DECISION ANALYSIS
Vol. 10, No. 1, March 2013, pp. 82–97
ISSN 1545-8490 (print) | ISSN 1545-8504 (online)

http://dx.doi.org/10.1287/deca.1120.0263

WholeSoldier Performance Appraisal to Support Mentoring and Personnel Decisions

Robert A. Dees
McCombs School of Business, University of Texas at Austin, Austin, Texas 78712,
rob.dees@utexas.edu

Scott T. Nestler
Naval Postgraduate School, Monterey, California 93943,
scott.nestler@gmail.com

Robert Kewley
Department of Systems Engineering, United States Military Academy, West Point, New York 10996,
robert.kewley@usma.edu

We present a multiattribute model called WholeSoldier Performance that measures the performance of junior enlisted soldiers in the U.S. Army; currently there is no formal performance appraisal system in place. The application is unique to decision analysis in that we utilize a common constructed scale and single-dimensional value function for all attributes to match the natural framework of model users and based on operability concerns. Additionally, we discuss model validation in both the terms of decision analysis and psychometrics in models that are used for repeated or routine assessments and thus generate significant quantities of data. We highlight visualization of data for use to support mentoring and personnel decisions to better train, assign, retain, promote, and separate current personnel. Last, we address common cultural concerns related to performance appraisals in organizations by offering a method to standardize ratings and hold raters accountable for their responsibility to mentor subordinates as well as identify their performance to the larger organization.

Key words: value-focused thinking; performance appraisal; mentoring; personnel decisions; applications: military; practice
History: Received on October 7, 2011. Accepted by former Editor-in-Chief L. Robin Keller on December 14, 2012, after 2 revisions.

1. Introduction

Field Manual 1, The Army (Headquarters, Department of the Army 2005) codifies the vision for the U.S. Army; the opening paragraph emphasizes that "quality" soldiers are the army's most important resource. As such, the U.S. Army should take great effort to manage this resource wisely. We take managing soldiers wisely to mean making good personnel decisions regarding the recruitment, assignment, mentoring, training, retention, promotion, and separation of soldiers. To effectively pursue such decision making, the army must define and measure the quality of soldiers. Symons et al. (1982, p. 5) describe the definition of soldier quality as important, emotional, and elusive in that "quality itself is a qualitative descriptor and resists quantification in an age when quantifiable data is required for everything." Three decades later, similar conditions exist as the army faces significant budgetary and personnel cutbacks that include reducing the size of the active-duty force by 80,000 soldiers over the next five years (Mattson 2012); personnel decisions are of the utmost importance to allow the army to satisfy its mission in the decades ahead. The purpose of this paper is to outline the process that was followed to define a multiattribute model of WholeSoldier Performance, thereby providing a definition and measure of soldier quality such that leaders in the army can better mentor

soldiers and make personnel decisions while providing a framework and data for continued research. The application of the methodology is to military personnel, but there are clear parallels in academia, business, healthcare, sports, government, and other fields. In §1, we provide a brief context and background relating to measures of personnel performance in the army and business. Section 2 focuses on the model, visualization of data, and validation. Section 3 concludes and highlights directions of future work.

1.1. Army Background

Significant time and energy have been devoted to the study of soldier quality. With the inception of the All-Volunteer Force in 1974, high school diploma graduate status and the Armed Forces Qualification Test score (Rostker 2006) were congressionally mandated as the primary measures of quality. Similarly, there are dozens of psychometric measures and other tests that are proposed or utilized in the recruit population to provide information in recruiting decisions. Although these measures may provide information to understand the uncertain potential of recruits before they enter the service, they do not measure realized performance inside the organization. Realized performance has value; indicators of recruit potential are valued in recruiting decisions based only on their ability to predict future longevity or performance. Although recruiting measures are very important, our focus is on defining and measuring the performance of soldiers to support decisions regarding personnel once they are in the army.

Currently, there is no standard measure of performance utilized in the junior enlisted soldier population, who make up nearly half of all army personnel. Quarterly performance counseling is conducted, but the counseling form[1] does not include any quantifiable information and is maintained locally in a paper file. Although immediate supervisors closely interact with and understand the performance of soldiers under their authority, there is currently no mechanism for this knowledge to be aggregated and communicated to the larger organization. In general, it takes several years for a young soldier to be promoted to the rank of sergeant, when he or she would begin to receive Noncommissioned Officer Evaluation Report ratings. This leaves policy makers "nearly blind to merit" (Kane 2011). Our study was initiated by leaders at the U.S. Army Recruiting Command in 2008 to address this concern.

Other researchers have considered measures of performance for junior enlisted soldiers inside the army; most notably, Schinnar et al. (1988) employed data envelopment analysis to develop performance indices for four specific jobs in the army based on job-knowledge tests, hands-on tests, school knowledge tests, and supervisor ratings. We employ a multi-attribute decision analysis model that incorporates organizational preference to define soldier performance and collect supervisor ratings across all jobs in the army while retaining the flexibility to incorporate specific measures for specific jobs. Schinnar et al. (1988) noted that their work is exploratory and descriptive; we carry on in the same spirit within a prescriptive decision analysis framework and offer a low-cost, broadly applicable tool for regular supervisor assessment of soldier performance across all jobs in the army.

The U.S. Army does collect performance information on officers and noncommissioned officers. Currently, the Officer Evaluation Report[2] only has one meaningful "block check," in which senior raters (two levels above the rated officer) generally only categorize performance as "above center of mass" or "center of mass"; it is better than an absence of quantifiable information, but does not differentiate well. The Noncommissioned Officer Evaluation Report[3] incorporates ratings in five areas (competence, physical fitness/military bearing, leadership, training, and responsibility/accountability) with four levels each and one overall rating with three levels. Although the army arguably modeled its objectives in

[1] The Developmental Counseling Form can be found at http://armypubs.army.mil/eforms/pdf/A4856.pdf.

[2] The form can be found at http://armypubs.army.mil/eforms/pdf/A67_9.pdf. Based on the culture of the organization, Part VII.b is the only area that is truly used to differentiate performance, and generally only the top two blocks are used.

[3] The form can be found at http://armypubs.army.mil/eforms/pdf/A2166_8.pdf. Parts IV.b–f and V provide quantifiable differentiation of performance.

this report, it stops short of a value model to reflect preferences over objectives. Additionally, the rating levels are relatively unclear (excellence, success, needs some improvement, and needs much improvement) and could easily be redesigned to reduce ambiguity. Both reports are subject to factors that encourage raters to inflate their ratings leading to a measure of culture rather than performance. We provide a method to address these concerns with WholeSoldier Performance.

1.2. Related Work

In business, companies have employed a "balanced scorecard" (Kaplan and Norton 1992) approach that complements traditional financial measures and translates organizational mission, vision, and strategy into an actionable "set of objectives and measures, agreed upon by all senior executives, that describe the long-term drivers of success" (Kaplan and Norton 1996, p. 76). To align employees' individual performances with the firm's overall strategy, "the organization's high-level strategic objectives and measures must be translated into objectives and measures for operating units and individuals" through the use of a personal scorecard at the individual level (Kaplan and Norton 1996, p. 80). Furthermore, many companies have linked individual compensation to performance by "assigning weights to each objective and calculating incentive compensation by the extent to which each weighted objective was achieved" (Kaplan and Norton 1996, p. 82). Although Kaplan and Norton (1996) do not advocate aggregation of this nature, Keeney (2000) concluded that "decision analysis provides a logical foundation for, procedures to implement, and models to use a balanced scorecard approach." In this way, WholeSoldier Performance can be considered as a personal scorecard that is logically supported by a multiattribute model to communicate the organization's vision to individual soldiers, to facilitate mentoring through goal setting and performance review, and to quantifiably support a broad class of personnel decisions.

2. WholeSoldier Performance Modeling

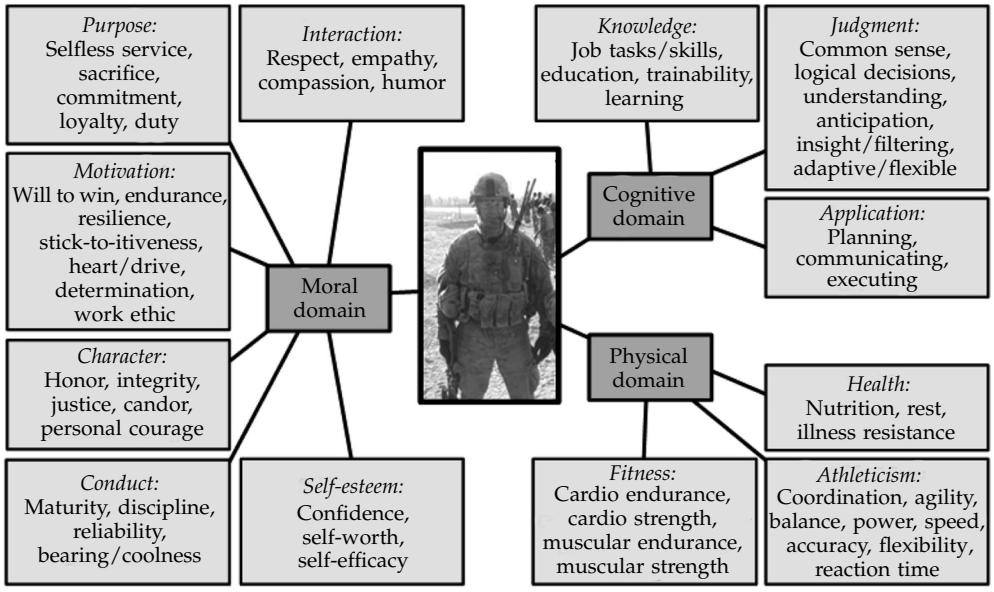
Value-focused thinking (VFT) is a leading philosophical approach to building value hierarchies in decisions with multiple attributes (Keeney 1992) and is underpinned by the mathematical methodology of multiple attribute decision analysis (Keeney and Raiffa 1976). The central idea of a VFT analysis under certainty is to define attributes and measures in a value hierarchy and then represent preferences with a quantitative value function.

2.1. Problem Structuring

As a starting point, we consulted with individuals in many relevant academic departments and centers at the United States Military Academy. In the military research community, we consulted with individuals from the Army Research Institute, RAND Corporation, and others involved in the U.S. Army Accessions Command research consortium. In particular, we found synergy with the human dimension study (Headquarters, Department of the Army 2008, p. 16; italics added for emphasis) designed as a point of departure for research into "the performance, reliability, flexibility, endurance, and adaptability of an Army made up of Soldiers" and accepted its conclusion that "the Army will require extraordinary strength in the *moral*, *cognitive*, and *physical* components of the human dimension."

To develop a value hierarchy, we spent a year interviewing hundreds of army personnel including recruiters, drill sergeants, squad leaders, platoon sergeants, platoon leaders, first sergeants, company commanders, command sergeant majors, battalion and brigade commanders, and special forces team leaders. For reference, there are approximately 10 soldiers in a squad, 30 in a platoon, 100 in a company, and 800 in a battalion. The interviews were effectively a lengthy exercise in affinity diagramming (Parnell 2007), a problem structuring technique to gather and group large amounts of language data on attributes in applications with multiple stakeholders. We asked each interviewee to first spend time generating an exhaustive list of desirable attributes in soldiers and then group them, while emphasizing the properties of completeness, nonredundancy, decomposability, operability, and small size (Keeney and Raiffa 1976, Kirkwood 1997). Operability, which Kirkwood (1997, p. 18) defined as a property of a model "that is understandable for the persons who must use it," and small size are particularly relevant to military leaders, because any performance assessment system

Figure 1 WholeSoldier Performance Attribute Groups Hierarchy



must resonate with all army personnel and not create an undue organizational burden in implementation. This exercise in affinity diagramming and hierarchy refinement led to the value hierarchy shown in Figure 1. Here, italicized headings are the attribute groups in the moral, cognitive, and physical domains; other words often grouped together during affinity diagramming are recorded under these headings to provide context.

The WholeSoldier Performance Counseling Form is provided in the appendix and shows how we capture the spirit of each of the 12 attribute groups by using all of the words in Figure 1. For example, the words maturity, discipline, reliability, bearing, and coolness all provide context for the type of conduct desired from soldiers. When evaluating soldiers in realistic settings, leaders thus use their mind as an informal synthesizer (Keeney and Raiffa 1976) when considering the words in an attribute group.[4]

2.2. Preference Elicitation

An additive value function is the simplest and most commonly used aggregation method in multiattribute decision analysis (Keeney and Raiffa 1976):

$$v(x_j) = \sum_{i=1}^{n} w_i v_i(x_{ij}),$$

where $v(x_j)$ is the total value of alternative j; $i = 1$ to n are the 12 attribute groups; x_{ij} is alternative j's score on attribute group i; $v_i(x_{ij})$ is the single-dimensional value of alternative j on attribute group i; and w_i is the weight of attribute group i. This approach is consistent with Keeney and Raiffa's (1976, p. 115) discussion of looking for "natural attribute groups" as in Figure 1 to reasonably utilize additivity at the aggregate level.

2.2.1. Common Constructed Scale. For several reasons, we decided it was advantageous to utilize a single constructed scale for assessment of performance in all 12 attribute groups. First, consultation with leaders exposed the existence of a common framework for discussion of performance regardless of the attribute in question. Second, the operability of the model is enhanced for the large and diverse population when all measures are on the same scale. Third, it allows the model to be applied to all soldiers, regardless of job, while retaining the flexibility for leaders to further specify the meaning of scale levels for particular jobs. Last, although direct natural measures arguably exist

[4] Descriptions of the attributes within the mental framework used by leaders in the U.S. Army are provided at http://www.robdees.com/uploads/1/0/6/5/10651736/wholesoldier_performance_attribute_group_descriptions.pdf.

for the "Fitness" and "Knowledge" attribute groups in particular, the use of a common constructed scale requires leaders to fulfill their responsibility of directly assessing and mentoring soldiers by providing and discussing performance ratings.

While developing a common scale based on interviews and observed discussions, we noted that leaders first categorize a soldier's performance as good, neutral, or bad. Also, within the "good" and "bad" categories of performance, leaders naturally split performance into three sublevels by using common modifiers that relate to frequency of behavior, severity of impact, or other common jargon as illustrated by the seven levels in Table 1.

To measure performance within the mental framework that is natural for decision makers, we utilize this seven-point scale for all attribute groups. For example, it is routine to hear comments such as "that guy is solid; he is highly motivated most of the time" to refer to individual performance (Level 5). The neutral category (Level 3) is often the easiest to recognize in statements like "when completing tasks, she has enough knowledge to get by but is consistently mediocre and doesn't learn very quickly." On the negative side, comments like "his conduct is consistently undisciplined; he is a 'problem soldier' but I'm not ready to give up on him just yet" are also common (Level 1). Level 0 is a message to the organization that a soldier should be separated from the army based on performance; we found that leaders in the army generally felt that roughly 10% of their subordinates were performing at Level 0 in at least one attribute group. Of note, Table 1 is used as an ordinal scale in automated data collection. In implementation as in the appendix, the numerical labels are hidden from the user to eliminate opportunities for confusion concerning ordinal and cardinal relationships.

2.2.2. Common Single-Dimensional Value Function. Unlike Likert (1932) style instruments, we do not assume equal spacing of levels or linear returns to scale. In multiattribute decision analysis, single-dimensional value functions measure returns to scale (Kirkwood 1997). Typically, attributes in applications have different units and ranges, thus requiring distinct single-dimensional value functions. Because it is a unique feature in our application, we emphasize that the common constructed scale is an operable mental framework for value judgments to be used by leaders in the army while considering performance regardless of the attribute group. For example, in the assessment "his low motivation makes him a problem soldier," the words "problem soldier" communicate the value level within the organizational culture. As such, the common single-dimensional value function was elicited over the common scale itself, and only then were behavioral descriptions elicited to map specific behavior to scale levels. Because the value judgments are inherent in the scale for the leaders evaluating subordinates, there is only one single-dimensional value function drawn over the scale.

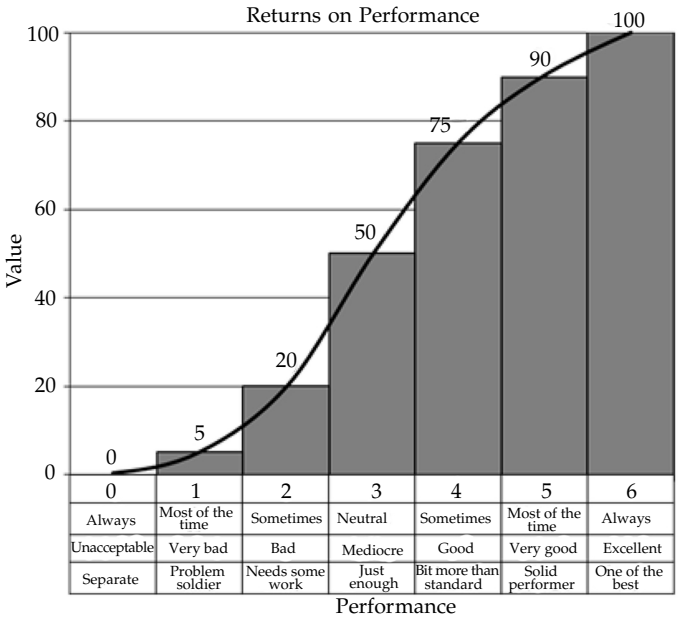
Senior leaders felt that an *S*-shaped value function (Parnell et al. 2011) was appropriate over the common scale, indicating that marginal value is achieved more rapidly in the middle of the scale rather than at the ends. For further investigation, we spent two days in a focus group setting with 96 platoon leaders and platoon sergeants from Third Brigade Combat Team, First Cavalry Division. These personnel were selected by their superiors as respected leadership teams at the platoon level. On the first day, we confirmed the value hierarchy and elicited a value model; the second day was spent conducting an initial data collection. Figure 2 displays the elicited discrete value function and also shows a corresponding continuous function to more clearly communicate the shape.

Because data on an interval rather than an ordinal scale are desired, and to facilitate elicitation

Table 1 Constructed Scale for WholeSoldier Objective Groupings

	Bad			Neutral	Good		
	0	1	2	3	4	5	6
Frequency	"Always"	"Most of the time"	"Sometimes"	"Neutral"	"Sometimes"	"Most of the time"	"Always"
Impact	"Unacceptable"	"Very bad"	"Bad"	"Mediocre"	"Good"	"Very good"	"Excellent"
Category	"Separate"	"Problem soldier"	"Needs some work"	"Just enough"	"Bit more than standard"	"Solid performer"	"One of the best"

Figure 2 Elicited *S*-Shaped Value Function



with the large group, we assume weak difference independence (Dyer and Sarin 1979) to generate a measurable multiattribute value function. First, after basic instruction concerning value functions and return to scale, the group confirmed the appropriateness of an *S*-shaped value function. We set the endpoints of the value scale to 0 and 100 by convention and then iteratively developed the value function with the group by discussing and confirming ratios of intervals on the value scale. To do this, we had each of the 48 leadership teams independently discuss a value function while guiding them through the process. Next, we facilitated a group discussion and adjusted the value function on a screen until consensus was reached. For instance, the leaders concurred that moving from level 2 to level 3 rating offers twice the return as moving from level 1 to level 2. Although this approach did not allow for formal analysis of consistency between the teams after the group discussion, consensus was easily reached, and it did overcome the challenges of the large group and the time afforded. We note that there are diminishing returns to positive performance, but that the increasing returns in moving from negative to neutral performance are more pronounced. A "Problem soldier" offers only minimally more value than a soldier that falls in the "Separate" category. Last, there is not a large difference in value between a "Solid performer" and "One of the best," but this difference was confirmed to be twice the magnitude of the value difference between the two most negative levels.

2.2.3. Behavioral Description of Scale Levels for Attribute Groups. Behavior is typically observed in small revelations over time by immediate supervisors. After eliciting a natural common scale and a single-dimensional value function over the scale, we moved to elicit specific behavioral mappings onto the scale for each attribute group. This was intuitive for the leaders who would use the model, and the descriptions of behavior serve to clarify the levels on the common scale for each attribute group. Along with the common scale, we provide these clarifying descriptions of positive and negative behavior shown in the appendix, the WholeSoldier Counseling Form. Because we are using a constructed verbal scale and single-dimensional value function to quantify leaders' insights on an interval scale, the model is still somewhat subject to different people's interpretations of the words used. But we have clarified far beyond the simple descriptions—e.g., "Success" and "Excellence"—used in the Noncommissioned Officer Evaluation Report or other commonly employed Likert-style instruments (1932) that provide relatively unclear ordinal ranges (often assumed as interval) between descriptions "strongly agree" to "strongly disagree." When using the WholeSoldier Counseling Form, assessors expressed great comfort with the constructed scale, the behavioral descriptions, and their ability to assess levels of performance.

Table 2 Elicited Swing Weights

Moral 56% (%)						Cognitive 26% (%)			Physical 18% (%)		
Purpose	Motivation	Interaction	Conduct	Character	Self-esteem	Judgment	Application	Knowledge	Fitness	Athleticism	Health
10	9	9	10	10	8	9	9	8	6	6	6

2.2.4. Weights. In the additive model, swing weights sum to one and are the value achieved by moving the score on an attribute group from its least preferred to most preferred level (Kirkwood 1997).

We elicited swing weights (Table 2) in the same focus group of 96 platoon leaders and platoon sergeants by using the weighting process described by Kirkwood (1997) with each pair of leaders. We first considered the increments in value that would occur by increasing each of the attribute groups from the least preferred to the most preferred level. Then we asked the leaders to scale each of the value increments as a multiple of the smallest value increment, or to make $(n-1)$ pairwise ratio comparisons and obtain weights by using the requirement to sum to unity.

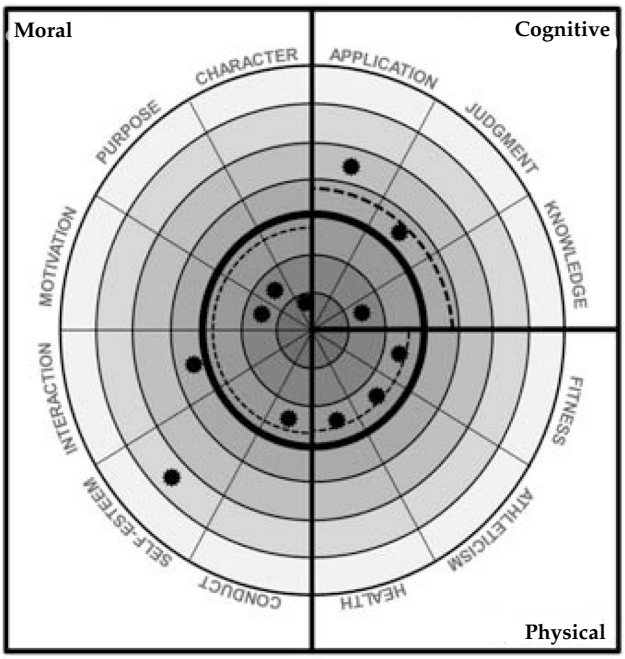
Finally, we aggregated the swing weights using simple averaging and presented them to the group to reach consensus. The general sentiment was that "If these kids show up with heart, then I can train their bodies and minds," and so the 56% weight on the moral domain was corroborated. At the attribute group level, they concurred that purpose, conduct, and character are weighted slightly more than the other attribute groups. Overall, the elicited swing weights were viewed as reflecting the organizational preference of leaders at the platoon level where junior enlisted soldiers are employed, observed, and assessed by leaders.

2.3. Initial Test and Data Visualization for Use in Practice

With a complete value model, we facilitated an initial data collection using WholeSoldier Performance with soldiers ($n = 195$) from the Third Brigade Combat Team, First Cavalry Division. We present several visualizations and possible uses of this data to facilitate mentoring, personnel decisions, and rater accountability in the process of soldier assessment.

2.3.1. Mentoring. The first benefit of WholeSoldier Performance assessment is improvement in a rater's ability to *mentor* a subordinate. We developed the WholeSoldier Target (Figure 3) to display the rater's assessments in a single graphic we refer to as the subordinate's "shot group." A tight shot group near the center of the target indicates strong performance, not unlike the evaluation of a soldier's marksmanship. The dotted arc segments generated in each domain represent the value achieved in each respective domain, and the bold circle denotes the overall WholeSoldier Performance achieved. Variations of the target were considered, including reflecting the weights in the size of each "wedge" of the target and spacing the "rings" on the value scale or the assessment scale. Because the purpose of the graphic is primarily to summarize assessments and support mentoring discussions with a general audience, and also based on the desire to retain flexibility for leaders to discuss preference in specific contexts, we decided on a simpler representation without reflection of weights and on the scale in which assessments are made.

Figure 3 Infantryman #24 WholeSoldier Target



With the WholeSoldier Target, it is easy both to mentor a soldier and understand performance with much higher fidelity than with any currently existing system. While using the target shown in Figure 3, a leader expressed the following (summarized) sentiments to his subordinate, Infantryman #24:

> Based on the past few months, I have some feedback for you. In the moral domain, I greatly appreciate your character and the fact that you are both selfless in purpose and highly motivated to accomplish the mission. Your conduct is mature, but I have noticed that you sometimes have problems interacting with the team.

> Additionally, some things that you have said and done indicate that you don't have high confidence or feel that you are a valuable team member. You have the required knowledge, but it seems like you have difficulty using this knowledge to make decisions in situations that are constantly changing. This is also reflected in the fact that you sometimes need to better plan and execute once a decision is made. Relating this back to the moral domain, I think you understand these difficulties and that this drives your low-self esteem. Over the next months, we will work together on your judgment, application, and team interaction. I think that this will help to boost the perception you have of yourself and help the team to better accomplish the mission. Finally, you continue to be one of the stronger guys in the platoon when it comes to physical stuff to include rest and nutrition; keep it up.

When looking at a WholeSoldier Target, we often get a sense that we know the soldier in question and believe the mentoring benefits alone are enough to justify broad implementation. Through the lens of experience, army leaders are able to identify and understand the performance of particular soldiers through the target graphic. During this initial implementation, the WholeSoldier Target has prompted some of the best discussions of individual performance and proactive leader strategies for improvement that we have ever observed as army officers.

2.3.2. Decision Support. Unlike any other current system, WholeSoldier Performance allows the army to visualize the holistic performance of all soldiers rather than relying on disparate indicators that provide information only on limited subsets of individuals in populations. For instance, the army currently tracks individual indicators like disciplinary action and meritorious awards, but these measures only identify small subsets of individuals rather than providing information on all soldiers. Figure 4 summarizes three platoons' WholeSoldier Performance data; each row corresponds to a soldier and provides attribute group ratings along with calculated WholeSoldier Performance. A three-color scale (green, yellow, red) with gradation is used to indicate performance from best to worst, respectively, and the soldiers are rank ordered based on the WholeSoldier Performance column.

WholeSoldier population data can be used to support a variety of decisions concerning current personnel. Leaders can determine those individuals that are best qualified or most in need of individual *training* and measure the return on investment of training and education programs. To develop soldiers across multiple dimensions, the army can *assign* them to jobs that would help them develop in areas of weakness or jobs that reinforce strengths. Currently, the army only offers flat-rate *retention* (reenlistment) incentives to soldiers in a given job; Wardynski et al. (2009–2010) have shown this to be a failed retention strategy in the officer domain. With WholeSoldier Performance, the army can offer individualized reenlistment bonuses or other incentives to retain the people they want for the jobs they need.

WholeSoldier Performance also facilitates *promotion*. For instance, if a soldier displays moral and physical performance but is lacking in the cognitive domain, then leaders may desire to delay his or her advancement to the rank of sergeant. We do not advocate that rank ordering populations by scores should replace decisions by boards, but rather that the model can allow boards to focus in on those individuals near a boundary between "promote" and "do not promote." Last, the population data in Figure 4 show that the army can use WholeSoldier Performance to *separate* poor performers as needed based on lack of merit; this is particularly relevant in the upcoming period of personnel drawdown. In sum, WholeSoldier Performance allows the army to understand, visualize, and rank order the performance of individuals in populations to better train, assign, retain, promote, and separate current personnel.

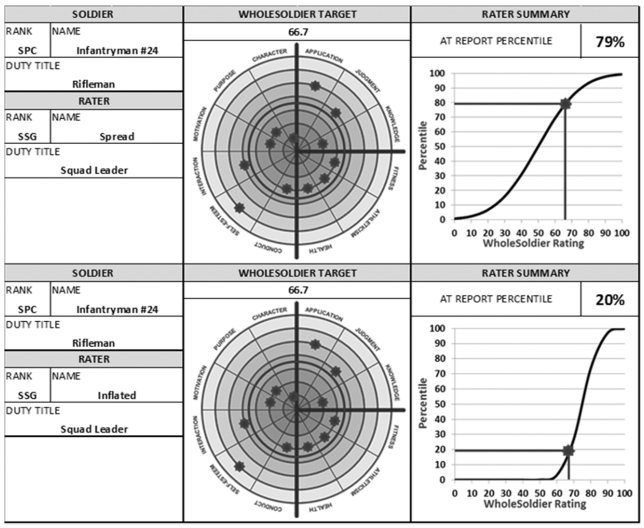
2.3.3. Rater Accountability. In the U.S. Army and other organizations, performance assessment systems are often subject to concerns such as supervisors just checking a box to minimize the time invested in assessment, gaming the system to make everyone look good, or inflating reports (Hamilton 2002). All three concerns result in individuals being indistinguishable to the organization in rating data, and all three are the consequence of misaligned leader incentives combined with a failure of raters to fulfill their responsibility to objectively rate performance. We propose that visualization of a rater's distribution of past ratings (Figure 5) provides a tool to incentivize a culture of truth through transparency. The top panel of Figure 5 displays rating information from a hypothetical "spread" rater and the bottom panel from an

Figure 4 WholeSoldier Population Data for Three Infantry Platoons

DOMAIN	MORAL						COGNITIVE			PHYSICAL				
DOMAIN WEIGHT	56%						26%			18%				
OBJECTIVE GROUPING	Purpose	Motivation	Interaction	Conduct	Character	Self-Esteem	Judgment	Application	Knowledge	Health	Athleticism	Fitness	WholeSoldier Performance	WholeSoldier Rank
OBJECTIVE WEIGHT	10%	9%	9%	10%	10%	8%	9%	9%	8%	6%	6%	6%		
	6	6	6	6	6	6	6	6	6	6	6	6	100.0	1
	4	5	6	5	6	6	6	6	6	6	6	6	93.5	2
	5	5	4	5	6	4	5	5	6	6	4	5	83.5	3
	5	5	5	5	6	4	5	4	4	6	5	5	81.8	4
	5	5	4	5	6	5	5	5	5	6	5	2	81.5	5
	6	5	5	2	5	6	5	5	5	6	4	3	79.3	6
	4	3	5	5	6	5	6	5	5	3	3	3	75.8	7
	4	5	4	4	5	4	5	4	5	5	5	5	75.7	8
	2	2	5	5	5	5	6	6	6	1	3	5	72.2	9
	4	5	4	4	5	5	3	4	4	5	4	5	71.7	10
	4	5	4	4	5	3	4	4	4	5	5	4	70.5	11
	5	4	5	5	6	4	3	3	5	2	2	5	70.2	12
	4	5	4	4	5	4	3	4	3	5	5	4	69.0	13
	4	4	4	4	4	4	4	4	4	4	5	5	68.7	14
	5	4	5	3	5	4	3	3	4	4	4	5	67.8	15
	5	5	3	4	5	2	3	4	4	4	5	2	64.8	16
	4	5	3	3	4	4	5	4	3	5	4	2	64.2	17
	4	5	3	4	4	4	3	4	3	4	5	3	63.8	18
	4	4	3	4	5	3	4	5	3	4	3	1	61.7	19
	5	5	3	4	6	1	2	3	5	4	4	1	61.5	20
	3	4	4	3	4	4	4	3	5	3	3	2	59.2	21
	4	5	3	3	3	2	3	3	4	5	5	3	58.7	22
	3	3	3	3	3	3	3	4	4	4	5	5	57.8	23
	4	5	4	3	2	4	3	2	2	5	4	4	57.0	24
	4	4	4	3	4	1	3	4	2	6	6	0	56.8	25
	4	3	1	1	2	5	4	3	3	6	6	6	56.8	26
	4	3	3	3	4	3	3	4	3	4	3	3	55.8	27
	3	4	4	3	4	3	2	3	3	2	4	3	53.2	28
	3	3	2	3	4	2	2	1	2	6	6	6	52.0	29
	2	2	4	3	4	4	4	3	3	1	3	3	50.8	30
	3	3	3	3	3	3	3	3	3	3	3	3	50.0	31
	3	3	3	3	3	3	3	3	3	2	3	3	49.0	32
	0	1	4	2	3	3	4	6	6	1	2	3	48.8	33
	3	3	2	3	3	3	3	3	4	1	3	3	47.8	34
	3	4	2	3	3	2	2	3	2	3	4	3	46.8	35
	3	2	2	3	3	2	4	4	4	1	1	3	46.0	36
	3	4	2	1	2	3	4	4	1	4	4	1	45.3	37
	3	2	2	1	3	1	2	3	2	5	5	5	44.2	38
	2	3	4	2	2	3	2	3	3	2	3	3	44.0	39
	3	3	3	1	1	3	3	3	3	3	3	3	43.3	40
	4	3	1	1	1	1	2	3	2	4	6	3	40.5	41
	2	4	4	1	2	3	2	3	2	3	0	1	38.5	42
	1	1	1	0	1	3	4	4	4	3	3	3	36.7	43
	1	2	2	2	1	1	1	2	4	2	5	4	34.8	44
	2	3	2	3	3	1	0	1	2	1	0	3	30.3	45
	0	0	4	0	0	3	3	3	3	1	1	1	26.0	46
	1	2	2	1	0	2	1	2	1	2	0	0	19.8	47
	0	0	0	0	0	0	1	1	1	1	1	1	7.3	48
	0	0	0	0	0	0	0	1	0	0	0	0	1.5	49

"inflated" rater. The two targets, both resulting in a WholeSoldier Performance of 66.7, are identical, but their meanings are different when given by different raters. On the top right, we display the distribution of the raters' past assessments and the rated subordinate's percentile rank with respect to all others. With the spread rater, the rating places the soldier in the 79th percentile, whereas the same rating from the inflated rater places the soldier in the 20th percentile. Providing individuals with their raters' distribution

Figure 5 Standardization of Ratings



significantly reduces the opportunity for a discrepancy between the mentoring discussion and the subordinate's performance relative to others. As such, it offers a cultural incentive for truth in performance assessments while discouraging gaming and inflation.

Performance rating distributions also allow the organization to hold raters accountable for their responsibility to correctly differentiate among the performance of individuals. A spread distribution clearly shows more differentiation than a narrow one. Of greater interest, correct differentiation by a rater can be analyzed retrospectively in light of future performance ratings given by different raters. Raters whose performance assessments prove to be predictive of future performance in the organization can be rewarded. In this way, WholeSoldier Performance not only facilitates mentoring and decisions concerning the rated individual, but also provides the organization a mechanism to assess, incentivize, and make decisions regarding raters.

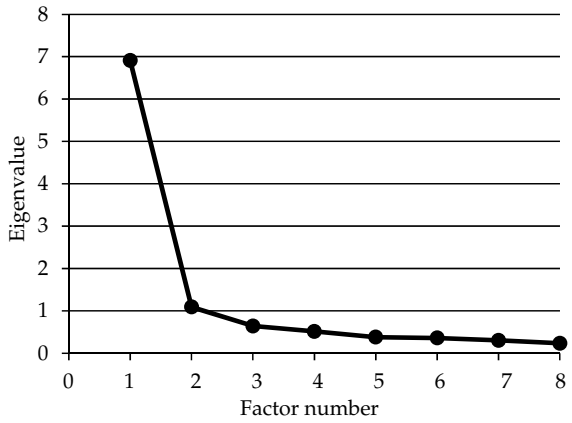
2.4. Model Validation

In general, decision analysis models are validated through concurrence or consensus that the model reflects the preferences of the decision maker or group. We received consensual support from both senior decision makers and large numbers of lower-level stakeholders at every stage of modeling. Additionally, the Military Operations Research Society awarded this work the Barchi Prize in 2010 as the best research effort in the military community presented at the previous year's symposium.[5] General Dempsey, former chief of staff of the U.S. Army and the current chairman of the Joint Chiefs of Staff, offered that "the Army thirsts for such a mentoring tool that is useful for evaluations" (Dempsey 2009).

One reason for a focus on validation via interaction with decision makers is that most implementations of multiattribute decision analysis occur when a significant decision among a relatively small number of alternatives is made once. For example, the military has used multiattribute analyses to support one-time decisions concerning materiel acquisitions, future concepts, force mix, training plans, etc. (Parnell 2007). In models like WholeSoldier Performance that are meant for routine assessment and continuous decision support over time, data are generated, and there are unique opportunities to confirm the assessed model with tools from the field of psychometrics. Cronbach's alpha (1951) is the standard measure of the internal consistency or reliability of a measure, is scaled between 0 and 1, and is interpreted as the percentage of time the measure will be reliable in practice. Cronbach (1951, p. 297) stated that the "reliability coefficient demonstrates whether the test designer was correct in expecting a certain set of items to yield interpretable statements about individual differences." In our initial data collection on the 12 attribute groups, we observed a Cronbach's alpha of 0.945, categorized as "excellent" in the field and suggesting the retention of a single factor in factor analysis. The Kaiser–Meyer–Olkin measure of sampling adequacy (Kaiser 1970) is 0.917 for our data set, which Kaiser (Dziuban and Shirkey 1974, p. 539) categorized as "marvelous." Additionally, Bartlett's test of sphericity (Bartlett 1950) yields a significance of 0.000, indicating that the data are appropriate for factor analysis. Fabrigar et al. (1999) argued that if the assumption of multivariate normality is severely violated in the data, then a principal factor method should be applied; we employ principal axis factoring. The first four eigenvalues are 6.913, 1.095,

[5] Barchi Prize information is available at http://www.mors.org/recognize_excellence/richard_ h_barchi_prize.aspx.

Figure 6 Scree Plot



0.648, and 0.518. Based on the scree test (Cattell 1966), we retain one factor because there is only one dimension to the left of the elbow, as shown in Figure 6.

With one factor, 62.8% of the variance is accounted for. Although we might have expected to see three distinct factors based on the three domains in the value hierarchy, the data clearly support one factor, which we call WholeSoldier Performance. It is interesting to note that the swing weighting procedure considers attributes only at the bottom level of a hierarchy, and as such they are considered independent of the number of domains at a higher level in the hierarchy. All of the factor loadings, which represent item correlations with the underlying factor, are above 0.6, suggesting that all items (attribute groups) should be retained. DiStefano et al. (2009) discuss various methods of using factor loadings to generate an overall score; summing item scores and weighting item scores with factor loadings are both discussed. Similar to Kirkwood's (1997) discussion of weights in decision analysis, they point out that summing item scores blindly assumes equal weight; we normalize the loadings to sum to one as in the additive value model. The normalized loadings (Table 3) reflect the elicited weights nearly identically.

In the context of psychometric measurement, elicited swing weights are statements about how important the items are relative to a single underlying factor (total value) a priori; we find it compelling that experts without data and factor analysis on collected data yielded nearly the same weights. Along with the concurrence of stakeholders, we take this application of factor analysis to our data as additional validation of the attribute groups themselves along with their associated weights.

3. Conclusion

3.1. Summary

The Command Sergeant Major of the Army, Raymond F. Chandler III, recently stated that commanders and their noncommissioned officers will have the biggest impact in deciding who will stay and who will go in the upcoming drawdown, and provides guidance that these leaders should use the WholeSoldier concept in making decisions (Mattson 2012); we provide a model to implement this view. In the army officer domain, Wardynski et al. (2009–2010) outline a *talent management system* to help the army achieve its overall objectives and discuss an information technology solution. They propose that the central activities are accessing (includes screening, vetting, and culling), developing, retaining, and employing talent. In their terms, we propose that there must also be an underlying *talent measurement system* like WholeSoldier Performance to support these activities. We recommend that the army replace the current developmental counseling form used to counsel soldiers with the WholeSoldier Performance Counseling Form to routinely and quantifiably assess the performance of soldiers; it can be implemented for relatively low cost in an information technology solution to facilitate automated generation of visualizations that support mentoring and

Table 3 Normalized Loadings and Swing Weights

	Purpose	Motivation	Interaction	Conduct	Character	Self-esteem	Judgment	Application	Knowledge	Fitness	Athleticism	Health
Normalized loading (%)	10	9	9	10	10	9	9	9	8	6	6	5
Swing weight (%)	10	9	9	10	10	8	9	9	8	6	6	6

also provide data to better train, assign, retain, promote, and separate soldiers.

3.2. Related Efforts

Currently, WholeSoldier Performance is framing a rewriting of the human dimension study that initially "spent a lot of attention on materiel but not on the person we were putting in the uniform" (Tan 2012). Outside the military, the first author developed a WholeSurgeon model for the Mayo Clinic that has been implemented to assess the performance of surgical residents.

3.3. Future Work

We present WholeSoldier Performance as a model to reflect the preferences of the U.S. Army for the performance of all soldiers. To support decisions related to soldiers in specific jobs, we see two areas of future work. First, proponents responsible for the management of specific jobs may further refine descriptions of behavior related to a job or possibly refine the value model to include some natural measures along with the constructed ones. For instance, proponents might specify mappings between physical fitness test scores and WholeSoldier "fitness" ratings for combat versus noncombat soldiers or choose to include job specific tests to measure "knowledge." Second, proponents might desire to develop and utilize a revision to the WholeSoldier Performance weights to reflect varying emphases in different jobs within the army. Theoretically, this also provides opportunity for research into specific multiattribute models that are nested within the framework of a general model.

The focus of this paper is to define and measure performance, such that the army can make better personnel decisions regarding current soldiers, and future research efforts can design models using recruit measures to predict performance. Researchers are currently able to predict longevity of service to some degree, but are unable to predict performance levels because of the lack of performance data collected routinely across the entire force. With WholeSoldier Performance, we offer such future studies a response variable for use in longitudinal studies of recruit measures that are known before recruiting decisions are made. This requires theoretical investigation into the aggregation of performance data collected over time by different raters. We propose that transforming WholeSoldier Performance scores into percentile ranks as in §2.3.3 might be viewed as a logical way to account for the effect of multiple raters, but aggregation of multiple ratings over time to produce a single value for use in a predictive model is a separate issue warranting deeper investigation. Factors for consideration might include duration of the performance report, the specific job performed during each reporting period, and whether recent reports should receive more weight than older ones. Such predictive modeling with WholeSoldier Performance as a response variable will allow army leaders to better understand the impacts on soldier performance when adjusting recruiting policy, which was the original need expressed at the outset of this work. This is also theoretically related to the distinction between preference and prediction models along with their interaction as addressed by Butler et al. (2006), but with the added benefit of having data to support the establishment or refinement of the predictive model.

Performance appraisals, particularly those in large organizations, provide large amounts of data that support repeated decisions. We utilize the psychometric tool of exploratory factor analysis to gain insight into the validity of retaining all 12 attribute groups and their associated weights. With broad implementation and more data, confirmatory factor analysis would also be appropriate. Traditionally, decision makers validate multiattribute models, but are continually concerned with the validity of the model and any updates that should be made over time. We see a rich opportunity to further investigate the validation and refinement of multiattribute models that generate large amounts of assessment data.

Acknowledgments

The authors thank Jim Dyer, John Butler, Greg Parnell, Ralph Keeney, and two reviewers for their insightful comments. They thank COL Gary Volesky and CSM James Pippin for facilitating their initial WholeSoldier data collection effort and all of the many soldiers, noncommissioned officers, and officers that spent countless hours in consultation.

Appendix. WholeSoldier Counseling Form

WholeSoldier Performance Counseling Form

PRINCIPAL PURPOSE:	To assist leaders in conducting and recording counseling data pertaining to subordinates.
ROUTINE USES:	For subordinate leader development. Leaders should use this form at least quarterly.
DISCLOSURE:	Counseling data will be recorded in the Soldier's online file.

PART I - ADMINISTRATIVE DATA

Soldier Name (Last, First, MI)	Soldier Rank	Soldier Pos'n	Soldier MOS	Soldier AKO	Date
Infantryman # 24	**PFC**	**Rifleman**	**11B**		
Leader Name (Last, First, MI)	**Leader Rank**	**Leader Pos'n**	**Leader MOS**	**Leader AKO**	**Organization**
	SFC	**Platoon Sergeant**	**11B**		

PART II - EVALUATION OF PERFORMANCE

SCALE	BAD			NEUTRAL	GOOD		
Frequency	"Always"	"Most of the time"	"Sometimes"	"Neutral"	"Sometimes"	"Most of the time"	"Always"
Impact	"Unacceptable"	"Very Bad"	"Bad"	"Mediocre"	"Good"	"Very Good"	"Excellent"
Category	"Separate"	"Problem Soldier"	"Needs some work"	"Just Enough"	"Bit more than standard"	"Solid Performer"	"One of the Best"

MORAL DOMAIN

PURPOSE (Why): Selfless Service, Sacrifice, Commitment, Loyalty, Duty

○	○	○	○	○	●	○
Not a team player and displays selfish attitude. Tends to put personal desires before others and unit mission.			**Marginal.**	**Committed to performing duties even when sacrifice is required. Selfless member of the team with loyalty to mission and unit.**		

Examples:

MOTIVATION (Effort): Will to Win, Endurance, Resilience, Heart, Drive, Determination, Work Ethic

○	○	○	○	○	●	○
Lacks determination and drive to get the job done. Doesn't respond well to tough conditions or bounce back from setbacks.			**Marginal.**	**Possesses the will to win and puts forth best effort. Won't quit and positively responds to setbacks. Inspires Motivation in others.**		

Examples:

CHARACTER (How): Honor, Integrity, Justice, Candor, Personal Courage

○	○	○	○	○	○	●
Looks for loopholes and lacks integrity to be trusted. Won't take a stand for what is right or take responsibility for mistakes.			**Marginal.**	**Can be trusted to do and stick up for what is right. Accepts and strives to correct mistakes. Tells whole truth even when painful.**		

Examples:

CONDUCT (Personal): Maturity, Discipline, Reliability, Bearing, Coolness

○	○	○	○	●	○	○
Needs constant supervision and has problems leading a balanced life. Disrespectful and loses bearing/coolness.			**Marginal.**	**Performs well without supervision and within intent. Mature lifestyle and coolness/bearing under stress is example for others.**		

Examples:

INTERACTION (External): Respect, Empathy, Compassion, Humor

○	○	○	●	○	○	○
Cynical, negative, or inconsitent towards others. Doesn't exert effort to interact with others and/or is awkward in interaction.			**Marginal.**	**Positive, respectful, outgoing, and humorous. Makes others comfortable to share ideas/issues and adds to team atmosphere.**		

Examples:

SELF-ESTEEM (Internal): Confidence, Self-Worth, Self-Efficacy

○	●	○	○	○	○	○
Lacks confidence and is unsure of ability to accomplish mission/goals. Thinks of excuses when failure may happen.			**Marginal.**	**Displays confidence in interactions and execution of tasks. Understands value to team, isn't afraid to fail, and believes he/she is up to the task.**		

Examples:

Appendix. Continued

SCALE	BAD			NEUTRAL	GOOD		
Frequency	"Always"	"Most of the time"	"Sometimes"	"Neutral"	"Sometimes"	"Most of the time"	"Always"
Impact	"Unacceptable"	"Very Bad"	"Bad"	"Mediocre"	"Good"	"Very Good"	"Excellent"
Category	"Separate"	"Problem Soldier"	"Needs some work"	"Just Enough"	"Bit more than standard"	"Solid Performer"	"One of the Best"

COGNITIVE DOMAIN

KNOWLEDGE (Information): Job Tasks/Skills, Education, Trainability, Learning

○	○	○	○	○	●	○
Untrainable and has shown an inability to learn. Lacks the technical competence to complete tasks			Marginal.	Knows tasks two levels up. Capable of higher learning. Soldier is an intelligent, life long learner.		

Examples:

JUDGMENT (Reasoning): Common Sense, Logic, Insight, Understanding, Anticipation, Adaptive, Flexible

○	○	○	●	○	○	○
Displays lack of good judgment. Does not apply common sense, or recognize important factors in varying situations.			Marginal.	Makes good decisions in routine situations and new ones. Sees the big picture and what is important. Can change course of action when needed.		

Examples:

APPLICATION (Action): Planning, Communicating, Executing

○	○	●	○	○	○	○
Soldier is continually reliant on others. Can't handle more than one task a time or lead others in a plan. Doesn't get the job done.			Marginal.	Able to apply knowledge/judgment to complete complex tasks. Able to organize team to execute multiple tasks in support of mission.		

Examples:

PHYSICAL DOMAIN

FITNESS (Traditional): Cardio Endurance, Cardio Strength, Muscular Endurance, Muscular Strength

○	○	○	○	●	○	○
Does not meet established Army standards. Cannot carry his/her share of the load. Poor performance in unit PT.			Marginal.	Carries more than his/her share of the load. Exceeds Army standards and excels during PT.		

Examples:

ATHLETICISM (Functional): Coordination, Agility, Balance, Power, Speed, Accuracy, Flexibility, Reaction Time

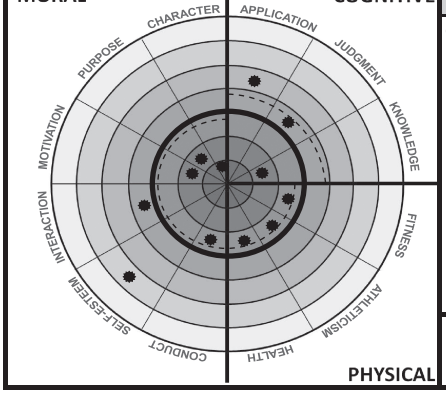
○	○	○	○	●	○	○
Soldier moves awkwardly and is unathletic in tasks requiring coordination. Soldier cannot fight, or live up to unforseen physical challenges.			Marginal.	Soldier is an athlete and can perform under a variety of conditions. Can transfer ability to nearly any task during the mission.		

Examples:

HEALTH (Balance): Nutrition, Rest, Resistance to Illness

○	○	○	○	●	○	○
Unhealthy habits contribute to poor performance. Regularly on profile or at sick call. Fails to meet bodyfat % standards.			Marginal.	Not hindered by sickness/injury. Demonstrates balance in rest, nutrition, and personal habits. Maintains a reserve and meets demands.		

Examples:



PART III - PLAN OF ACTION

Comments:

WholeSoldier Performance (0 to 100):

References

Bartlett MS (1950) Tests of significance in factor analysis. *British J. Psych.* 3(2):77–85.

Butler JC, Dyer JS, Jia J (2006) Using attributes to predict objectives in preference models. *Decision Anal.* 3(2):100–116.

Cattell RB (1966) The scree test for the number of factors. *Multivariate Behav. Res.* 1(2):245–276.

Cronbach LJ (1951) Coefficient alpha and the internal structure of tests. *Psychometrika* 16(3):297–334.

Dempsey M (2009) Interview by Rob Dees. Personal notes, December 9. West Point, NY.

DiStefano C, Zhu M, Mindrila D (2009) Understanding and using factor scores: Considerations for the applied researcher. *Practical Assessment, Res. Eval.* 14(20):1–11.

Dyer JS, Sarin RK (1979) Measurable multiattribute value functions. *Oper. Res.* 27(4):810–822.

Dziuban C, Shirkey E (1974) When is a correlation matrix appropriate for factor analysis? Some decision rules. *Psych. Bullet.* 81(6):358–361.

Fabrigar LR, Wegener DT, MacCallum RC, Strahan EJ (1999) Evaluating the use of exploratory factor analysis in psychological research. *Psych. Methods* 4(3):272–299.

Hamilton CR (2002) The effects of multiple constraints on the Army's new officer evaluation report. United States Marine Corps Command and Staff College, Quantico, VA.

Headquarters, Department of the Army (2005) Field manual 1, The Army. Headquarters, Department of the Army, Washington, DC.

Headquarters, Department of the Army (2008) The U.S. Army Study of the Human Dimension in the Future 2015–2024. TRADOC Pamphlet 525-3-7-01, Training and Doctrine Command, Headquarters, Department of the Army, Fort Monroe, VA.

Kaiser HF (1970) A second generation little jiffy. *Psychometrika* 35(4):401–415.

Kane T (2011) Why our best officers are leaving. *The Atlantic* (Jan/Feb). http://www.theatlantic.com/magazine/archive/2011/01/why-our-best-officers-are-leaving/308346/.

Kaplan RS, Norton DP (1992) The balanced scorecard—Measures that drive performance. *Harvard Bus. Rev.* 70(1):71–79.

Kaplan RS, Norton DP (1996) Using the balanced scorecard as a strategic management system. *Harvard Bus. Rev.* 74(1):75–85.

Keeney RL (1992) *Value-Focused Thinking: A Path to Creative Decisionmaking* (Harvard University Press, Cambridge, MA).

Keeney RL (2000) Designing a balanced scorecard. Presentation, INFORMS Annual Meeting, Institute for Operations Research and the Management Sciences, November 6, San Antonio, TX.

Keeney RL, Raiffa H (1976) *Decision Making with Multiple Objectives: Preferences and Value Tradeoffs* (Wiley, New York).

Kirkwood CW (1997) *Strategic Decision Making: Multiobjective Decision Analysis with Spreadsheets* (Duxbury Press, Pacific Grove, CA).

Likert R (1932) A technique for the measurement of attitudes. *Archives of Psych.* 22(140):1–55.

Mattson J (2012) Army to enforce standards, retain quality soldiers during drawdown. Accessed January 22, 2013, http://www.army.mil/article/75726/.

Parnell GS (2007) Value-focused thinking. Loerch A, Rainey L, eds. *Methods for Conducting Military Operational Analysis: Best Practices Throughout the Department of Defense*, Chap. 19 (Military Operations Research Society).

Parnell GS, Driscoll PJ, Henderson DL, eds. (2011) *Decision Making in Systems Engineering and Management*, 2nd ed. (John Wiley & Sons, Hoboken, NJ), 331–339.

Rostker B (2006) *I Want You! The Evolution of the All-Volunteer Force* (RAND Corporation, Santa Monica, CA), 314–317.

Schinnar AP, Wood L, Kaveroj PD, Nord RD, Schmitz EJ (1988) Recruit quality, soldier performance, and job assignment. DTIC accession number ADA197082, University of Pennsylvania, Philadelphia.

Symons RW, Bilberry RW, Caram MH, Luallin JS, Mason RS (1982) What is a quality soldier? DTIC accession number ADA118851, U.S. Army War College, Carlisle Barracks, PA.

Tan M (2012) Army seeks new recruiting, training strategies. *Army Times* (May 13). http://www.armytimes.com/news/2012/05/army-human-dimension-seeks-new-recruiting-training-strategies-051312w/.

Wardynski C, Lyle D, Colarusso M (2009–2010) Towards a US Army Officer Corps strategy for success. Six monograph series on talent management, Strategic Studies Institute, U.S. Army War College, Carlisle Barracks, PA.

Robert A. Dees is a Ph.D. candidate in risk analysis and decision making in the Department of Information, Risk, and Operations Management (IROM) within the McCombs School of Business at the University of Texas at Austin. Previously, he served as an assistant professor teaching decision analysis and management science in the Department of Systems Engineering at the United States Military Academy at West Point. He holds an M.S. in industrial and systems engineering from Texas A&M University and a B.S. in engineering management from the United States Military Academy. He is a Major in the United States Army, where he serves as an operations research analyst. His research interests include the theory and practice of decision analysis. In particular, he is interested in personnel decisions, the use of data in decision analysis, and logical updating of decision analysis models. Additionally, he has worked on research involving probabilistic scoring, Monte Carlo simulation, gaming for training, and in efforts to advocate decision analysis as an essential skill for junior military leaders. In the Army, he has held the leadership positions of infantry platoon leader, scout platoon leader, and infantry company commander.

Scott T. Nestler is an assistant professor in the Operations Research Department at the Naval Postgraduate School, where he has been teaching courses in risk benefit analysis, business modeling and analysis, and statistics. He received his Ph.D. in business and management (major: management science, minor: finance) in 2007 from the Robert H. Smith School of Business, University of Maryland, College Park. Additionally, he has an M.S. in applied mathematics (operations research) from the Naval Postgraduate School and a B.S. in civil engineering from Lehigh University. His research interests include decision and risk analysis, simulation modeling, and data analysis and visualization. He is a member of the INFORMS Analytics Certification Task

Force and the program committee for the INFORMS Conference on Business Analytics and Operations Research. He is a Colonel in the United States Army, where he serves as an operations research analyst. His assignments include a stint as the Chief of Strategic Assessments for Multi-National Force–Iraq; service in the Pentagon as a force structure analyst; and two assignments to the U.S. Military Academy at West Point, as assistant professor and Director, Center for Data Analysis and Statistics in the Department of Mathematical Sciences and as a senior researcher in the Operations Research Center in the Department of Systems Engineering. Earlier in his career, he served as a leader in PATRIOT missile units, including during Operations Desert Shield and Storm.

Robert Kewley is the department head and a professor in the Department of Systems Engineering at the United States Military Academy at West Point. He earned a Ph.D. in information and decision sciences and an M.S. in industrial and managerial engineering, both from Rensselaer Polytechnic Institute in Troy, NY, and a B.S. in mathematics from the United States Military Academy. He is a Colonel in the United States Army, with experience as a leader in Armor units, including service in Operations Desert Shield and Storm. He has also served as an operations research analyst in both Iraq and Afghanistan supporting campaign planning and counter-IED analysis. His research interests include simulation interoperability and command and control architectures.