, too, one must either devote a substantial amount of time initially to develop the expert template(s), or absorb the ongoing cost associated with ratings of responses. In any case, we should note that we believe that both the tacit knowledge and mental models measures developed by ARI have struck a nice balance in terms of grounding vs. generalizabilty. Both development efforts constructed multiple forms for use with leaders at different organizational levels. While falling short of the "core" dimension theme with supplemental scales that we have advocated, this limited generalizability approach has enabled the researchers to both focus their assessments efforts while not overly confining the use of the measures.

The ARI assessments of leader behaviors (e.g., CPR, AZIMUTH) have been designed for limited applications. As we discussed in Section 5, we believe that the framework or infrastructure for gathering 360 type ratings of leader behaviors could be developed in a fairly

generic fashion allowing for more customized applications in terms of what dimensions are evaluated, by whom, and for what purpose(s), in any given application. Whereas the MLQ instrument affords widespread comparability across settings, it is not designed to hone in on specific requirements of Army leadership positions nor to direct developmental feedback efforts. It, or comparable assessments, are useful for research purposes and for making comparisons across settings, hierarchical levels, etc., but that comparability comes at the expense of applicability to any given circumstance.

Research Protocols

In terms of research protocol, we found that most ARI efforts followed a common approach. First, most started with a good foundation in theory and a description of the larger framework within which the specific effort was targeted. Then, whether it was a prediction or assessment effort, some attention was devoted to identifying the underlying dimensions of leadership to be focused upon. Next, a large number of potential items, observations, etc. (i.e., indicators) of the relevant domain were generated and distilled. Herein lies a weakness of the prototypic method. There was typically a disconnect between the a priori specification of intended underlying dimensions, the indicator generation, and the indicator confirmation. The modal strategy appears to be to generate a large number of potential indicators and then to employ both judgmental techniques and exploratory quantitative data reduction analyses to "reveal" underlying dimensions. In contrast, an apriori approach would first specify the intended dimensions and then generate indicators of those specific dimensions. Next, depending on the number and potential redundancy of indicators, expert judgments could be solicited to combine, refine, and focus the preliminary set of items as related to their intended underlying dimensions. Finally, data can be collected from a preliminary sample that represents the intended boundaries of generalizability for use of the assessment device. Confirmatory analytic techniques can then be applied to test the extent to which the indicators map to their intended underlying dimensions. No doubt some revision will be necessary, and the stability of the resulting structure can be evaluated using additional developmental samples.

The paragraph above describes a fairly standard measurement development protocol. In fairness to the ARI researchers, we believe that they often try to accomplish "too much" in any particular study. That is, there is often an attempt to develop or refine measures while addressing more substantive relations with other variables of interest. While laudable, this dual focus tends

to detract from both aims. The inclination is to "shotgun" the measurement effort in order to ensure that adequate coverage of the domain will be achieved. But, this approach, combined with the use of exploratory data reduction techniques, yields instruments that are not comparable from one study to the next and limits the evolution of knowledge. Now, we fully recognize that different research questions, field applications, and so forth imposed demands on every research investigation. What we advocate, however, is the development of more standardized assessments that can be used intact in a number of different investigations. To achieve this, we recommend the following.

First, a theory or common framework of leader effectiveness needs to be adopted. This is not to say that every study needs to subscribe to a particular theoretical position, but it would hasten the evolution of knowledge if all ARI studies of leadership could at least be described in terms of how they represent certain facets of a given theory. While, naturally, the theory that researchers believe best fits the U.S. Army of the 21st Century is the best candidate for this function, what is more important is that some common yardstick be adopted.

Second, an updated job analysis of Army leadership positions is warranted for the identification of dimensions that are common across positions and those that have more limited representation. Third, an analysis of the important knowledge, skills, abilities, and other attributes important for performing those dimensions should be conducted. Fourth, criteria measures of effective performance of those dimensions should be developed. Given the multiple uses of feedback, a 360 rating framework focused on leader behaviors would likely pay high dividens here. However, other indices of effectiveness should also be considered and incorporated (see below). Fifth, there is a need to move beyond exploratory data analytic methods to more confirmatory techniques. Perhaps the biggest advantage of doing so lies not so much in the statistical tests and model fit indices, as it does in the demands it places on investigators. These analyses require that researchers formulate an a priori framework for the measures they are testing. Sixth, additional explanatory variables should be incorporated to identify the limits of generalizability and potential moderators of relations.

The recommendations in the paragraph above are not new grand insights or revolutionary. Rather, they hearken to a call for getting back to the basics before moving forward. Research scientists are intrinsically and extrinsically rewarded for developing new measures, testing new or innovative ideas, and essentially for moving forward in to uncharted

territory. However, if each study in a program of research introduces a new twist or "refinement" of an assessment technique, then progress is actually stunted not enhanced. As we have mentioned throughout this report, if attention were devoted to establishing measures of core dimensions of Army leadership (whether those be predictors or assessments), along with more specific dimensions for given applications, in the aggregate, ARI research would be facilitated as each new study would have a better foundation from which to begin. This approach would, then, free resources for expanded inquiries incorporating other factors.

Expanding the Framework

The framework that we depicted in Figure 1 was intended to be an organizational device for the measures that we reviewed. Our review of the ARI literature from the past 10 years, however, revealed that most work focused on leader KSAOs and behaviors. Only a few studies addressed other influences depicted such as the task and operational environments, follower characteristics, or effectiveness (i.e., outcome) measures. Tenets of SST suggest that different variables will be important for leader effectiveness depending on the leaders' career stages and level in the organization. Beyond that focus, however, very few studies have considered situational influences on leader effectiveness. Moreover, follower characteristics have been virtually ignored. Clearly the Army of the 21st Century will differ from what we have seen in the past. The shear number of troops and officers will diminish yet the demands on them will increase. While the number of men and women serving will decrease, their average abilities and expectations will surely go up as compared to previous generations. Technological sophistication has changed, and will continue to change, how battles are fought in the future. While some features of effective leadership are timeless, such as the ability to inspire and motivate troops, history has demonstrated that technology changes the nature of warfare and what makes for effective leadership. These factors warrant far more attention as ARI works to understand and enhance leadership in the Army of the future.

There is also a serious need to develop the criteria side of ARI research investigations. Far too many of the leader assessment studies "validated" some measure of, for example, leader knowledge, by correlating scores on it with participants responses on a different type of test (e.g., a situational exercise). Whereas such studies do provide evidence of construct validity for the measure in question, they do not substitute for criterion related validity coefficients. Furthermore, when actual criteria measures have been employed, they have been limited to

ratings of leaders' behaviors. As illustrated in Figure 1, a vast number of effectiveness criteria such as unit performance (e.g., combat effectiveness and resource availability) and subordinates' reactions (e.g., morale, confidence in leadership, re-enlistment rates) have yet to be incorporated. We caution to add that using some of these indices, such as unit performance, may impose limits on the research designs that can be employed and the applicable generalizations, but they better approximate ultimate criteria and are of great interest to line units.

Army HR Practice & Leadership Research

In times of diminishing budgets and demands to do more with less, it is important to leverage leadership research with ongoing human resource (HR) programs in the Army. This alignment should highlight two factors. First, it is widely accepted that different leader attributes are important at different career stages and hierarchical levels. ARI research that samples across these stages can inform practice as to what specific features are most critical at which times. In terms of the research implications of this approach, it also suggests that some variables are rendered moot for some purposes. For example, Zaccaro's (1996) summary of SST theory suggest that acute cognitive abilities skills are presumed to be possessed by all high ranking officers such that what differentiates effective and ineffective executive leadership is attributable to other factors as such as proclivity. Note that this would suggest that indexing leaders' attributes such as cognitive capacity would be important if one was interesting in predicting who would rise to senior officer levels, but would be far less informative if one were interested in predicting effectiveness among executive officers. Therefore, there is a natural synergy between what the focus of certain research investigations should be given their purpose, and how they can inform practice in terms of providing developmental focus, critical feedback dimensions, and so forth.

The second theme linking ARI leadership research and practice involves the imbeddedness of investigations. Many of the efforts we reviewed had clear linkages with ongoing Army activities (e.g., the CPR, AZIMUTH, Special Forces & Biodata). Embedding research investigations in ongoing activities always necessitates some compromises due to administrative demands and constraints, and multiple data purposes. However, it also enhances the relevance of the research both to the line units and to the participants. We see numerous benefits from making ongoing research investigations relevant to the units providing the data. Whether it be ongoing leader development, training programs, or field exercises, to the extent

that data collected are seen as valuable to the officers and solders involved, the ease with which it is collected and the quality of the resulting indices will be enhanced. Having said the above, we realize that many more basic research investigations simply cannot be woven in to the fabric of ongoing activities, at least not in their developmental phases. We submit, however, that gaining access for these more basic and developmental activities will be easier in the context of ongoing efforts that are valued by the line and training units. Such a demarcation of efforts would also clarify the value of different studies for the Army units.

In summary, this report has chronicled a great deal of ARI leadership assessment work from the past 10 years. Much has been developed and learned. We suggest, however, that ARI is at a critical juncture and should pause to consider its strategic directions for future leadership research. In one sense, we advocate a more limited focus and integrated "back to the basics" emphasis. On the other hand, we encourage an expansion to consider a wider array of variables such as situational and follower attributes that moderate the effectiveness of leader behaviors in different circumstances. We also recommended greater embedding on research activities in ongoing Army activities and a cross-fertilization between research and practice.

# Section 7: References

**NOTE. * References represent sources of featured measures.**


Arlin, P. K. (1990). Wisdom: The art of problem finding. In R. J. Sternberg (Ed.), Wisdom: Its nature, origins, and development (pp. 230-243). New York: Cambridge University Press.

Atwater, L., Avolio, B. J., & Bass, B. M. (1991). A retrospective/prospective view of leadership development, emergence, and performance. Alexandria, VA: U. S. Army Research Institiute for the Behavioral and Social Sciences.

Atwater, L. E., Dionne, S. D., Avolio, B. J., Camobreco, J. F., & Lau, A. W. (1996). Leader attributes and behaviors predicting emergence of leader effectiveness. (Tech. Report 1044). Alexandria, VA: U. S. Army Research Institute for the Behavioral and Social Sciences.

Atwater, L., Lau, A., Bass, B., Avolio, B., Camobreco, J., & Whitmore, N. (1994). The content, construct, and criterion-related validity of leader behavior measures. (Research Note 95-01). Alexandria, VA: U. S. Army Research Institute for the Behavioral and Social Sciences.

Avolio, B. J., Dionne, S. D., Atwater, L. E., Lau, A., Camobreco, J., Whitmor, N., & Bass, B. (1996). Antecedent predictors of a "full range" of leadership and management styles. (Tech. Report 1040). Alexandria, VA: U. S. Army Research Institute for the Behavioral and Social Sciences.

Avolio, B. J., & Howell, J. M. (1992). The impact of leader behavior and leader-personality match on satisfaction and unit performance. In K. E. Clark, M. B. Clark, & R. Campbell (Eds.) Impact of leadership. Greensboro, NC: The Center for Creative Leadership.

Baer, J. M. (1988). Long-term effects of creativity training with middle school students. Journal of Early Adolescence, 8, 183-193.

Barber, H. F. (1990). Some personality characteristics of senior military officers. In K. E. Clark & M. B. Clark (Eds.), Measures of leadership (pp. 441-454). Greensboro, NC: Center for Creative Leadership.

Barrata, J. E., & McManus, M. A. (1991). Development of predictors for AIM (Assessment Inventory for Management) (Tech. Rep. MRR2-1991). Hartford, CT: Life Insurance Marketing and Research Association.

Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: A meta-analysis. Personnel Psychology, 44, 1-25.

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

* Bass, B. M. (1996). A new paradigm of leadership: An inquiry into transformational leadership. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Bass, B. M. (1996, proposal). Platoon readiness as a function of transformational/transactional leadership, squad mores and platoon cultures. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Bollen, K. A. (1989). Structural equations with latent variables. New York: Wiley.

Bond, M. H., Nakazato, H., & Shiraishi, D. (1975). Universality and distinctiveness in dimensions of Japanese person perception. Journal of Cross Culture Psychology, 6, 346-357.

Borgatta, E. F. (1964). The structure of personality characteristics. Behavioral Science, 123, 8-17.

Boyatzis, R. R. (1982). The competent manager: A model for effective performance. New York: Wiley.

Briggs, S. R. (1989). The optimal level of measurement for personality constructs. In D. M. Buss & N. Cantor (Eds.). Personality psychology: Recent trends and emerging directions. New York: Springer-Verlag.

* Briggs-Myers, I., & McCauley M. H. (1985). Manual: A guide to the development of the Myers Briggs Type Indicator. Palo Alto, CA: Consulting Psychologists Press.

Bruner, J. (1966). Toward a theory of instruction. New York: Norton.

Bycio, P., Hackett, R. D., & Allen, J. S. (1995). Further assessments of Bass's (1985) conceptualization of transactional and transformational leadership. Journal of Applied Personnel Psychology, 46, 823-850.

* Campbell, D. (1991). Manual for the Campbell Leadership Index (CLI). Minneapolis, MN: National Computer Systems, Inc.

Campbell, J. P., Dunnette, M. D., Lawler, E. E., & Weick, K. E. (1970). Managerial behavior, performance and effectiveness. New York: McGraw-Hill.

Cannon-Bowers, J. A., Salas, E., & Converse, S. A. (1993). Shared mental models in expert team decision making. In N. J. Castellan, Jr. (Ed.), Current issues in individual and group decision making. Hillsdale, N.J: Lawrence Erlbaum.

Cannon-Bowers, J. A., Salas, E., & Converse, S. A. (1990). Cognitive psychology and team training: Training shared mental models of complex systems. Human Factors, 33(12), 1-4.

Cantor, N., & Kilhstrom, J. F. (1987). Personality and social intelligence. Englewood Cliffs, NJ: Prentice Hall.

Carlsson, W. B., & Gorman, M. E. (1992). A cognitive framework to understand technological creativity: Bell, Edison, and the telegraph. In R. J. Weber & D. N. Perkins (Eds.), Inventive minds: Creativity in technology. New York: Oxford University Press.

Carlyn, M. (1977). An assessment of the Myers-Briggs Type Indicator. Journal of Personality Assessment, 41, 461-473.

Cattell, R. B. (1946). The description and measurement of personality. Yonkers, NY: World Book.

* Center for Creative Leadership (1996) Benchmarks assessing leadership skills and enhancing the development process. Greensboro, NC: Center for Creative Leadership.

Chi, M. T. H., Glaser, R., & Rees, E. (1982). Expertise in problem solving. In R. Steinberg (Ed.), Advances in the psychology of human intelligence (Vol. 1) (pp. 7-75). Hillsdale, NJ: Lawrence Erlbaum Associates.

Connelly, M. S., Marks, M. A., & Mumford, M. D. (1993). An integrated dimensional structure of wisdom. Paper presented at the annual meeting of the Eastern Psychological Association, Washington, D. C.

Connelly, M. S., Zaccaro, S. J., & Mumford, M. D. (1992, August). Evidence for trait-based cross-situational leadership. Paper presented at the meeting of the American Psychological Association, Washington, D.C.

Converse, S. A., Cannon-Bowers, J. A., & Salas, E. (1991). Mental models: Theory and some methodological issues. Proceedings of the Human Factors Society 35th Annual Meeting (pp. 1417-1421).

Converse, S. A., & Kahler, S. E. (1992, August). Shared mental models, team performance, and knowledge Acquisition. Paper presented at the annual meeting of the American Psychological Association, Washington, D.C.

Cook, J. D., Hepworth, S. J., Wall, T. D., & Warr, P. B. (1981). The experience of work. London: Academic Press.

Cooke, N. J. (1992). Predicting judgement time from measures of psychological proximity. Journal of Experimental Psychology: Learning, Memory, and Cognition, 18(3), 640 - 653.

Cooke, N. J., Durso, F. T., & Schvaneveldt, R. W. (1986). Recall and measures of memory organization. Journal of Experimental Psychology: Learning, Memory, and Cognition, 12(4), 538 - 549.

* Costa, P. T., Jr.,& McCrae, R. R. (1992). Revised NEO Personality Inventory (NEO PI-R) and NEO Five Factor Inventory (NEO-FFI) professional manual (pp. 1-65). Florida: Psychological Assessment Resources.

Costa, P. T., Jr., & McCrae, R. R. (1990). Personality disorders and the five-factor model of personality. Journal of Personality Disorders, 4, 362-371.

* Costa, P. T., Jr., & McCrae, R. R. (1988). From catalog to classification: Murray's needs and the five factor model. Journal of Personality & Social Psychology, 50, 116-123.

* Costa, P. T., Jr., & McCrae, R. R. (1987). The NEO Personality Inventory manual. Odessa, FL: Psychological Assessment Resources.

* Costa, P. T., Jr., & McCrae, R. R. (1985). The NEO Personality Inventory manual (pp. 1-65). Odessa, Florida: Psychological Assessment Resources.

Costa, P. T., Jr., McCrae, R. R., & Dye, D. A. (1991). Facet scales for Agreeableness and Conscientiousness: A revision of the NEO Personality Inventory. Personality and Individual Differences, 12, 887-898.

Costa, P. T., Jr., McCrae, R. R., & Holland, J. L. (1984). Personality and vocational interests in an adult sample. Journal of Applied Psychology, 69, 390-400.

Digman, J. M. (1990). Personality structure: Emergence of the five factor model. Annual Review of Psychology, 41, 417-440.

Dillion, J. T. (1982). Problem finding and solving. Journal of Creative Behavior, 16, 97-111.

Dweck, C. S. (1986). Motivational processes affecting learning. American Psychologist, 41, 1040-1048.

Eberhardt, B. J., & Muchinsky, P. M. (1982). An empirical investigation of the factor stability of Owens' biographical questionnaire. Journal of Applied Psychology, 67, 138-145.

Evans, M. G. (1970). The effects of supervisory behavior in the path-goal relationship. Organizational Behavior and Human Performance, 5, 277-298.

Eysenck, H. J., & Esyenck, B. G. (1964). Manual for the Eysenck Personality Inventory. London: University Press.

Fallesen, J. J., Michel, R. R., Lussie, J. W., & Pounds, J. (1996). Practical thinking: Innovation in battle command Instruction (ARI Tech. Rep. 1037). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Farr, J. L. (1982). A five factor system of leadership: A follower model of leadership. Greensboro, NC: Farr Associates.

Fiedler, F. E. (1978). The contingency model and the dynamics of the leadership process. In L. Berkowitz (Ed.), Advances in Experimental Social Psychology. New York: Academic Press.

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

Fiedler, F. E., & Garcia, J. E. (1987). New approaches to effective leadership: Cognitive resources and organizational performance. New York: Wiley.

Fleishman, E. A. (1975). Toward a taxonomy of human performance. American Psychologist, 30, 1127-1149.

Fleishman, E. A., & Quaintance, M. K. (1984). Taxonomies of human performance: The description of human tasks. Orlando, FL: Academic Press.

Fleishman, E. A., Mumford, M. D., Zaccaro, S. J., Levin, K. Y., Korotkin, A. L., & Hein, M. B. (1991). Taxonomic efforts in the description of leader behavior: A synthesis and functional interpretation. The Leadership Quarterly, 2(4), 245-287.

Fredericksen, N., & Evans, F. R. (1974). Effects of models of creative performance on ability to formulate hypotheses. Journal of Educational Psychology, 66, 67-82.

French, J., & Raven, B. H. (1959). The bases of social power. In D. Cartwright (Ed.), Studies in social power. Ann Arbor, MI: Institute for Social Research.

Gardner, W. L., & Martinko M. J. (1996). Using the Myers Briggs Type Indicator to study managers: Literature review and research agenda. Journal of Management, 22(1), 45-83.

Geiwitz, J. (1993). A conceptual model of metacognitive skills (ARI Tech. Rep. 51-1). Alexandria, VA: Army Research Institute for the Behavioral and Social Sciences.

Getzels, J. W., & Csikszentmihalyi, M. (1976). The creative vision: A longitudinal study of problem finding in art. New York: Wiley.

Goldberg, L. R. (1992). The development of markers for the big five factor structure. Psychological Assessment, 4, 26-42.

Goldsmith, T. E., & Johnson, P. J. (1990). A structural assessment of classroom learning. In R. W. Schvaneveldt (Ed.), Pathfinder associative networks studies in knowledge organization. Norwood, N.J: Ablex Publishing Corporation.

* Gough, H. G. (1988). California psychological administer's guide. Palo Alto, CA: Consulting Psychologists Press.

* Gough, H. G. (1957). Manual for the California Psychological Inventory. Palo Alto, CA: Consulting Psychologists Press.

Gough, H. G. (1952). On making a good impression. Journal of Educational Research, 46, 33-42.

Gough, H. G., & Heilbrun, A. B., Jr. (1983). Adjective Check List manual. Palo Alto, CA: Consulting Psychologists Press.

Gough, H. G., & Lanning, K. (1986). Predicting grades in college from the California Psychological Inventory. Educational and Psychological Measurement, 46, 205-213.

Guilford, J. P. (1967). The nature of human intelligence. New York: McGraw Hill.

* Guilford, J. P., & Guilford, J. S. (1980). Consequences: Manual of instructions and operations. Orange, CA: Sheridan Psychological Services.

Guilford, J. P., & Hoepfer, R. (1971). The analysis of intelligence. New York: McGraw-Hill.

Hakel, M. D. (1974). Normative personality factors recovered from ratings of personality descriptors: The beholder's eye. Personnel Psychology, 27, 409-421.

Harris, M. M., & Schaubroeck, J. (1988). A meta-analysis of self-supervisor, self-peer, and peer-supervisor ratings. Personnel Psychology, 41, 43-62.

Hathaway, S. R., & McKinley, J. C. (1967). Manual for the Minnesota Multiphasic Personality Inventory. New York: Psychological Corporation.

Hausser, D. L. (1980). Comparison of different models for organizational analysis. In E. E. Lawler, D. A. Nadler, & C. Cammann (Eds.), Organizational assessment: Perspectives on the measurement of organizational behavior and the quality of work life (pp. 132-161). New York: Wiley.

Hemphill, J. K., & Coons, A. E. (1957). Development of the Leader Behavior Description Questionnaire. In R. M. Stogdill & A. E. Coons (Eds.), Leader behavior: Its description and measurement (pp. 6-38). Columbus, OH: Ohio State University.

Hinsz, V. (1995). Mental models of groups as social systems: Considerations of specification and assessment. Small Group Research, 26, 200-233.

Hinton, B. L. (1968). Environmental frustration and creative problem solving. Journal of Applied Psychology, 52, 211-217.

Hogan, R. (1990). Personality and personality measurement. In M. D. Dunnette & L. Hough (Eds.), Handbook of industrial and organizational psychology (vol. 2). Palo Alto, CA: Consulting Psychologists Press.

* Hogan, R. (1986). The Hogan Personality Inventory manual. Minneapolis, MN: National Computer Services.

Hogan, R. (1971). Personality characteristics of highly rated policemen. Personnel Psychology, 24, 679 - 686.

Hogan, R., Hogan, J., & Busch, C. (1984). How to measure service orientation. Journal of Applied Psychology, 69, 157-163.

Holland, J. L. (1985). The SDS Professional Manual- 1985 Revision. Odessa, FL: Psychological Assessment Resources.

* Holt, K. E., & Hazucha, J. F. (1991). The PROFILIER Technical Report. Minneapolis, Minnesota: Personnel Decisions.

Holyoak, K. (1984). Mental models in problem solving. In J. R. Anderson & S. M. Kosslyn (Eds.), Tutorials in learning and memory: Essays in honor of Gordon Bower. New York: Freeman.

Hooijberg, R., & Quinn, R. E. (1992). Behavioral complexity and the development of effective managers. In R. L. Philips & J. G. Hunt (Eds.), Strategic leadership: A multiorganizational-level perspective (pp. 16-25). London: Quorum Books.

Hoover, S. M., & Feldhusen, J. F. (1990). The scientific hypothesis formulation ability of gifted ninth-grade students. Journal of Educational Psychology, 82, 838-848.

* Horvath, J. A., Forsythe, G. B., Sweeney, P. J., McNally, J. A., Wattendorf, J., Williams, W. M., & Sternberg, R. J. (1994). Tacit knowledge in military leadership: Evidence from officer interviews (ARI Tech. Rep. 1018). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Horvath, J. A., Sternberg, R. J., Forsythe, G. B., Sweeney, P. J., Bullis, R. C., Williams, W. M., & Dennis, M. (1996). Tacit knowledge in military leadership: Supporting instrument development (ARI Tech. Rep. 1042). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Horvath, J. A., Williams, W. M., Forsythe, G. B., Sweeney, P. J., Sternberg, R. J., McNally, J. A., & Wattendorf, J. (1994). Tacit knowledge in military leadership: A review of the literature (Tech. Rep. 1017). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

House, R. J. (1971). A path-goal theory of leader effectiveness. Administrative Science Quarterly, 16, 321-339.

Jackson, D. N. (1976). Jackson Personality Inventory: Inventory manual. Port Huron, MI: Research Psychologists Press.

Jackson, D. N. (1984). Personality Research Form manual (3rd ed.). Port Huron, MI: Research Psychologists Press.

Jacobs, T. O., & Jaques, E. (1991). Executive Leadership. In R. Gal & A. D. Manglesdoff (Eds.), Handbook of military psychology. Chichester: England University.

Jacobs, T. O. & Jaques, E. (1990). Military executive leadership. In K. E. Clark & M. B. Clark (Eds.), Measures of leadership. Greensboro, NC: Center for Creative Leadership.

* Jacobs, T. O., & Jaques, E. (1989). Executive leadership (ARI Tech. Rep.). Alexandria, VA: Army Research Institute.

Jacobs, T. O., & Jaques, E. (1987). Leadership in complex systems. In J. Zeidner (Ed.), Human productivity enhancement organizations, personnel and decision making (pp.7-65). New York: Praeger.

Jacobs, T., O., & Lewis, P. (1992). Leadership requirements in a stratified system. In R. L. Philips & J. G. Hunt (Eds.), Strategic leadership: A multiorganizational-level perspective (pp. 16-25). London: Quorum Books.

Jaques, E. (1985). Development of intellectual capability. In F. R. Link (Ed.), Essays on the intellect (pp. 107-141). Alexandria, VA: Association for Supervision and Curriculum Development.

Jaques, E. (1975). A general theory of bureaucracy. London: Heinemann Educational Books.

Jaques, E., & Clement, S. D. (1991). Executive leadership : A practical guide to managing complexity. Arlington, VA: Cason Hall.

John, O. P. (1989). Towards a taxonomy of personality descriptors. In D. M. Buss & N. Cantor (Eds.), Personality psychology: Recent trends and emerging directions. New York: Springer-Verlag.

Johnson-Laird, P. (1983). Mental Models. Cambridge, MA: Harvard University Press.

Jung, C. G. (1971). Psychological types (H. G. Baynes Trans., revised by R. F. C. Hull). Princton, NJ: Princeton University Press.

Kaplan, R. E. (1986). The warp and woof of the general manager's job. In B. Schneider & D. Schoorman (Eds.), Facilitating work effectiveness. Lexington, MA: Lexington Books.

Katkovsky, W. (1992). [Review of the Leatherman Leadership Questionnaire]. In J. J. Kramer & J. C. Connoley (Eds), The eleventh mental measurement yearbook (pp. 465 - 466). Lincoln, NE: Buros Institute of Mental Measurements.

Katz, D., & Kahn, R. L. (1978). The social psychology of organizations (2nd Ed.). New York: Wiley.

* Keene, S. D. , Halpin, S. M., & Spiegel, D. K. (1996). The Leader Azimuth Check: A leader self development tool. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

* Kegan, R. (1982). The evolving self: Problems and process in human development. Cambridge, MA: Harvard University Press.

Kerr, S., & Jermier, J. M. (1978). Substitutes for leadership: Their meaning and measurement. Organizational Behavior and Human Performance, 22, 375-403.

* Kilcullen, R. N., White, L. A., Mumford, M. D., & O'Connor, J. A., (1995). Predicting the effectiveness of public sector supervisors using biodata. Paper presented at the annual meeting of the Society of Industrial and Organizational Psychology, Orlando, FL.

Kleinman, D. L., & Serfaty, D. (1989). Team performance assessment in distributed decision making. In R. Gilson, J. P. Kincaid, & B. Goldiez (Eds.), Proceedings: Interactive networked simulation for Training Conference (pp. 22-27). Orlando, FL: Institute for Simulation and Training.

Knowlton, W., & McGee, M. (1994). Strategic leadership and personality making the MBTI relevant. Washington, DC: National Defense University Industrial College of the Armed Forces.

Kouzes, J. M., & Posner, B. Z. (1987). The leadership challenge. How to get extraordinary things done in organizations. San Francisco, CA: Jossey Bass.

Kraiger, K. (1993). Further evidence of structural assessment as a valid measure for training evaluation. Paper presented at the annual meeting of the American Psychological Asociation, CA.

Kuhn, T. S. (1970). The structure of scientific revolutions. Chicago, IL: University of Chicago Press.

Laskey, K. B. , Leddo, J. M., & Bresnick, T. A. (1990). Executive thinking and decision skills: A characterization and implications for training. (Research Note 91-07). Alexandria, VA: U. S. Army Research Institute for Behavioral and Social Sciences.

Latham, G. P., Saari, L. M., Pursell, E. D., & Campion, M. A. (1980). The situational interview. Journal of Applied Psychology, 65, 442-431.

Lautenschlager, G. J., & Shaffer, G. S. (1987). Reexamining the component stability of Owens's biographical questionnaire. Journal of Applied Psychology, 72, 149-152.

Lewis, P. M. (1995). Conceptual capacity and officer effectiveness. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Lewis, P. M. (1993). Career Path Appreciation (CPA) data reduction and analysis (ARI Tech. Rep. 983). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Lewis. P, & Jacobs, T. O. (1992). Individual differences in strategic leadership capacity: A constructive/developmental view. In R. L. Philips & J. G. Hunt (Eds.), Strategic leadership: A multiorganizational-level perspective. London: Quorum Books.

Livneh, H., & Livneh, C. (1989). The five factor model of personality: Is evidence for its cross media premature? Personality and Individual Differences, 10, 75 - 80.

London, M., & Beatty, R. W. (1993). 360-degree feedback as a competitive advantage. Human Resource Management, 32, 353-372.

Mael, F. A., & Schwartz, A. C. (1991). Capturing temperament constructs with objective biodata (ARI Tech. Rep. 939). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Markessini, J. (1991). Executive leadership in a changing world order: Requisite cognitive skills (ARI Res. Rep.). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Mathieu, J. E., & Day, D. V. (1997). Assessing processes within and between organizational teams: A nuclear power plant example. In M. T. Brannick, E. Salas, & C. Prince (Eds.), Team performance assessment and measurement: Theory, methods, and applications. Mahwah, NJ: Lawrence Erlbaum.

McCall, M. W., Spreitzer, G. M., & Mahoney, J. (1996). Prospector's user's guide. Greensboro NC: Center for Creative Leadership.

McCarley, N. G. & Carskadon, T. G. (1983). Test-retest reliabilities of scales and subscales of the Myers-Briggs Type Indicator and of criteria for clinical interpretive hypotheses involving them. Research in Psychological Type, 6, 24-36.

McCrae, R. R., & Costa, P. T., Jr. (1989). Reinterpreting the Myers-Briggs Type Indicator from the perspective of the five factor model of personality. Journal of Personality, 57, 17-40.

McCrae, R. R., & Costa, P. T., Jr. (1987). Validation of the five factor model of personality across instruments and observers. Journal of Personality and Social Psychology, 52, 81-90.

McCrae, R. R., & Costa, P. T., Jr. (1985). Updating Norman's "adequate taxonomy": Intelligence and personality dimensions in natural language and in questionnaires. Journal of Personality and Social Psychology, 49, 710 - 721.

McCrae, R. R., Costa, P. T., Jr., & Busch, C. M. (1986). Evaluating comprehensiveness in personality systems. The California Q-set and the five factor model. Journal of Personality, 54, 430-466.

McGrath, J. E. (1976). Small group research. Austin, TX: Holt, Rinehart, &Winston.

McIntyre, R. M., Jordan, P., Mergen, C., Hamill, L., & Jacobs, T. O. (1993). The construct validity of the CPA: Report on three investigations (ARI Rep. No. TCN 92-017). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Miner, J. B. (1965). Studies in management education. New York: Springer Publishing.

Mintzberg, H.(1973). The nature of managerial work. New York: Harper & Row.

Mobley, M. I., Doares, L. M., & Mumford, M. D. (1992). Process analytic models of creative capacities: Evidence for the combination and reorganization process. Creativity Research Journal, 5, 125-155.

Modjeski, R. B., & Michael, W,. B. (1983). An evaluation by a panel of psychologists of the reliability and validity of two tests of critical thinking. Educational and Psychological Measurement, 43, 1187 - 1197.

Mohammed, S. (1995). Pathfinder as a mental model measurement technique. Presented at a symposium at the Academy of Management: Vancouver, British Columbia, Canada.

* Motowidlo, S. J., Dunnette, M. D., & Carter, G. W. (1990). An alternative selection procedure the low fidelity simulation. Journal of Applied Psychology, 75,(6), 640-647.

Motowidlo, S. J., & Tippins, N. (1993). Further studies of the low fidelity simulation in the form of a situational inventory. Journal of Occupational and Organizational Psychology, 66, 337-344.

Mount, M. K., & Barrick, M. R. (1995). The big five personality dimensions: Implications for research and practice in human resources management. Personnel and Human Resources Management, 13, 153-200.

* Mumford, M. D., Baughman, W. A, Supinski, E. P., Costanza, D. P., & Threlfall, K. V. (1993). Cognitive and metacognitive skill development: Alternative measures for predicting leadership potential. U. S. Army Research Institute for Behavioral and Social Sciences and Management Research Institute Inc., Maryland.

Mumford, M. D., O' Connor, J., Clifton, T. C., Connelly, M. S., & Zaccaro, S. J. (1993). Background data constructs as predictors of leadership behavior. Human Performance, 6, 151-195.

Mumford, M. S., & Stokes, G. S. (1992). Developmental determinants of individual action: Theory and practice in applying background data measures. In M. D.

Dunnette & L. M. Hough (Eds.), Handbook of industrial and organizational Ppsychology: Vol 3 (pp. 61-138). Palo Alto, CA: Consulting Psychologists Press.

Mumford, M. D., Yoarkin-Levin, K., Korotkin, A. L., Wallis, M. R., & Marshall-Mies, J. (1986). Characteristics relevant to performance as an Army leader: Knowledge, skills, aptitudes, other characteristics and generic skills (ARI Res. Note 86-24). Bethesda, MD: Advanced Research Resources Organization.

* Mumford, M. D., Zaccaro, S. J., Harding, F. D., Fleishman, E. A., & Reiter-Palmon, R. (1991). Cognitive and temperament predictors of executive ability: Principles for developing leadership capacity. Alexandria, VA: U.S. Army Research Institute for Behavioral and Social Sciences.

* Myers, I. B., & McCaulley, M. H. (1985). Manual: A guide to the development and use of the Myers-Briggs Type Indicator. Palo Alto, CA: Consulting Psychologists Press.

Myers, I. B., & Myers, P. B. (1980). Gifts differing. Palo Alto, CA: Consulting Psychologists Press.

Nickerson, R. S. (1990). Technology in education: Possible influences on context, purpose, content, and methods. In R. S. Nickerson and P. P. Zodhiates (Eds.), Technology and Education in 2020. Hillsdale, NJ: Lawrence Erlbaum.

Novak, J. D. (1990). Concept maps and Vee diagrams: Two metacognitive tools to facilitate meaningful learning. Instructional Science, 19, 29-52.

Nunally, J. C. (1978). Psychometric Theory (2nd ed.). New York: McGraw-Hill.

Nunally, J. C., & Bernstein, I. H. (1994). Psychometric theory (3rd ed.) New York: McGraw-Hill.

Oldham, G. R. (1976). The motivational strategies used by supervisors: Relationships to effectiveness indicators. Organizational Behavior and Human Performance, 15, 66-86.

Office of Institutional Research (1989). Evaluation of the sequential leader development system: Content validity. West Point, NY: U. S. Military Academy.

Okuda, S. M., Runco, M. A., & Berger, D. E. (1991). Creativity and the finding and solving of real-world problems. Journal of Psychoeducational Assessment, 9, 45-53.

Owens, W. A. (1976). Background data. In M. D. Dunnette (Ed.), Handbook of industrial and organizational psychology. Chicago: Rand McNally.

Owens, W. A. (1971). A quasi-actuarial prospect for individual assessment. American Psychologist, 26, 992-999.

Owens, W. A. (1969). Cognitive, noncognitive, and environmental correlates of mechanical ingenuity. Journal of Applied Psychology, 53, 199-208.

Owens, W. A. (1968). Toward one discipline of scientific psychology. American Psychologist, 23, 782-785.

* Owens, W. A., & Schoenfeldt, L. F. (1979). Toward a classification of persons. Journal of Applied Psychology Monograph, 65, 569-607.

Pfeffer, J., & Salancik, G. R. (1978). The external control of organizations: A resource dependent perspective. New York: Harper & Row.

Podsakoff, P. M. & Organ, D. W. (1986). Self-reports in Organizational Research: Problems and Prospects. Journal of Management, 12, 531-544.

* Posner, B. Z., & Kouzes, J. M. (1996). Psychometric properties of the Leader Practices Inventory - Updated. Educational and Psychological Measurement, 53(1), 191-199.

Posner, B. Z., & Kouzes, J. M. (1990). Leadership Practices: An Alternate to the Psychological Perspective. In K. E. Clark & M. B. Clark (Eds.), Measures of leadership (pp. 205-215). West Orange, NJ: Leadership Library of America.

Redmond, M. R., Mumford, M. D., & Teach, R. (1993). Putting creativity to work: Effects of leader behavior on subordinate creativity. Organizational Behavior and Human Decision Processes, 55, 120-151.

Repapi, M., Gough, H. G., Lanning, K., & Stefanis, C. (1983). Predicting academic achievement among Greek secondary school students from family background and California Psychological Inventory scores. Contemporary Educational Psychology, 8, 181-188.

Rouse, W. B., & Morris, N. M. (1986). On looking into the black box. Prospects and limits in the search for mental models. Psychological Bulletin, 100, 349-363.

Schriesheim, C. A., Kinicki, A. J., & Schriesheim, J. F. (1979). The effect of leniency on Leader Behavior Descriptions. Organizational Behavior and Human Performance, 23, 1-29.

Schriesheim, C. A., & Stogdill, R. M. (1975). Differences in factor structure across three versions of the Ohio State leadership scales. Personnel Psychology, 28, 189-206.

Schvaneveldt, R. W. (1990). Proximities, networks, and schemata. In R. W. Schvaneveldt (Ed.), Pathfinder associative networks studies in knowledge organization, Norwood, NJ: Ablex Publishing Corporation.

Schvaneveldt, R. W., Durso, F. T., & Dearholt, D. W. (1989). Network structures in proximity data. The Psychology of Learning and Motivation, 24, 249-284.

Schvaneveldt, R. W., Durso, F. T., Goldsmith, T. E., Breen, T. J., Cooke, N. M., Tucker, R. G., & Maio, J. C. (1985). Measuring the structure of expertise. International Journal of Man-Machine Studies, 23, 699-728.

Schwager, E. H., & Evans, K. L. (1996). An exploration of the construct validity of a Leadership Behavior Rating System (ARI Tech. Rep. 1041). Alexandria, VA: U. S. Army Research Institute for the Behavioral and Social Sciences.

Shaffer, G. S., Saunders, V., & Owens, W. A. (1986). Additional evidence for the accuracy of biographical data: Long-term retest and observer ratings. Personnel Psychology, 39, 791-809.

Shiflett, S., Eisner, E., & Inn, A. (1981). The identification and measurement of military team functions using videotaped segments of team activities (ARRO Final Report 3068). Washington, DC: Advanced Research Resources Organization.

Shiflett, S., Price, S., & Schemmer, F. M. (1981). Toward a taxonomy of team functions concepts: Operations and issues (ARRO Interim Report 3068). Washington DC: Advanced Research Resources Organizations.

Shorris, E. (1981). The oppressed middle: Politics of middle-management/scenes from corporate life. New York: Anchor Press/Doubleday.

Siegler, R. S., & Richards, J. D. (1982). The development of intelligence. In R. J. Sternberg (Ed.), Handbook of human intelligence. New York: Cambridge University Press.

Sipps, G. J., & Alexander, R. A. (1987). The multifactorial nature of extraversion-introversion in the Myers-Briggs Type Indicator and Eysenck personality inventory. Educational and Psychological Measurement, 47, 543-552.

Sipps, G. J., Alexander, R. A., & Friedt, L. (1985). Item Analysis of the Myers-Briggs Type Indicator. Educational and Psychological Measurement, 45, 789 - 796.

Smilansky, J. (1984). Problem solving and the quality of invention: An empirical investigation. Journal of Educational Psychology, 76, 377-386.

Smith, G. M. (1967). Usefulness of peer ratings of personality in eduactional reseach. Educational and Psychological Measurement, 27, 967-984.

Smith, J., & Baltes, P. B. (1990). Wisdom related knowledge: Age/cohort differences in response to life-planning problems. Developmental Psychology, 26, 494-504.

Snow, R. E. (1993). Construct validity and constructed response tests. In R. E. Bennett & W. C. Ward (Eds.), Construction versus choice in cognitive measurement. (pp. 45-60). Hillsdale, NJ: Lawrence Erlbaum.

Snow, R. E., & Lohman, D. R. (1984). Toward a theory of cognitive aptitude for learning from instruction. Journal of Educational Psychology, 76, 347-375.

Spielberger, C. C., Jacobs, G., Crane, R., Russell, S., Westberry, L., Barker, L., Johnson, E., Knight, J., & Marks, E. (1979). Preliminary manual for the State-Trait Personality Inventory (STPI). Tampa: University of South Florida Human Resources Institute.

Stamp, G. P. (1988). Longitudinal research into methods of assessing managerial potential (ARI Tech. Rep. 819). Alexandria, VA: U. S. Army Research Institute for the Behavioral and Social Sciences.

* Stamp, G. P. (1986). Career Path Appreciation and Career Path Mapping: Measurement of capacity unit publication. Alexandria, VA: U. S. Army Research Institute for the Behavioral and Social Sciences.

Sternberg, R. J. (1990). Wisdom and its relations to intelligence and creativity. In R. J. Sternberg (Ed.). Wisdom: Its nature, origin and development (pp. 230-243). New York: Cambridge University Press.

Sternberg, R. J. (1985). Implicit theories of intelligence, creativity, and wisdom. Journal of Personality and Social Psychology, 49, 606-627.

Sternberg, R. J. (1985). Beyond IQ. New York: Cambridge University Press.

Sternberg, R. J. (1988). A three-faceted model of creativity. In R. J. Sternberg (Ed.), The Nature of Creativity: Contemporary Psychological Perspectives (pp. 125-147). Cambridge, MA: Cambridge University Press.

Sternberg, R. J., & Wagner, R. K. (1993). The g-ocentric view of intelligence and job performance is wrong. Current Directions in Psychological Sciences, 2, 1-4.

Sternberg, R. J., Wagner, R. K., & Okagaki, L. (1993). Practical intelligence: The nature and role of tacit knowledge in work and school. In H. W. Resse & J. M. Puckett (Eds.), Mechanisms of everyday cognition. Hillsdale, NJ: Lawrence Erlbaum.

Stewart, S. R, Kilcullen, R., & Hopkins, J. E. (1994). Strategic Leader Development Inventory (SLDI). Alexandria, VA: U.S. Army Research Institute for Behavioral and Social Sciences.

Stodgill, R. M. (1974). Handbook of leadership: A survey of the literature. New York: Free Press.

* Stodgill, R. M. (1963). Manual for the Leader Behavior Description Questionnaire - Form XII. Columbus: Ohio State University, Bureau of Business Research.

Stodgill, R. M., Goode, O. S., & Day, D. R. (1965). The leader behavior of university presidents. Columbus, OH: Ohio State University, Bureau of Business Research.

* Stout, R. J., Salas, E., & Kraiger, K. (1997). The role of trainee knowledge structures in aviation team environments. The International Journal of Aviation Psychology, 7, 235-250.

Super, D. E. (1962). The structure of work values in relation to status, achievement, interests and adjustment. Journal of Applied Psychology, 46, 231-239.

* Ternman, L. M., & Oden, M. (1959). The gifted group at mid-life. Stanford, CA: Stanford University Press.

Van Fleet, D. D., & Yukl, G. A (1986). Military leadership: An organizational behavior perspective. Greenwich, CT: JAI Press.

Velsor, E. V., & Leslie, J. B. (1991). Feedback to managers, Volume II: A review and comparison of sixteen multi-rater and feedback instruments. Greensboro, NC: Center for Creative Leadership.

* Wagner, R. K. (1987). Tacit knowledge in everyday intelligent behavior. Journal of Personality and Social Psychology, 52, 1236-1247.

Wagner, R. K., & Sternberg, R. J. (1990). Street Smarts. In K. E. Clark & M. B. Clark (Eds.), Measures of leadership (pp. 493-504). West Orange, NJ: Leadership Library of America.

Wagner, R. K., & Sternberg, R. J. (1985). Practical intelligence in real-world pursuits: The role of tacit knowledge. Journal of Personality and Social Psychology, 49, 436-458.

* Watson, G., & Glaser, E. M. (1964). Watson-Glaser Critical Thinking Appraisal manual: Form YM and ZM. New York: Harcourt Brace and World.

Wernimont, P. E., & Campbell, J. P. (1968). Signs, samples, and criteria. Journal of Applied Psychology, 52, 372-376.

Williams, W. M., & Sternberg, R. J. (in preparation). Success acts for managers. New York: Harcourt Brace and Company.

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

Zaccaro, S. J. (1996). Models and theories of executive leadership. A conceptual/empirical review and integration. Alexandria, VA: U.S. Army Research Institute for Behavioral and Social Sciences.

Zaccaro, S. J., Foti, R. J., & Kenny, D. A. (1991). Self monitoring and trait based variance in leadership: An investigation of leader flexibility across multiple group situations. Journal of Applied Psychology, 76(2), 308 - 315.

Zaccaro, S. J., Gilbert, J. A., Thor, K. K., & Mumford, M. D. (1991). Leadership and social intelligence: Linking social perceptiveness and behavioral flexibility to leader effectiveness. Leadership Quarterly, 2, 317-331.

* Zaccaro, S. J., Marks, M., O'Conner-Boes, J., & Costanza, D. (1995). The nature and assessment of leader mental models (MRI Report 95-3). Alexandria, VA: U.S. Army Research Institute for Behavioral and Social Sciences.

* Zaccaro, S. J., White, L., Kilcullen, B., Parker, C. W., Williams, D., & O'Conner-Boes, J. (1997). Cognitive and temperament predictors of Army civilian leadership. (Draft). Bethesda, MD: Management Research Institute.