

## Context-Sensitive Attribute and Competency Assessment

**Jayne L. Allen**  
Army Research Institute  
Ft. Moore, GA  
jayne.l.allen3.civ@army.mil

**Randy J. Brou**  
Army Research Institute  
Ft. Moore, GA  
randy.j.brou.civ@army.mil

**Krista L. Ratwani**  
Aptima, Inc.  
Woburn, MA  
kratwani@aptima.com

**Frederick J. Diedrich**  
Independent Consultant  
Middlebury, VT  
frederick.diedrich@gmail.com

**Scott M. Flanagan**  
Sophia Solutions, LLC  
Durham, NC  
scott@sophiaspeira.com

**Tatiana H. Toumbeva**  
Aptima, Inc.  
Orlando, FL  
ttoumbeva@aptima.com

### ABSTRACT

Over their careers, Soldiers must exercise a variety of attributes and competencies to effectively adapt to the varying contextual demands of emerging operational conditions. It is challenging to determine when Soldiers have been adequately prepared to meet such demands because, as described in Brou et al. (2022), Soldier development is often nonlinear and context-dependent. Thus, rather than focusing on whether a Soldier “has” an attribute or competency, the assessment challenge becomes better understanding the likelihood of attribute-related performance under varying conditions (i.e., if in condition X, then ...). The purpose of this paper is to present a method to diagnose the degree to which individuals’ attribute-related performance changes across varying contextual demands. We examine how situational judgment test (SJT) items and associated rubrics can be used to reliably assess the adaptive exercising of attributes (e.g., empathetic, patient) in an instructional setting. SJT items were constructed to systematically explore responses across conditions that varied in both problem domain (e.g., calling for fire amid uncertainty, effectively working with a struggling student) and the intensity of contextual stressors (e.g., time pressure, threat level, relationship status). Using the SJT items and rubrics, we reliably measured attribute-related performance across contextual manipulations. Findings from a sample of 345 responses across SJT variants demonstrated that agreement between raters was 97% within one point on a five-point scale. Intraclass coefficients ranged from .74 to .81 across attributes reflecting excellent levels of interrater reliability (Cicchetti, 1994). Collectively, these findings suggest that the SJT items and associated rubrics provide a reliable approach to assessing the likelihood of attribute-related performance under varying conditions. Future work will use this method to explore development across time and context by enabling the assessment of individuals’ patterns of strengths and weaknesses. These assessments will in turn provide insights into the creation of targeted instructional interventions.

### ABOUT THE AUTHORS

**Jayne L. Allen** is a Research Psychologist with the U.S. Army Research Institute for the Behavioral and Social Sciences. She holds a Ph.D. in Personality and Social Psychology, as well as a Cognate in the Science of College Teaching, from the University of New Hampshire.

**Randy J. Brou** is the Chief of the Fort Moore Research Unit of the U.S. Army Research Institute for the Behavioral and Social Sciences. He holds a Ph.D. in Applied Cognitive Science from Mississippi State University.

**Krista L. Ratwani** is a Principal Scientist and the Senior Vice President of Research and Engineering at Aptima. Her work focuses on assessment and feedback tools to facilitate learning and development throughout careers. She holds a Ph.D. in Industrial-Organizational Psychology from George Mason University.

**Frederick J. Diedrich** is a consultant who focuses on methods of instruction and assessment designed to deliberately support competency and attribute development. He holds a Ph.D. in Cognitive Science from Brown University.

**Scott M. Flanagan** is a retired U.S. Army Master Sergeant having served 20 years on active duty in Special Forces and Special Mission Unit assignments. As a consultant, he has primarily focused on training and leader development in support of Army learning and assessment strategies.

**Tatiana H. Toumbeva** is a Senior Scientist and Team Lead in the Training, Learning, and Readiness Division at Aptima, Inc. with expertise in assessment and training development and validation. She holds a Ph.D. in Industrial/Organizational Psychology from Bowling Green State University.

## Context-Sensitive Attribute and Competency Assessment

**Jayne L. Allen**  
Army Research Institute  
Ft. Moore, GA  
jayne.l.allen3.civ@army.mil

**Randy J. Brou**  
Army Research Institute  
Ft. Moore, GA  
randy.j.brou.civ@army.mil

**Krista L. Ratwani**  
Aptima, Inc.  
Woburn, MA  
kratwani@aptima.com

**Frederick J. Diedrich**  
Independent Consultant  
Middlebury, VT  
frederick.diedrich@gmail.com

**Scott M. Flanagan**  
Sophia Solutions, LLC  
Durham, NC  
scott@sophiaspeira.com

**Tatiana H. Toumbeva**  
Aptima, Inc.  
Orlando, FL  
ttoumbeva@aptima.com

### INTRODUCTION

Over their careers, Soldiers must exercise a variety of attributes and competencies to effectively adapt to the varying contextual demands of emerging operational conditions. With the advent of the Army's new Operations doctrine (FM 3-0; U.S. Department of the Army, 2022), multidomain operations (MDO) has been established as the Army's operational concept involving complex, large-scale operations where Soldiers must not only adapt to changing contexts, but also exploit advantageous conditions as they emerge (ADP 6-22, U.S. Department of the Army, 2019). The tenets of MDO (e.g., agility, convergence, endurance, and depth) guide the operational imperatives of recognizing and anticipating change. The context in which tactical level units may find themselves involves continuous observation by adversaries where there are degraded command and control capabilities, diminished freedom of maneuver, and no safe havens. These contextual demands not only challenge the ability to survive on today's battlefield but also the ability to fight back with tactics and technical capabilities that may be evolving or new. The consequences of not exploring attribute development across these varying contexts may have negative implications for a leader's ability to effectively transition to the complex demands of MDO. It is challenging to determine when Soldiers have been adequately prepared to perform effectively across contexts because, as described in Brou et al. (2022), assessments related to development do not often capture the complexities associated with performance. To be effective, assessments must capture the nonlinear nature and context dependencies of performance.

For example, imagine that a new Officer Candidate School (OCS) candidate trains with a map and compass in the gently rolling terrain and thick vegetation at Ft. Moore, GA. They pass the land navigation test in this context where numerous skills and attributes are displayed (e.g., adaptability, situational awareness, navigation skills). In contrast, now imagine that as a newly commissioned officer, they find themselves attending the Infantry Basic Officer Leader Course (IBOLC) as a Platoon Leader (PL) on a training mission. In this hypothetical condition, the 'enemy' possesses the use of GPS tracking and the ability to jam radio communications. The task of being able to navigate with a map and compass is now embedded in a strategic context where the Soldier must maneuver given new constraints imposed on them: there is an enemy, and the enemy can pinpoint their location and prevent radio communication. The Soldier must decide to navigate to a position that is tactically advantageous to the mission at hand while also being able to effectively communicate without being compromised. The question is whether they will display the attributes, competencies and skills as they did in the OCS land navigation course. Is it reasonable to expect that the PL "has" the ability to exercise sound judgment while navigating with agility under new contextual constraints? How certain can we be that the PL will be successful in this context, and—beyond that—when they are deployed, and both the enemy and consequences are real? If we do not develop and use methods to better understand the attributes, skills, and competencies our warfighters possess, they may too frequently find themselves in situations for which they were not ready, ultimately compromising mission effectiveness.

With these types of complexities in mind, Brou et al. (2022) presented a theoretical framework to account for contextual influences on attribute development. Building on work from a variety of domains (e.g., Rose, 2016; Shoda et al., 2007; Thelen & Smith, 1994), Brou et al. made two fundamental claims regarding the likely complexity of Soldier development. First, they argued that Soldier development with respect to leader attributes is unlikely to be

simple and linear. For instance, individual development of the Army Values in Basic Combat Training (BCT) is not always linear (Toumbeva et al., 2019). Similarly, leader development at the United States Military Academy is not simply a pattern of continual growth, instead showing individual plateaus (e.g., Bartone et al., 2007). More generally, in the domain of skill acquisition, such plateaus are common along with “leaps” and “dips” (Gray & Lindstedt, 2017). Second, Brou et al. argued that like human behavior more generally, the display of leader attributes is likely to be context dependent. The argument was rather than “having” an attribute, it was more likely that display of attribute-related behavior depended on context in an if-then manner (if in context X, then behavior Y; e.g., Mischel & Shoda, 1995). Notably, Army leaders must operate in a variety of contexts, and moreover, the complexity of these contexts changes for leaders as they progress through their careers (e.g., Jacobs & Jacques, 1987). Hence, as discussed in Brou et al., it may not be appropriate to think about Soldiers as “having” attributes like the ability to build trust, a tolerance for ambiguity, or patience. Rather, leveraging if-then relationships, the challenge becomes better understanding the likelihood of whether a Soldier will be able to behave in accordance with an attribute under different conditions.

Given this theoretical account, the purpose of the research described here is to present a method to diagnose the degree to which individuals’ attribute-related performance changes across varying contextual demands. The first requirement for any viable assessment method is to detect performance changes reliably; the second requirement is the flexibility to systematically manipulate contextual demands. Accordingly, Brou et al. (2022) suggested that Situational Judgment Tests (SJTs), which have been shown to be effective in predicting behavior (e.g., Motowidlo et al., 2006), could satisfy both requirements. Most SJTs include a brief description of a scenario vignette followed by a “what would you do” question. Importantly, the scenario descriptions can be systematically manipulated to explore contextual effects. For instance, if we want to assess the display of patience across conditions, the scenario description might vary the threat level and time available to accomplish important training. With this in mind, we present an assessment approach that employs SJTs along with corresponding rubrics for the coding of responses. The goal of this paper is to present this foundational approach with a focus on the creation of the method and an exploration of its associated reliability. Future work will employ this approach to test the impact of context on attribute performance over Soldier development, thereby expanding on the concepts presented in Brou et al.

## **DEVELOPMENT OF ASSESSMENT METHODS**

This section describes the development of several SJT items to assess the attributes of interest under various conditions and also presents a behavioral rubric designed to aid in the assessment of performance within the SJT items. Together, these materials were developed to allow for a systematic and reliable assessment of attribute performance under various contexts.

### **Situational Judgment Test Development**

Employing an SJT methodology has the potential to help provide evidence to (a) assess how context impacts the display of attributes, (b) determine the relative importance of contextual variables on the display of attributes, and (c) identify patterns of attributes across Soldiers and across contexts. To meet those goals, assessment items that produced differences in responses in a priori defined ways needed to be developed. As an example, given their range of duties, it may be informative to collect evidence that a PL is empathetic in the context of a medical situation, as well as if that leader can be empathetic in the context of teaching a struggling counterpart. It may or may not be the case that behavior is consistent, and if not, it becomes necessary to understand the details of the situation to facilitate growth (i.e., what might the leader find salient and why). Constructing SJT scenarios to simply be “realistic” would not be the most effective approach because of the likelihood of confounding the differences in responses with the various elements of content of the scenario. Instead, items that constrain various confounding factors need to be developed to precisely conclude which situational elements have an impact and which do not. As such, the scenarios have the potential to collectively provide evidence with respect to attributes in a manner that is highly nuanced (Brou et al., 2022).

For the purposes of this research, the factorial survey (Rossi & Anderson, 1982) or experimental vignette method (EVM; Atzmüller & Steiner, 2010) approach was determined to be the most appropriate framework in which to derive conclusions about the importance of situational elements on attribute expression via SJTs. This was because factorial surveys, or EVMs, are scenario based and manipulate situational factors to understand causal effects of those factors on decision making. For the purposes of this research, the terms factorial survey and EVM are considered synonymous and EVM is used to refer to both. Moreover, although the term “vignette” is often used in the factorial survey or EVM

literature, this research uses the term “situation” to refer to the overarching context in which situational elements are manipulated. The term “scenario” refers to the specific instantiation of the manipulated elements and their levels within a situational context.

EVMs have been used to assess a range of attributes and outcomes such as clinical judgments (Ludwick & Zeller, 2001), perceptions of mental illness (Thurman et al., 1988), institutional compliance (Raaijmakers et al., 2015), sexual harassment responsibility (Pierce et al., 2000), and personnel selection decisions (Podsakoff et al., 2011). The EVM approach is based on experimentally manipulating situational element combinations (i.e., “dimensions” and “levels”) across situations to gather data on differences in response likelihoods. This information can be key for understanding where development may be necessary, for example observing sub-optimal responses given certain situational element combinations. Moreover, EVM forces developers to consider all possible situational element combinations and, therefore, the possible range of scenarios that a Soldier may encounter.

In the following section, the development of the specific situations to include the manipulation of key situational elements that should elicit differences in the expression of the attribute elements (i.e., responses) is described. Scenarios were created that were nearly identical in terms of content except for the manipulated situational elements which may produce different responses (i.e., scenario variants).

### **Developing Situations and Variants**

The selection of the situation is vital because it provides the overarching structure that binds the situational elements, target attributes, and target population together. The baseline situations developed for this work were derived by engaging in an iterative development process that included obtaining feedback from Professional Military Education (PME) instructors and other Army subject matter experts (SMEs). Four final baseline situations were developed for this research that included calling for fire, teaching a security assessment process, providing medical care, and conducting a site survey. These situations were developed to have various tradeoffs and nuances that contributed to there being no clear single “right” answer, but rather challenged Soldiers to make and justify their responses amid complexity. For the purposes of this paper, we illustrate the approach using the assessment process example, which specifically challenged Soldiers to teach a new security assessment process to coalition partners when a student is struggling to learn (Struggling Student).

Following the development of the baseline situations, the research team used the situational elements framework developed by Brou et al. (2022) to determine the elements to experimentally manipulate across the situations to produce scenarios. The implementation of the situational elements creates a level of standardization across the scenarios such that they form “variants” which would make the scenarios look similar to each other except for the manipulated situational element combinations. For this research, three elements were chosen to manipulate within the situations that put varying amounts of pressure on attribute-related actions and behaviors: (1) threat level, (2) time pressure, and (3) relationship tension (akin to the Trust element in Brou et al.’s framework). Based on discussions with the instructors and other Army SMEs, these three elements should impact how a Soldier responds to a given situation or scenario. The number of manipulated elements was purposefully kept small to reduce the overall complexity of representing these elements in each scenario. Similarly, we limited the number of levels of each situational element to two (high and low) to help reduce the number of scenarios that would need to be constructed. See Table 1 for how each level of each situational element was operationalized using the Struggling Student scenario.

**Table 1. Situational Element Levels Taxonomy for Struggling Student**

Situational Elements	Level	Example Situational Element by Level
<b>Threat</b> - Does the target individual/learner involved feel threatened (psychologically or physically)?	High	“Decentralized insurgents in the area of operations”
	Low	“No known insurgents in the immediate area of operations”
<b>Time Pressure</b> - Is time available for task completion tight?	High	“Busiest market day begins in the morning”
	Low	“Busiest market day is in a few days”
<b>Relationship Tension</b> - Do the individuals involved trust each other at the specific point in time of the situation? What is the relationship quality/trust level?	High	“Recently the relationship has been strained”
	Low	“Work well together”

The Struggling Student scenario is included below wherein the situational elements were manipulated to the high level (i.e., high time, high relationship tension, high threat levels). These manipulations are **bolded** within the scenario. A low pressure (i.e., time, threat, and relationship tension) variant is very similar to the high level one with the exception of replacing the bolded items with low-level elements based on the examples displayed in Table 1.

As an advisor, you are tasked with assisting coalition partners in an assessment of the security of a unit’s compound that is located next to a busy market area in a village with little distance from multiple shops and vendors. **It is 1600 and the busiest market day of the week begins in the morning** [*time pressure*]. **You’ve been working with the coalition commander for a few months but recently the relationship has been strained** [*relationship tension*].

**The commander tells you that he is concerned about vulnerabilities because of decentralized insurgents in the area of operations (AO)** [*threat*]. He wants you to teach his staff a process that they can use now and in the future for how to conduct a good assessment to inform security enhancements. Currently you don’t have access to typical supplies you would need such as concertina wire and T-barriers. They have a process they have used in the past.

You start to take the staff through the new assessment process, and it’s clear that members of the staff do not understand. One Soldier starts the process but keeps reverting to their original approach rather than the one you are teaching. The commander loses his patience, berates the Soldier in front of others, and demands that he quickly follow your instructions.

### Response Instructions

Following each scenario, three open-ended prompts were constructed: (1) What would you do in this situation? (2) How did you come to that decision? and (3) What are the consequences of your decision? If any negative consequences, what would you do about it? Such open-ended response questions were used to capture the thought process of the respondents more fully compared to more traditional response formats. SJTs typically follow a multiple-choice format with predetermined courses of action that a participant could select (cf., Motowidlo et al., 1997); however, such an approach may be limited in detecting idiosyncrasies in decision-making processes around attribute-related behavior, especially in the context of ill-defined problems. Open-ended responses should allow for a more nuanced understanding of what someone does or does not understand and their developmental needs. In addition, in the context of instruction, they could help generate discussion around decision-making processes in complex situations to foster critical thinking and learning as students consider the implications of their choices. Furthermore, because

there are not yet any specific hypotheses regarding how situational elements may impact the display of attributes, the open-ended responses enabled the capture of multiple attributes of interest; forced choice options would necessarily need to be designed around the display of targeted attributes.

### Rubric Development

In order to help reliably assess performance within each SJT item response, we developed behavioral-based rubrics that utilize a 1 to 5 Likert-type scale. A general rubric for each attribute was developed with input from PME instructors and training developers for use across a variety of instructional situations (e.g., live exercises with role players, tactical decision exercises in the classroom, etc.). In the context of the work presented here, coders employed this general rubric to make an overall assessment of each response. Given the goals of the specific Program of Instruction (POI) involved in this work, the instructors assessed students on a range of attributes that were defined in a manner specific to their instructional goals. While many of the attributes were similar in focus to those that appear in the Leader Requirements Model (LRM; ADP 6-22, U.S. Department of the Army, 2019), there were subtle differences in emphasis given instructional objectives. As an example, Table 2 shows attribute “Empathetic.” Below we show results for a sample of the attributes addressed that cover a range of areas that have similarities to various items across the broad categories of Character, Presence, Intellect, Leads, Develops, Achieves as employed in the LRM.

**Table 2. General Attribute Rubric for Empathetic**

1	2	3	4	5
Does not ask about, listen to, or acknowledge others' perspectives	<i>Mix of 1 and 3</i>	Listens to and acknowledges others but may not always probe for additional information	<i>Mix of 3 and 5</i>	Proactively requests, actively listens to, and tactfully acknowledges others' perspectives (e.g., demonstrates understanding; asks clarifying questions, provides comments or words of support)
Ignores, dismisses, or belittles the opinions and practices of others	<i>Mix of 1 and 3</i>	Accounts for the opinions and practices of relevant parties but may take too much time doing so	<i>Mix of 3 and 5</i>	Accounts for the opinions and practices of relevant parties when taking actions without compromising the mission timeline
Provides negative feedback to partners, peers, leaders, or subordinates in inappropriate times/places (e.g., embarrasses others)	<i>Mix of 1 and 3</i>	Provides feedback and acknowledges the impact but does not always anticipate consequences	<i>Mix of 3 and 5</i>	Consistently interjects or provides feedback to partners, peers, leaders, or subordinates at appropriate time and place while allowing others to save face
Uses a one-size-fits-all approach when assigning roles/tasking to partners, peers, or subordinates	<i>Mix of 1 and 3</i>	Considers strengths and weaknesses of partners, peers, or subordinates when assigning roles or tasking	<i>Mix of 3 and 5</i>	Actively looks for opportunities to utilize strengths and develop weaknesses of partners, peers, or subordinates when assigning roles or tasking
Allows others to fail unnecessarily	<i>Mix of 1 and 3</i>	Helps others who are struggling but not at the optimal time (e.g., too late, too early)	<i>Mix of 3 and 5</i>	Anticipates when others may struggle and helps at the "right" time (e.g., before trust and rapport is damaged)

To complement the generic rubric, a specific rubric was developed for each scenario based on the general attribute rubric. The scenario-specific rubric contains more concrete examples of how the attribute may manifest itself within each scenario. To develop the specific rubric, the attributes were decomposed into themes (facets). The facets reflect

the rows in the general rubric. For instance, as shown in the scenario-specific example for Struggling Student in Table 3, Empathetic was broken down into Perspective Seeking, Perspective Taking, Help at the Right Place and Time, and Aware of Others' Strengths and Weaknesses; such themes are reflected in the general rubric displayed in Table 2. Each scenario was reviewed to identify specific behaviors that respondents may describe in relation to each facet. The outlined example behaviors provided a tool for coders to resolve the situation depicted in the scenario in a manner that related to each attribute facet.

**Table 3. Examples from the Empathetic Scenario-Specific Rubric for Struggling Student**

<b>Perspective Seeking</b>	<b>Perspective Taking</b>	<b>Help at the Right Place/Time</b>	<b>Aware of Others' Strengths/Weaknesses</b>
<i>Where/How</i>	<i>Where/How</i>	<i>Where/How</i>	<i>Where/How</i>
Trying to figure out why the Soldier is shutting down	Acknowledgment that the U.S. process is new/ different from what the coalition partners usually do; it may not work for them and/or it may take time	Pulling the commander aside to discuss	Helping the student to problem solve/learn the new process
Trying to figure out why the commander is getting so upset	Acknowledging the feelings/perspective of someone else (e.g., commander) and/or threat situation	Pulling the student aside to discuss the challenges; following up with the student after the situation is resolved	Adjust instructional approach
			Mention of helping the commander learn/develop

## DATA COLLECTION OVERVIEW

To test the SJT methodology, the developed SJT items were used to collect data from 173 Soldiers currently enrolled in a PME course. Each participant completed two of the four baseline SJT scenarios, resulting in a total of 345 responses across all scenario variants.<sup>1</sup> Participants were asked to read through the first scenario, respond to the open-ended questions, and then move on to the second scenario. The data collection process took approximately 30 minutes.

Pairs of trained coders subsequently used the developed rubrics to score the responses. In total, six different coders took part in the process, with pairs being randomly assigned. Each coder rated responses independently.

## RESULTS

Findings from a sample of 345 responses across SJT variants demonstrated that agreement between raters was 97% within one point on a five-point scale. A one-way, random effects model with multiple raters (average ratings) was used to calculate the intraclass coefficients (ICCs). Intraclass coefficients ranged from .74 to .81 across attributes reflecting excellent levels of interrater reliability (Cicchetti, 1994). Collectively, these findings suggest that the SJT items and associated rubrics provide a reliable approach to assessing the likelihood of attribute-related performance under varying conditions. Tables 4 and 5 show the percent agreement and ICCs for an illustrative set of attributes. As shown in Table 4, for the depicted attributes, raters either had complete agreement or were within one point of each other for the large majority of ratings (97.43%). Importantly, agreement and interrater reliability were high no matter the attribute. Although course specific, the attributes showcased in Tables 1 and 2 relate to a range of attributes represented in the Army's LRM (U.S. Department of the Army, 2019). The method employed in this research was

<sup>1</sup> One participant only completed one scenario.



able to consistently rate the display of attributes that would be considered more cognitive in nature (e.g., sound judgment) and those considered more affective (e.g., empathetic).

**Table 4. Percent Agreement within Rater Pairs Across Attributes**

Attribute	Frequency and Percent Breakdown based on Rating Differences			
	0 point	1 point	2 points	3 points
Empathetic	171 (49.6 %)	166 (48.1%)	8 (2.3%)	0
Mature	192 (55.7%)	146 (42.3%)	7 (2.0%)	0
Situationally Aware	178 (51.6%)	155 (44.9%)	10 (2.9%)	2 (.6%)
Sound Judgment	183 (53.0%)	153 (44.3%)	9 (2.6%)	0
Disciplined	188 (54.5%)	150 (43.5%)	7 (2.0%)	0
Initiative	191 (55.4%)	144 (41.7%)	10 (2.9%)	0
<b>Average Percent</b>	<b>53.30%</b>	<b>44.13%</b>	<b>2.45%</b>	<b>.12%</b>

*Note.* N = 345 responses for each attribute. A five-point scale was used to rate SJT responses on each attribute. No ratings were 4 points apart.

**Table 5. Interrater Reliability Across Attributes**

Attribute/Competency	ICC	95% Confidence Interval		F	df1, df2	Sig.
		Lower Bound	Upper Bound			
Empathetic	.79	.74	.83	4.67	344, 345	<.001
Mature	.76	.71	.81	4.21	344, 345	<.001
Situationally Aware	.74	.68	.79	3.90	344, 345	<.001
Sound Judgment	.78	.72	.82	4.46	344, 345	<.001
Disciplined	.76	.70	.80	4.10	344, 345	<.001
Initiative	.77	.71	.81	4.27	344, 345	<.001

## DISCUSSION AND FUTURE RESEARCH

Brou et al. (2022) presented a theoretical framework to capture the complexity of Soldier development. Key to this account are the claims that attribute development in Soldiers is likely to be nonlinear and characterized by context dependencies. Accordingly, the assessment challenge is to understand, over time, changes in the likelihood of whether a Soldier will behave in accordance with an attribute under different conditions, uncovering if-then dependencies. Building on this theory, the purpose of the work presented here was to present a method capable of supporting the diagnosis of the degree to which individuals' attribute-related performance changes over time and across varying contextual demands.

In this manuscript, we outlined the use of SJTs as a viable assessment method. Leveraging an EVM-based approach, we illustrated the potential for how situational elements such as time, threat, and relationship pressures can be used to systematically probe contextual influences across situations. In addition, we showed how responses can be coded through use of rubrics to score open-ended prompts with respect to attribute-related behaviors. Results indicated that such coding can be done reliability, with good agreement between coders across a range of attributes including Empathetic, Mature, Situationally Aware, Sound Judgment, Disciplined, and Initiative. Collectively, the result of the work presented here is a methodological approach that can enable the systematic and reliable exploration of contextual dependencies.

While this work is promising, it remains the case that we have yet to explore developmental changes with respect to contextual dependencies. Thus far, there is a theoretical account (Brou et al., 2022) and a methodological toolbox as presented here, which will enable subsequent investigation. Looking forward, the next step is to use this approach to explore if, indeed, we can find contextual dependencies, and if so, whether we can uncover changes in patterns of if-then relationships as Soldiers develop over time. For instance, imagine that we look at a junior officer as they progress from OCS to IBOLC to their time as a PL. We might find that their attribute related scores progress in a universally

positive manner over time with respect to the various situations (calling for fire, teaching a security assessment process, providing medical care, and conducting a site survey). Similarly, we might find that they are better able to contend with higher pressures related to situational factors such as time, threat, and relationship. On the other hand, we may not find that this is true. Instead, it may be that some attributes progress more quickly than others, such that situationally aware progresses differently than empathetic, and this in turn may depend on the situation (e.g., providing medical care vs. helping a struggling student). Similarly, it might be the case that time pressure has a larger effect on empathetic than situationally aware, etc. In fact, there is no reason to believe that development will be uniformly positive for any individual attribute, for we should expect U-shaped functions and/or plateaus as hallmarks of change (e.g., Bartone et al., 200; Gray and Lindstedt, 2017; Toumbeva et al., 2019). Moreover, in our hypothetical case, it is worth noting that a PL is in fact a junior leader. It may be that they struggle in individual ways as they contend with situations like calling for fire versus performing a site survey, and that they are differentially affected by threat in each case. Perhaps once that PL gets to the Maneuver Captains Career Course (MCCC) they might be more capable of perceiving and acting in ways consistent with various attributes as they encounter subtle contextual influences across situations. The point is, however, that we simply do not know, which reflects the pressing need for longitudinal studies that explore Soldier attribute development across contexts.

Overall, the contribution of the work presented here is a starting point. We have presented a method that should allow us to systematically and reliably explore contextual influences over Soldier development. Future work will provide data on these issues. To extent we are successful, we will move from an analysis of “having” or “not having” an attribute to knowing when a Soldier is likely to thrive or not given the contextual complexities of their operational reality. In turn, this understanding will enable targeted instructional interventions to further support readiness in complex, large-scale, multidomain operations.

## REFERENCES

- Atzmüller, C., & Steiner, P. M. (2010). Experimental vignette studies in survey research. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 6(3), 128–138.
- Bartone, P. T., Snook, S. A., Forsythe, G. B., Lewis, P., & Bullis, R. C. (2007). Psychosocial development and leader performance of military officer cadets. *The Leadership Quarterly*, 18, 490-504.
- Brou, R. J., Allen, J. L., Ratwani, K. L., Diedrich, F. J., Toumbeva, T. H., & Flanagan, S. M. (2022). Not that straightforward: Examining and enhancing Soldier development. *Journal of Military Learning*, October 2022, 38-53.
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6(4), 284–290.
- Gray, W.D. and Lindstedt, J.K. (2017). Plateaus, Dips, and Leaps: Where to Look for Inventions and Discoveries During Skilled Performance. *Cognitive Science*, 41, 1838-1870.
- Jacobs, T. O., & Jaques, E. (1987). Leadership in complex systems. In J. Zeidner (Ed.), *Human productivity enhancement* (pp. 7-65). Praeger.
- Ludwick R., & Zeller, R. A. (2001). The factorial survey: An experimental method to replicate real world problems. *Nursing Research*, 50(2), 129-33.
- Mischel, W., & Shoda, Y. (1995). A cognitive-affective systems theory of personality: Reconceptualizing situations, dispositions, dynamics, and invariance in personality structures. *Psychological Review*, 102, 246-268.
- Motowidlo, S. J., Hooper, A. C., & Jackson, H. L. (2006). Implicit policies about relations between personality traits and behavioral effectiveness in situational judgment items. *Journal of Applied Psychology*, 91(4), 749-761.

- Motowidlo, S. J., Borman, W. C., & Schmit, M. J. (1997). A theory of individual differences in task and contextual performance. *Human Performance*, 10, 71–83.
- Pierce, C. A., Aguinis, H., & Adams, S. K. R. (2000). Effects of a dissolved workplace romance and rater characteristics on responses to a sexual harassment accusation. *Academy of Management Journal*, 43(5), 869–880.
- Podsakoff, N. P., Whiting, S. W., Podsakoff, P. M., & Mishra, P. (2011). Effects of organizational citizenship behaviors on selection decisions in employment interviews. *Journal of Applied Psychology*, 96(2), 310–326.
- Raaijmakers, A., Vermeulen, P., Meeus, M. & Zietsma, C. (2014). I need time! Exploring pathways to compliance under institutional complexity. *The Academy of Management Journal*, 58(1), 85-110.
- Rossi, P. and Anderson, A. (1982) The Factorial Survey Approach: An Introduction. In P. Rossi and S. Nock (Eds.), *Measuring Social Judgments: The Factorial Survey Approach*. Sage, Beverly Hills, CA, 1-25.
- Rose, T. (2016). *The end of average: Unlocking our potential by embracing what makes us different*. Harper Collins.
- Shoda, Y., Cervone, D. & Downey, G. (2007). *Persons in context: Building a science of the individual*. The Guilford Press.
- Thelen, E., & Smith, L. B. (1994). A dynamic systems approach to the development of cognition and action. MIT Press.
- Thurman, Q. C., Lam, J. A., & Rossi, P. H. (1988). Sorting out the cuckoo's nest: A factorial survey applied to the study of popular conceptions of mental illness. *Sociological Quarterly*, 29, 565-588.
- Toumbeva, T.H., Diedrich, F.J., Flanagan, S.M., Naber, A., Reynolds, K., Shenberger-Trujillo, J., Cummings, C., Ratwani, K.L., Ubillus, G., Nocker, C., Gerard, C.M., Uhl, E.R., & Tucker, J.S. (2019). *Assessing character in U.S. Army Initial Entry Training* (ARI Technical Report 1373). U.S. Army Research Institute for the Behavioral and Social Sciences. (DTIC No. AD1077839)
- U.S. Department of the Army (2022). *Operations* (FM 3-0). Retrieved March 16, 2022, from [https://armypubs.army.mil/epubs/DR\\_pubs/DR\\_a/ARN36290-FM\\_3-0-000-WEB-2.pdf](https://armypubs.army.mil/epubs/DR_pubs/DR_a/ARN36290-FM_3-0-000-WEB-2.pdf)
- U.S. Department of the Army (2019). *Army Leadership and the Profession* (ADP 6-22). Retrieved March 16, 2022, from [https://armypubs.army.mil/epubs/DR\\_pubs/DR\\_a/ARN20039-ADP\\_6-22-001-WEB-0.pdf](https://armypubs.army.mil/epubs/DR_pubs/DR_a/ARN20039-ADP_6-22-001-WEB-0.pdf)

## DISCLAIMER

The research described herein was sponsored by the U.S. Army Research Institute for the Behavioral and Social Sciences, Department of the Army (Contract No. W911NF220031). The views expressed in this presentation are those of the author and do not reflect the official policy or position of the Department of the Army, DOD, or the U.S. Government.

Controlled by: DAPE-ARI CUI Category: Unclassified Distribution Statement: A POC: Randy J. Brou, 706-575-8264
--