UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

CONSTRUCT TRANSPARENCY AND THE PSYCHOMETRIC PROPERTIES OF A

MULTIDIMENSIONAL SITUATIONAL JUDGMENT TEST

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

OLIVIA D. COOPER
Norman, Oklahoma
2014

CONSTRUCT TRANSPARENCY AND THE PSYCHOMETRIC PROPERTIES OF A
MULTIDIMENSIONAL SITUATIONAL JUDGMENT TEST

A DISSERTATION APPROVED FOR THE
DEPARTMENT OF PSYCHOLOGY

BY

_____
Dr. Eric Day, Chair

_____
Dr. Michael Buckley

_____
Dr. Shane Connelly

_____
Dr. Michael Mumford

_____
Dr. Robert Terry

Table of Contents

List of Tables

Abstract

Reported internal consistencies of SJTs are often low, and empirical examination of the structural properties of SJT scores are rarely examined. This paper addressed the need for an empirical investigation of factors that affect the validity evidence of SJT scores by examining the extent to which construct transparency influenced the internal structure of a multidimensional SJT. Methods for increasing construct transparency that have previously been applied in the assessment center literature were adapted for use in the present study. Two conditions—experimental versus control—were used to investigate the influence of construct transparency on a recently developed multidimensional SJT designed to measure multiple sensemaking skills important to leadership. Various psychometric properties were examined and compared for the two conditions using a sample of 383 participants. The study found that there were no significant differences between the two conditions for either estimates of scale intercorrelations or scale reliabilities, and the patterns of correlations with external scales were very similar. Participants in the construct transparency condition received higher scores on all scales compared to participants in the control condition, and evidence of structural validity slightly favored the scores of the construct transparency condition. However, half of the scales lacked supportive validity evidence. The results of this study are discussed in terms of the extent to which multidimensional SJTs of sensemaking skills may be dependent upon case-based knowledge and experience as well as the usefulness of making the content of multidimensional SJTs transparent to test takers.

Construct Transparency and the Psychometric Properties of a Multidimensional

Situational Judgment Test

Situational judgment tests (SJTs) are often used as low-fidelity simulations that present respondents with situations and possible courses of actions to handle those situations (Whetzel & McDaniel, 2009). SJTs have been established as valid predictors of job performance (Christian, Edwards, & Bradley, 2010). However, researchers are increasingly concerned with their features (i.e., item stems, response options, and response instructions) and with determining what constructs they measure that predict job performance (Christian et al., 2010; Sackett & Lievens, 2008). One explanation for why SJTs predict job performance is the theory of behavioral consistency (Motowidlo, Dunnette, & Carter, 1990). By obtaining an example of someone's current behavior, one can predict how someone will behave in the future. However, despite evidence of predictive validity, reported estimates of internal consistency for SJT scores are frequently low and the structural validity of scores is rarely examined. Some researchers seem to downplay the importance of SJT internal structure, especially in terms of how low estimates of internal consistency are to be expected due to the influence of multiple constructs (Whetzel & McDaniel, 2009). However, similar to issues regarding the structure of assessment center (AC) scores (Arthur, Day, & Woehr, 2008; Howard, 2008; Lance, 2008), concerns regarding the internal structure of SJT scores are important for better understanding their psychometric properties and ultimately their validity.

Building upon theory and empirical evidence in the AC literature, the purpose of the present study was to examine the extent to which construct transparency influenced

the structural properties of a multidimensional SJT, particularly in terms of the internal consistency and factor structure of scale scores. Methods for increasing transparency that have been used in the AC literature were adapted in this study. Presently, few researchers have applied construct transparency outside of AC research. The present study used two conditions: (1) control and (2) construct transparency. The SJT was a revised version of a recently developed multidimensional SJT designed to measure multiple sensemaking skills important to leadership. This SJT was comprised of several independent scenarios each followed by several questions that were each designed to measure a different sensemaking skill. Thus, the same sensemaking skills were measured across different scenarios. Previous research (Cooper et al., 2013) with this SJT showed mixed support for its internal structure. Specifically, although confirmatory factor analyses (CFAs) showed support for the multidimensional nature of the scores, support for the specific a priori scales was not found for every scale. Moreover, estimates of scale internal consistency were low, ranging from .32 to .68.

This paper first discusses SJTs and the current need to examine the factors affecting the construct-related validity of SJT scores. Then, the literature on construct transparency is discussed as well as the logic for extending this research and applying it to SJTs. Next, the literature regarding leadership and sensemaking is reviewed as well as the constructs measured in the current SJT that were drawn from it, and the relevance of transparency for such constructs is discussed. Lastly, the methodology used to examine the influence of construct transparency applied to SJTs is presented.

Situational Judgment Tests

SJTs can be defined as low fidelity simulations that present respondents with scenarios involving a challenge, problem, or dilemma that require respondents to answer questions related to what they think should be done to deal with the situation (Ployhart & MacKenzie, 2010). The formatting of SJTs may vary; however the basic content consists of item stems and response options. The item stems of SJTs present respondents with the challenging scenario while the response options present the respondents with options as to how they should deal with the scenario.

Methods for choosing response options can vary. Respondents may be instructed to select one response option, to select the best and worst response options, to choose what they would most and least likely do, to rank order all response options, or to rate response options on a Likert scale (Motowidlo, Crook, Kell, & Naemi, 2009; Ployhart & MacKenzie, 2011). The different types of response instructions have been shown to affect the construct-related and criterion-related validity of test scores (McDaniel, Hartman, Whetzel, & Grubb, 2007; Ployhart & Ehrhart, 2003; Ployhart & MacKenzie, 2011). For example, behavioral tendency questions, which ask respondents to select response options that reflect what they would (or would not do) in a given situation, have been shown to have higher correlations than knowledge instructions with personality constructs, and to be more susceptible to "faking" (i.e., selection of response options that are more socially desirable; Whetzel & McDaniel, 2009). On the other hand, knowledge instructions, which ask respondents to select what they consider to be the most (or least) effective response for a given situation, have been shown to have higher correlations than behavioral tendency instructions with cognitive ability, and to

be less susceptible to faking (Whetzel & McDaniel, 2009). SJTs that use knowledge instructions have also been shown to have stronger criterion-related validity than behavioral tendency instructions (Ployhart & MacKenzie, 2011).

Thus, it is commonly recommended that knowledge instructions be used when the focus of measurement is on cognitive or skill-based constructs (i.e., measures of maximal performance) (Ployhart & MacKenzie, 2011). Therefore, knowledge instructions were used in the present research given that this research was focused on the use of SJTs for the measurement of leadership sensemaking "skills" with real-world applications related to the assessment of skill proficiency in mind (e.g., training evaluation). Behavioral tendency instructions are more appropriate for the measurement of personality-based constructs and the prediction of typical performance (Ployhart & MacKenzie, 2011).

Situational Judgment Test: Method or Construct?

Similar to the history of ACs and their exercises, researchers have debated issues related to method versus construct variance in SJT scores. In particular, researchers have debated whether SJTs measure a single construct of situational judgment or whether SJTs are a measurement method that can be used to measure a variety of constructs. Evidence suggests that SJTs are measurement method that can be used to measure a variety of constructs (Christian et al., 2010). In addition, many researchers have found that a single general factor accounts for little of the variance in SJT scores (Schmitt & Chan, 2006). However, high performance on SJTs may also be partly attributed to experience or other third variables such as education or motivation (Calfee, 1993; Jensen, 1993; McClelland, 1993).

When viewed as a measurement method, SJTs have been shown to measure a number of constructs that are valid predictors of job performance (Christian et al., 2010). Applied social skills is one of the construct domains frequently measured with SJTs (Christian et al., 2010). Applied social skills encompass a broad category that includes interpersonal, teamwork, and leadership skills. The predominance of research on applied social skills is unsurprising due to the history between assessment and applied social skills that is reflected within the AC literature on managerial potential and leadership (Arthur & Day, 2011; Thornton & Byham, 1982).

Christian et al. emphasized the importance of reporting construct-level scores (as opposed to overall SJT scores) in order to provide evidence as to what constructs reflected in SJT scores predict job performance. Although leadership skills appear to be measured frequently in the SJT literature, researchers often do not explicitly state the types of leadership skills being measured. Hence, the present study attempted to clearly examine specific leadership skills within a multidimensional SJT.

Reliability and Structural Validity of SJTs

Although SJTs are commonly used because they are valid predictors of job performance and viable alternatives to high-fidelity behavioral simulations, they are not without their limitations. One major oversight in the SJT literature appears to be the lack of evidence supporting the internal psychometric properties. In this vein, a meta-analysis by McDaniel, Morgeson, Finnegan, Campion, and Braverman (2001) found that the reliability coefficients from 34 SJTs ranged from .43 to .94 ($M = .77$, $SD = .11$). However, the authors failed to indicate the type of reliability (e.g., internal consistency, test-retest, or parallel forms). This is something that is important to consider because

different types of reliability estimates vary substantially for SJTs. For example, Clause, Mullins, Nee, Pulakos, and Schmitt (1998) conducted a study with parallel forms of SJTs. The different forms of the SJTs had internal consistency coefficients ranging from .56 to .60 while the coefficients of congruence for the same forms of the SJTs ranged from .78 to .94. Also, Ployhart and Ehrhart (2003) reported internal consistencies ranging from .32 to .73 with corresponding test-retest estimates ranging from .20 to .92 (across six different response option formats of the same SJT).

On one hand, research on the reliability of SJT scores demonstrates the necessity for researchers to indicate the type of reliability coefficient used within their study. On the other hand, some researchers have argued that representing the reliability of SJTs with estimates of internal consistency may not be appropriate due to the fact that many SJTs are multidimensional in nature (Whetzel & McDaniel, 2009). Likewise, if evidence suggests that a SJT is homogeneous (i.e., those developed with a construct-oriented approach), estimates of internal consistency should be calculated and reported. If a SJT is multidimensional in nature, then it is important to clearly identify what the different constructs are. Accordingly, it is incumbent upon test developers and researchers to examine the reliability of SJT scales as well as the structural validity of the test scores (i.e., factorial validities). Subsequent validation efforts could then focus on testing hypotheses regarding how different dimensions measured by a multidimensional SJT are differentially related to other variables of interest such as job performance.

Transparency

Transparency can be defined as informing participants about the constructs measured by a test and informing them of the behaviors pertaining to the constructs (Kleinmann, Kuptsch, & Koller, 1996). In the AC literature, researchers have argued that transparency is an important issue because individuals differ in their ability to accurately recognize the constructs being assessed in each AC exercise, and in general participants' recognition of the constructs is poor (Kleinmann, 1993). This lack of transparency partly explains why behavior may be inconsistent across AC exercises as participants may believe that each exercise targets different constructs that involve different behaviors (Kolk, Born, & van der Flier, 2003). Researchers find that informing participants of the constructs being targeted leads to stronger evidence of validity for ACs. Specifically, participants behave more consistently across exercises because of their knowledge of the constructs (Kleinmann et al., 1996), and correlations between construct ratings across different exercises (i.e., monotrait-hetromethod correlations) are stronger when participants realize that the exercises target the same constructs (Kleinmann, 1993).

Outside the AC literature, research on construct transparency has primarily focused on the fakeablity of employment interviews (Allen, Facteau, & Facteau, 2004; Maurer, Solamon, & Lippstreu, 2008). Evidence from these studies suggests that informing participants of the constructs being measured does not result in higher scores due to faking or behaving in a socially desirable manner (Allen et al., 2004; Maurer et al., 2008).

Considering the similarities between SJTs and ACs as simulation-based assessment methods involving the measurement of a common set of constructs across multiple scenarios, it is plausible that transparency might have an influence on the psychometric quality of SJTs as it does with ACs. Similar to how transparency has influenced mono-trait correlations in ACs, one might expect that the correlations between items of the same SJT scale should increase because the participants are made aware that the same constructs are being measured in each of the scenarios (Kleinmann, 1993). Comparable to how transparency influences consistency in behavioral ratings of ACs, participants should be more consistent in how they respond across different scenarios if they are informed that the same constructs are measured in all of the SJT scenarios (Kolk et al., 2003). If the results previously found in transparency and ACs are further extended, it can be argued that informing participants of the constructs that are being measured will decrease the influence of extraneous factors on scores (Kolk et al., 2003; Smith-Jentsch, 2007). Transparency may be of particular importance for the constructs measured in the current multidimensional SJT given the nuanced, complex, and ambiguous circumstances in which these constructs are especially relevant.

Leadership and Sensemaking

Leadership can be defined as "the process of influencing others to understand and agree about what needs to be done and how to do it, and the process of facilitating individual and collective efforts to accomplish shared objectives" (Yukl, 2006, p. 8). Leadership is important to organizations because leaders can help build commitment, and the actions of leaders have important consequences for followers, organizations, and stakeholders (Yukl, 2006).

Situations and individual differences influence a person's behavior (Vroom & Jago, 2007), and so they should both be considered when assessing a leader's effectiveness. Leaders are frequently confronted with problems, including complex, multifaceted social situations. Thus, the construct of sensemaking seems directly relevant to leadership, because of the nuances and complexities within many social situations (Browning & McNamee, 2012).

Sensemaking can be defined as a "complex cognition that occurs when people are presented with ambiguous, high-stakes events" (Mumford et al., 2008, p. 317). When leaders are confronted with crisis situations, they resort to using mental models or complex cognitive structures that allow them to better handle and understand situations (Mumford et al., 2008; Weick, Sutcliffe, & Obstfeld, 2005). In models of sensemaking (Jensen, 2009; Mumford et al., 2008), the overarching idea is that different attributes of the situation affect the leader's impression of and the approaches taken to deal with the situation (Rouleau & Balogun, 2011).

The present research involved the application of SJTs as evaluation tools for new training and development programs for junior officers in the U.S. Army. The U.S. Army is increasingly concerned with meeting the new challenges faced by current Army leaders who must balance roles that include peace-keeping, humanitarian, and warfare operations, and which often involve collaborations among multiple branches of the U. S. military, military forces of foreign allies, as well as local forces and community leaders. To evaluate programs designed to train second lieutenants on building successful working relationships both within their Platoons as well as when involved in joint force operations, a low-fidelity multidimensional SJT was proposed as

one potentially viable criterion measure. Consequently, six specific skills were drawn

from the scientific literature on sensemaking and handling socially complex situations

involving the interests of multiple constituencies in which ethical dilemmas are

common: (1) cognitive reappraisal, (2) expressive suppression, (3) considering other

perspectives, (4) identification of key causes, (5) forecasting, and (6) ethical decision

making. These six skills overlap somewhat as each represents a different aspect of

sensemaking.

Sensemaking Leadership Skills in the Current Study

Emotion regulation refers to attempts to influence the experience and expression

of emotions (Gross, 1998). Emotion regulation has been shown to affect leadership for

team building, resolving conflicts, influencing others, and making decisions (Bono,

Foldes, Vinson, & Muros, 2007; George & Zhou, 2007; Hauck et al., 2010). There are

multiple types of emotion regulation strategies that can be used effectively with practice

and training to influence emotions (Lopes et al., 2011). The current SJT measured the

two specific strategies of cognitive reappraisal and expressive suppression. *Cognitive*

reappraisal was defined as a preemptive cognitive strategy whereby one reevaluates an

emotional situation to reduce the experience and emotional impact in order to accomplish

goals (Gross, 1998, 2001; Memedovic, Grisham, Denson, & Moulds, 2010; Yurtsever,

2008). Cognitive reappraisal allows leaders to turn an emotionally laden event into

something that is less emotion provoking so that they may continue to pursue their goals

in a more "rational" manner. *Expressive suppression* was defined as preventing the

expression of a felt emotion and behaving in such a way that others could not know what

one is feeling in order to accomplish goals (Brown & Mitchell, 2010; Gooty, Connelly,

Griffith, & Gupta, 2010; Gross, 1998, 2001; Heilman, Crişan, Houser, Miclea, & Miu,

10

2010; van't Wout, Chang, & Sanfey, 2010). It is important to note that research has shown that when one must display emotions that are not congruent with what is truly felt, emotional labor may result (Diefendorff & Richard, 2008; Gosserand & Diefendorff, 2005). Although expressive suppression can help leaders who have been confronted with a disappointing situation prevent expressing that disappointment to their followers so that everyone can stay on task, suppressing emotions may also result in negative repercussions such as job dissatisfaction and emotional exhaustion (Hülsheger, Alberts, Feinholdt, & Lang, 2012; Pugh, Groth, & Hennig-Thurau, 2011).

Considering other perspectives was defined as considering the perceptions, concerns, goals, and responsibilities of others who are affected by a situation (Brown & Mitchell, 2010; Kligyte, Marcy, Sevier, Godfrey, & Mumford, 2008; Kligyte, Marcy, Waples, et al., 2008; McEntire, Dailey, Osburn, & Mumford, 2006; Mumford et al., 2006; Munro, Bore, & Powis, 2005; Watley & May, 2004). Considering other perspectives allows leaders to identify and consider a variety of relevant information when making sense of a situation (Hauck et al., 2010; Kligyte, Marcy, Sevier, et al., 2008; Kligyte, Marcy, Waples, et al., 2008). Furthermore, considering how one's decisions impact others results in behavioral intentions that show greater ethicality (Watley & May, 2004).

Identification of key causes was defined as identifying causes or factors that influence important elements of a situation (Antes et al., 2010; Dutton & Duncan, 1987; Isenberg, 1986; Marcy & Mumford, 2007, 2010; Marta, Leritz, & Mumford, 2005; Osburn & Mumford, 2006; Van Der Linden, Sonnentag, Frese, & Van Dyck, 2001). Identifying the key causes of a situation helps leaders develop plans and solutions to problems (Dutton & Duncan, 1987; Isenberg, 1986; Marcy & Mumford, 2007; Osburn & Mumford, 2006; Van Der Linden et al., 2001), results in more effective planning (Isenberg, 1986; Marcy

11

& Mumford, 2007), and influences group performance and behavior (Marta et al., 2005).

Forecasting was defined as predicting likely outcomes or consequences of implementing an idea within a particular setting (Jones, 1991; Kligyte, Marcy, Waples, et al., 2008; Lonergan, Scott, & Mumford, 2004; Marcy & Mumford, 2007; Mumford et al., 2006, 2008; Mumford, Friedrich, Caughron, & Byrne, 2007; Shipman, Byrne, & Mumford, 2010). Forecasting allows leaders to explore a variety of decision options, each with its own consequence if implemented (Mumford et al., 2008). Forecasting has also been shown to share a positive relationship with ethical decision making (Mumford et al., 2006). Previous research with the present SJT (Cooper et al., 2013) operationalized forecasting in terms of thinking about downstream versus short-term consequences. However because the scale lacked supportive validity evidence, the literature on forecasting was carefully reconsidered. It was then decided to focus on extensiveness in forecasts in the present SJT. Extensiveness in forecasts refers to comprehensive and complete consideration of a situation and potential outcomes (Byrne, Shipman, & Mumford, 2010; Saffo, 2007; Shipman et al., 2010).

Ethical decision making was defined as a process by which an individual takes into consideration the situation, potential responses to the problem, and then chooses a course of action that minimizes harm yet is consistent with ethical guidelines by considering informal norms in addition to rules, policies, and guidelines (Antes et al., 2007; Brown & Mitchell, 2010; Carnevale, Inbar, & Lerner, 2011; Kligyte, Marcy, Sevier, et al., 2008; Mumford et al., 2008). Leaders must simultaneously consider a number of factors operating within a given situation and weighing the possible outcomes to make the most ethical and socially acceptable decision (Antes et al., 2007; Kligyte, Marcy, Sevier, et

al., 2008). And because leaders are often placed in ambiguous situations, their actions and decisions have repercussions for their followers.

Study Overview, Hypotheses, and Research Question

As previously discussed, the sensemaking leadership skills measured in the current SJT are critical to ambiguous, social situations where no clear right or wrong course of action is apparent. Because of the nuances of these complex situations, individuals tend not to be aware of or understand the common features that underlie these situations. Consequently, there is inconsistency in how individuals respond to such situations because they approach each situation based more on the nuances of the situation rather than the important features that are common across the situations. The core supposition of the present research was that providing individuals with a basic understanding of the skills most needed to effectively address these complex situations would lead to more consistency in how individuals approached such situations. In other words, construct transparency is particularly relevant for simulation-based tests that involve socially complex situations and are designed to measure multiple constructs.

Therefore, the purpose of the present study was to examine the influence of construct transparency on the structural properties of a multidimensional SJT by using two conditions: (1) a control and (2) a construct transparency. In the AC literature, construct transparency has been shown to increase behavioral consistency across exercises as evidenced by stronger monotrait-hetromethod correlations. And specifically, when evidence of structural validity is examined using CFA, there should be more support for a six-factor model in the construct transparency condition compared to the control condition. Therefore, the following hypotheses were examined.

13

*Hypothesis 1***:** SJT scale estimates of internal consistency will be higher

for the scores from the construct transparency condition compared to

those from the control condition.

*Hypothesis 2***:** Evidence of structural validity will be stronger for the

scores from the construct transparency condition compared to those from

the control condition.

In a previous study with the original version of the present SJT (Cooper et al.,

2013), correlations between the SJT scale scores and a wide variety of cognitive and

non-cognitive measures were examined. The three external measures that yielded the

most consistent or strongest relationships with SJT scales were (a) military background,

(b) general intelligence, and (c) verbal reasoning. As such, measures for these variables

were included in the present study to examine whether correlations with SJT scores

differed as a function of the transparency manipulation. Thus, the following research

question was examined.

Research Question: Will the correlations between the SJT scale scores

and the external measures of military background, general intelligence,

and verbal reasoning differ as a function of the transparency

manipulation?

Method

Participants

Participants were 402 undergraduate students from the University of Oklahoma

(OU). All students received credit towards a research requirement in a psychology

course. Because eight participants did not follow directions or had more than 10% of

the SJT data missing, their data were removed from analyses. Data for an additional 11 participants were removed from analyses because their scores on a basic test of recall indicated that they were not paying attention. The recall test consisted of 10 items and was constructed to assess participants' basic understanding of the military structure and content presented in the SJT. The items asked questions related to the role they played in the SJT and the types of leadership situations they were required to handle within the SJT scenarios. Participants were presented with three response options and they had to indicate which option was correct. Participants who scored a six or below were excluded from further analyses, leaving a final sample of 383 participants. Of the 383 participants, 60 had a military background which was defined as participating in a JROTC or ROTC program, serving in the military, or having at least one parent or guardian who served in the military. Participants were predominately female (74.4%), and the mean age was 18.58 years old ($SD = 1.15$). Sixty-eight percent of the participants self-reported their ethnicities as White, 9.7% as Asian, 7.0% as Black, and 4.2% as Latino.

Transparency Manipulation

The transparency manipulation was a PowerPoint presentation with recorded audio information that provided participants with definitions of the six sensemaking leadership skills used in the current study along with effective and ineffective examples for each skill. Appendix B shows the content of the transparency manipulation presentation. One hundred ninety-one participants (i.e., those in the construct transparency condition) viewed this presentation while the other 192 participants (i.e.,

those in the control condition) did not receive any information regarding the skills of interest. Participants were randomly assigned to one of the two conditions.

Procedures

All participants were told that the study was designed to examine differences in how people make decisions in various leadership situations. Participants first completed an informed consent form followed by a presentation that provided them with some basic information about leading a Platoon and the role they would assume in the SJT as a second lieutenant in the U.S. Army. Following this presentation, the transparency manipulation occurred. Approximately half of the participants immediately completed the SJT while the other half of the participants viewed the transparency manipulation presentation before completing the SJT. After completing the SJT, all participants completed a packet containing various timed (i.e., verbal reasoning, a test of recall, and manipulation check) and untimed (i.e., demographic information and military background) measures. After completing the packet, participants were thanked for their participation, given a form that provided more details about the purpose of the study, and dismissed.

Measures

SJT. The following provides a general description of the format of the SJT. Cooper et al. (2012) provides a more detailed review of the format and development process. The current SJT is a modified version of the SJT discussed in Cooper et al. (2012). The current SJT was comprised of eight self-contained scenarios. Each scenario contained an initial paragraph that presented a socially complex situation happening within the Platoon. To reduce confusion or response contamination between subsequent

scenarios, no two initial paragraphs involved the same central issue or personnel. The issues for each scenario are listed below.

1. An increase in hostilities by and against the soldiers

2. An increase in complaints regarding the carelessness of a squad of soldiers

3. The news of an extended deployment for the Platoon

4. The effect of combat stress on the soldiers

5. Substance abuse issues within the Platoon

6. Escalating conflicts between a Sergeant and a soldier within the Platoon

7. Problems regarding the leadership decisions of a Sergeant for his squad

8. Redeployment issues affecting the readiness of the Platoon

A team of four graduate students and two professors in the Industrial/Organizational Psychology program at OU developed the content of the test, with feedback from Army stakeholders. Before beginning the content development process, the students were trained and gained knowledge of the targeted skills from reading the scientific literature on sensemaking, the Army leadership field manual, and other relevant literature published by the U.S. Army. The SJT scenarios were developed by collapsing ideas and situations into common themes by reviewing approximately 60 transcribed interviews with experienced Army officers, recently developed officer training materials, and the scientific literature related to deployment and pre- and post-deployment issues affecting soldiers.

Individual graduate students independently developed initial scenario content, then reviewed and revised each other's content before receiving feedback from the professors. Based on feedback from the professors, further modifications were made. Finally, content was again presented to the professors and other graduate students for further review and modifications. This process was followed on a scenario-by-scenario basis with feedback from Army stakeholders before the entire test was finalized. In total, the Cooper et al. (2012) SJT contained 36 items with six items per scale and six response options per item. This SJT was modified for the current study by examining scale and item characteristics (e.g., response option discrimination values and substantive-content validity information) to reduce the response options from six to three and to identify response options in need of revision. Two additional scenarios, each with six items and three response options per item, were added to the modified SJT.

Each scenario consisted of six independent items, each item containing three response options, and each item targeting a different sensemaking skill. The response options for each item targeted a different level of effectiveness: low, medium, and high. The high response options generally focused on broader issues, picked up on nuances within the scenario, or reduced the potential negative effect of the given situation. Low responses generally focused on narrow or trivial issues, missed the nuances of the scenario, or increased or ignored the negative effects of the situation. The medium response options fell between the high and low options. To maintain independence, items and their response options were written such that responses to a particular item would not influence the responses made to later items within or following the scenario.

Participants were instructed to choose the best and worst options for each item. Selecting the high option as "best" and the low option for "worst" yielded +1 point each, selecting the low option as "best" and the high option for "worst" yielded –1 point each, and selecting the medium option yielded zero points. Thus, item scores could range from –2 to +2 points, with scale scores potentially ranging from –16 to +16 points.

Manipulation check. The manipulation check was designed to determine which participants were able to identify the six specific skills of interest in the SJT. Participants were presented with the correct skill set and three distractor sets. Participants were instructed to select the skill set that they believed the SJT was designed to measure. The purpose of this measure was to ensure that participants in the construct transparency condition were in fact better able to identify the constructs measured by the SJT.

Military background. The military background measure consisted of five items and was constructed to assess participants' connection to the military. The items asked demographic information related to the respondents' and their parents' involvement in military programs (i.e., JROTC and ROTC) and serving in the military. In total, 60 participants (45 females) denoted having some kind of military background.

General intelligence. General intelligence was measured by asking participants to self-report their ACT scores.

Verbal reasoning. Verbal reasoning was measured using the verbal reasoning component of the Employee Aptitude Survey (EAS; Ruch & Ruch, 1983). In the test, a series of facts and five conclusions are presented for the participant. The participant

must then indicate whether, based on the facts given, the conclusion is true, false, or uncertain. Coefficient alpha for the EAS was .74.

Results

Manipulation Check

To check if participants in the construct transparency condition were able to better identify the six specific skills of interest in the SJT, the selected skill sets for each condition were examined. Results showed a significant higher percentage of participants in the construct transparency condition (82.5%) selected the correct skill set compared to that of the control condition (13.4%), $\chi^2(3, N = 375) = 180.31, p < .01$.

Scale Descriptive Statistics

Scale means, standard deviations, and ranges were examined for each of the six scales and are shown in Table 1. For both conditions, scale means and standard deviations differed across the six scales. The cognitive reappraisal and identification of key causes scales had the highest means and lowest standard deviations, while the other four scales varied in terms of their means, standard deviations, and ranges. Specifically, means were lower for considering other perspectives and forecasting, and expressive suppression yielded the largest standard deviation. For every scale, scores in the control condition were lower than those of the construct transparency condition with statistically significant differences for cognitive reappraisal, $t(381) = -2.24, p < .05, d = -0.23$, considering other perspectives, $t(381) = -2.98, p < .01, d = -0.31$, forecasting, $t(381) = -2.06, p < .05, d = -0.21$, and expressive suppression, $t(381) = -3.37, p < .01, d = -0.35$. Although scores were higher in the construct transparency condition, standard deviations were comparable between the two conditions for every scale.

Scale Intercorrelations

Scale intercorrelations were also examined and are shown in Table 2. For both conditions, most scale intercorrelations were weakly to moderately correlated. Specifically, for the control condition, all scales were significantly intercorrelated with the exception of identification of key causes when correlated with cognitive reappraisal, $r = .08$, $p > .05$, and considering other perspectives, $r = .14$, $p > .05$. The largest scale intercorrelations were between forecasting and ethical decision making, $r = .48$, $p < .01$, and between cognitive reappraisal and expressive suppression, $r = .42$, $p < .01$. For the construct transparency condition, all scales were significantly correlated with the exception of cognitive reappraisal correlated with identification of key causes, $r = .10$, $p > .05$, and forecasting, $r = .14$, $p > .05$. The largest scale intercorrelation was between the cognitive reappraisal and expressive suppression scales, $r = .52$, $p < .01$. Nevertheless, in general the pattern and magnitude of the correlations were similar between the two conditions. In order to determine if the correlations significantly differed between the two conditions, scale intercorrelations were compared. Across the 15 comparisons, there were no significant differences between the two conditions for any of the scale intercorrelations.

Internal Consistencies

Hypothesis 1 predicted that SJT scale estimates of internal consistency would be higher for the scores from the construct transparency condition compared to those from the control condition. Spearman-Brown odd-even split-half reliability coefficients for the six scales are shown in Table 3. The internal consistency reliabilities varied greatly across the six scales for both conditions (from .08 to .86). There seemed to be a pattern

in that for both conditions, the considering other perspectives (reliability = .64 for both control and construct transparency conditions) and expressive suppression (control condition: reliability = .86; construct transparency condition: reliability = .80) scales yielded the highest internal consistencies than the other sensemaking scales. Only the expressive suppression scale yielded an internal consistency that would be considered acceptable (i.e., .70 for program evaluation or research purposes). In general, internal consistencies were comparable and the confidence intervals for every scale between the two conditions overlapped considerably. Thus, Hypothesis 1 was not supported.

Scale Structural Validity

Hypothesis 2 predicted that evidence of structural validity would be stronger for the scores from the construct transparency condition compared to those from the control condition. The SJT structure for both the control and construct transparency conditions were analyzed with structural equation modeling using SAS (SAS Institute Inc., 2008). Several possible models were tested using CFAs with maximum likelihood estimation and Proc CALIS. The fit indices of all models and relevant model comparisons are shown in Tables 4 and 5. Three models were examined a priori: (1) a one-factor model with all items loaded onto a single latent factor, (2) a two-factor correlated model with the cognitive reappraisal and expressive suppression items loaded onto an emotions latent factor and the considering other perspectives, identification of key causes, forecasting, and ethical decision making items loaded onto a sensemaking latent factor, and (3) a six-factor correlated model with items for each skill loaded onto their originally conceived latent factors. For the control condition, the six-factor model resulted in a non-positive definite matrix as the standardized correlation between the

forecasting and ethical decision making factors was greater than one. Although the χ^2 difference tests and the fit indices suggested a better fit for this model compared to both the one- and two- factor models, because of the non-positive definite matrix, accuracy of the fit indices are questionable.

For the construct transparency condition the same three a priori models were compared and results revealed the six-factor model fit the data best compared to both the one-factor and two-factor models. However, the fit indices, the CFI specifically, suggested that the six-factor model was not a good fitting model, $\chi^2(1065) = 11,485.62$, $p < .05$, CFI = .67, RMSR = .07, RMSEA = .05. Given the poor fitting models for both the control and construct transparency conditions, a fourth model was additionally examined consisting of three correlated factors such that the items from the forecasting, identification of key causes, and ethical decision making scales were excluded due to their extremely low reliabilities (i.e., low item intercorrelations). The fit of the three-factor model was then compared to the fit of its corresponding one-factor model.

As shown in Tables 4 and 5, fit indices as well as the χ^2 difference test indicated that the three-factor model fit the data better than the one-factor model for both the control, $\triangle \chi^2(3) = 193.58$, $p < .01$, and construct transparency, $\triangle \chi^2(3) = 128.39$, $p < .01$, conditions. However, the CFI values for both conditions indicated that the three-factor model did not fit the data (i.e., CFI values < .90). Although the results from the CFAs indicated that the correlated three-factor model did not fit the data well, it is important to note that the fit of the three-factor correlated model for the transparency condition almost reached levels of acceptability (CFI = .89, RMSR = .07, RMSEA = .04) whereas the fit of the control condition was not as close (CFI = .81, RMSR = .08,

RMSEA = .05). Thus, the data showed weak support for Hypothesis 2 which stated that evidence of structural validity would be stronger for the scores from the construct transparency condition compared to those from the control condition.

Correlations between SJT Scales and External Measures

A research question was proposed to investigate if correlations between the SJT scale scores and the external measures of military background, general intelligence, and verbal reasoning differed as a function of the transparency manipulation. The correlations and corresponding comparisons between the two conditions are shown in Table 6. In the control condition, all scales except expressive suppression were significantly positively correlated with general intelligence, while the same was true in the construct transparency condition for all scales except cognitive reappraisal and considering other perspectives. Similarly for verbal reasoning, all scales in the control condition were significantly correlated while for the construct transparency condition, all scales except expressive suppression were positively correlated. Overall, the patterns were similar between conditions such that the SJT scales were positively correlated with general intelligence and verbal reasoning for both the control and construct transparency conditions while the pattern for military background indicated that it was not correlated with the SJT scales in either condition. A series of comparisons of correlations was conducted between the two groups, and out of 18 comparisons, only one significant difference was found. These results indicated that correlations between the SJT scales and external measures did not differ as a function of the transparency manipulation.

Discussion

Whereas previous studies have rarely examined the structural properties of SJTs (Ployhart & Weekley, 2006), this paper addressed the need for an empirical investigation of factors that affect the validity evidence of SJTs. The present study examined the extent to which construct transparency influenced the internal structure of a multidimensional SJT. Methods for increasing construct transparency that had previously been applied in the AC literature (Kleinmann et al., 1996) were adapted for use in the present study in order to address two hypotheses and one research question: (1) that scale estimates of internal consistency would be higher for scores from the construct transparency condition, (2) that evidence of structural validity would be stronger for scores from the construct transparency condition, and (3) whether correlations between SJT scale scores and external measures differed as a function of the transparency manipulation.

The study found that there were no significant differences between the two conditions for either estimates of scale intercorrelations or scale reliabilities. It was also found that the patterns of correlations with external scales for the two conditions were very similar. One difference that emerged was that scores in the control condition were lower than those of the construct transparency condition for all scales. However, the standard deviations of the two conditions were similar, indicating that construct transparency did not restrict the variability of the scales. Although slightly more evidence of structural validity was found for scores from the construct transparency condition, some incremental fit indices (i.e., structural equation modeling fit indices that compare the target and null models) failed to reach values that are commonly viewed as

acceptable. Despite the low values for incremental fit indices, the evidence of structural validity favored the scores from the construct transparency condition. Lastly, the results showed that scale correlations with external measures did not differ as a function of the construct transparency manipulation. The following sections review the findings with respect to the viability of (a) using a construct-oriented approach to develop SJTs, (b) making the content of SJTs transparent to test takers, and (c) developing multidimensional SJTs of sensemaking skills. The limitations of the present study and suggestions and implications for future research are also discussed.

Construct-Oriented SJTs

Although current criterion-related evidence from the literature supports the use of SJTs as a measurement method, there is little construct-related evidence in support of SJTs (Christian et al., 2010). Therefore, a construct-oriented approach was taken in the present research. Past research suggests that the internal consistencies of SJTs are lacking as low estimates are often reported (Lievens & Patterson, 2011; Mcdaniel et al., 2001), and the structural validity of SJT scores are rarely examined. The present study attempted to use construct transparency as a means of increasing the estimates of internal consistency and improving the structural validity of a multidimensional SJT.

The construct-oriented approach used for the current SJT may be useful for other test developers in their research. This approach allowed for reporting of construct-level scores as opposed to an overall SJT score. Thus, from the data it was apparent that some constructs had different psychometric properties than others. For instance, the scales of cognitive reappraisal, considering other perspectives, and expressive suppression were found to be the scales that had higher internal consistencies compared

to the other scales of the SJT. There was also more evidence of structural validity found for these scales. Specifically, the three-factor model that included the items for these three scales had fit indices that were much closer to levels of acceptability compared to the fit indices of the six-factor model that included items for all six scales.

As previously discussed, the reported internal consistencies of SJTs are generally low. Out of 17 published papers that I found that reported internal consistencies, the average number of SJT items was 21 (range = 5 to 58 items) with an average internal consistency estimate of .60 (SD = .18; range = .20 to .92). Increasing the number of items to this average of 21 would result in the two additional scales of cognitive reappraisal and considering other perspectives having an internal consistency of at least .70. Thus, by merely increasing the number of items per scale, the reliabilities of the scales would be greatly improved and further examination of evidence of validity could be conducted.

Other researchers have also developed SJTs targeted to measure particular constructs other than job knowledge. For instance, a study by de Meijer, Born, van Zielst, and van der Molen (2010) used a more traditional method of creating SJTs (i.e., using subject matter experts to collect critical incidents and create response options; Ployhart & MacKenzie, 2011) in order to create an 11-item video-based SJT focused on the construct of integrity. Although the resultant SJT scale had an internal consistency of .69, the researchers found evidence to support the convergent and discriminant validity as well as the structural validity of the scores. Thus, these results for a SJT designed to measure a single construct are encouraging. In general, more researchers should continue to investigate novel methods for developing construct-oriented SJTs,

especially those designed to target multiple constructs, and likewise conduct construct-related validation studies.

Transparency and SJTs

Construct transparency has been examined in the AC and interview literatures as a means of minimizing the influence of extraneous factors on scores (Kolk et al., 2003; Smith-Jentsch, 2007). In the current study, the manipulation check showed that the methodology adapted from the AC literature to manipulate construct transparency was successfully extended in the context of a SJT study. However, the results were different than what has typically been found in the AC and interview literatures. For instance, in a study investigating the effect of transparency on structured interviews, Klehe, Konig, Ritchter, Kleinmann, and Melchers (2008) found that transparent interviews demonstrated improvement in construct validity compared to non-transparent interviews. Their results also showed that interviewees in transparent interviews received higher scores on all of the measured dimensions than interviewees in non-transparent interviews. Although the present study did not show much improvement in evidence of validity, this study did show that transparency resulted in higher participant scores. Despite the higher scores for the present study, the standard deviations of the scores were not different compared to those of the control condition, suggesting that increasing transparency did not produce uniformly high scores.

Also, considering that participants are informed what the test is measuring, construct transparency may increase reactions towards and acceptance of the test as well as increase perceptions of the fairness of the procedures (Schuler, 1993). Furthermore, in order to apply construct transparency methods, a test must be developed around the

constructs of interest. This would mean that as a result of applying construct

transparency methodology, the number of construct-oriented SJTs would rise which

would increase the instance of reporting of construct-level scores. This information

could then be used to help provide evidence as to what constructs reflected in SJT

scores predict job performance (Christian et al., 2010). In general, although making test

content more transparent may not necessarily improve validity, there may be other

potential benefits without any detriment to validity. More research on the use of

transparency techniques applied to SJTs while examining a range of outcomes is

needed.

SJTs and Sensemaking Leadership Skills

The general goal of the present research was to provide evidence of validity for

a multidimensional measure of sensemaking leadership skills. Some of the sensemaking

leadership scales received more psychometric support than others. The scales of

identification of key causes, forecasting, and ethical decision making lacked supportive

evidence the most. The reliabilities of these three scales were very low, and they

received little evidence of structural validity. The scales of cognitive reappraisal,

considering other perspectives, and expressive suppression showed higher scale

reliabilities as well as more evidence of structural validity. One interesting result from

this study was that despite all of the scales being conceptually related as they each

represent different aspects of sensemaking, the half of the scales that were

comparatively more interpersonal (i.e., socio-emotional) in nature received more

evidence of validity than the other half of the scales that were more cognitive in nature.

This broad distinction between the scales offers a potential explanation for why it is that

certain scales received more support than others. Research from the sensemaking literature suggests that case-based knowledge, which is knowledge gained from similar past experiences, is used substantially when engaging in sensemaking and when training others on how to use sensemaking strategies (Lundberg, 2000; Mumford et al., 2007, 2008). It may be that the more cognitive constructs of identification of key causes, forecasting, and ethical decision making require more case-based knowledge (Marcy & Mumford, 2010; Mumford et al., 2008; Shipman et al., 2010) than the sample of college students in the present study possessed, whereas the more interpersonal constructs may have been more easily relatable to the types of situations that college students often encounter (i.e., when they have to consider the viewpoints of others or regulate their emotions).

Limitations

Given the importance of case-based knowledge, and domain experience in general, to sensemaking, one salient limitation of this study was its use of participants with little military background. Although all participants were given some basic information about the military, without extensive case-based knowledge it is likely that the respondents would not be sensitive to the key themes and principles that are common across the ostensibly different situations described in the SJT. This may explain why there were particularly poor psychometric properties for the more cognitive scales. Therefore, it is recommended that in future studies the SJT be completed by a sample of participants who have the requisite mental models and are not influenced by preconceived notions (i.e., implicit theories) of military leadership. For instance, having a sample of military officers complete the measure and comparing the scores of that

sample to the current study's results would help to determine if various psychometric properties of the SJT are affected by domain expertise.

As previously stated, certain scales lacked supportive validity evidence. This may be in part because the participants were not trained on the skills used in the present study. The skills used in the present study are highly complex and are especially useful in ambiguous situations. In order to facilitate better understanding of what the constructs are and how they are effectively and ineffectively used in various situations, a study that integrates training into the transparency manipulation may be worthwhile. Previous research that has trained participants on sensemaking skills (Mumford et al., 2008) could be adapted for a construct transparency manipulation specifically by using case-based examples and encouraging participants to compare cases in order to extract the common principles varying across the cases (Gentner, Loewenstein, & Thompson, 2003). For instance, participants in the construct transparency condition could first watch a PowerPoint presentation that introduced them to the skills of interest by providing definitions and behavioral examples, followed by illustrated use of the skills across a variety of cases. The participants could then role play a number of cases and receive feedback on their use of the sensemaking skills. Such a training paradigm would provide participants with case-based knowledge on how to use the various sensemaking strategies which would provide them with a deeper understanding of the skills of interest. By using such a paradigm, improved psychometric properties of the SJT scales could be observed.

Another limitation was that the study used a SJT that was developed using a novel approach. Perhaps a more traditional approach for developing SJTs, while also

ensuring that particular dimensions of the SJT were focused on certain constructs (de Meijer et al., 2010), would show more supportive psychometric characteristics and results that were more in line with those from previous transparency studies in the literature.

Conclusion

The extent to which construct transparency affected various psychometric properties of a multidimensional SJT was examined. The results suggested that, overall, construct transparency did not have a much of an effect on the reliability estimates and structural validity evidence. However, there were some differences found. Participants in the construct transparency condition received higher scores compared to participants in the control condition, and evidence of structural validity slightly favored the scores of the construct transparency condition. Furthermore, only half of the scales showed promise in terms of conventional psychometric properties. Although the results did not show strong support for the psychometric properties of the multidimensional SJT sensemaking skills overall, the results were encouraging for the more interpersonal dimensions.

References

Allen, T. D., Facteau, J. D., & Facteau, C. L. (2004). Structured interviewing for OCB: Construct validity, faking, and the effects of question type. *Human Performance*, *17*(1), 1–24.

Antes, A. L., Brown, R. P., Murphy, S. T., Waples, E. P., Mumford, M. D., Connelly, S., & Devenport, L. D. (2007). Personality and ethical decision-making in research: The role of perceptions of self and others. *Journal of Empirical Research on Human Research Ethics*, *17*, 15–34.

Antes, A. L., Wang, X., Mumford, M. D., Brown, R. P., Connelly, S., & Devenport, L. D. (2010). Evaluating the effects that existing instruction on responsible conduct of research has on ethical decision making. *Academic Medicine*, *85*(3), 519–26. doi:10.1097/ACM.0b013e3181cd1cc5

Arthur, W., & Day, E. A. (2011). Assessment centers. In S. Zedeck (Ed.), *APA handbook of industrial and organizational psychology, Vol 2: Selecting and developing members for the organization* (pp. 205–235). Washington, DC: American Psychological Association.

Arthur, W., Day, E. A., & Woehr, D. J. (2008). Mend it, don't end it: An alternate view of assessment center construct-related validity evidence. *Industrial and Organizational Psychology*, *1*(1), 105–111. doi:10.1111/j.1754-9434.2007.00019.x

Bono, J. E., Foldes, H. J., Vinson, G., & Muros, J. P. (2007). Workplace emotions: The role of supervision and leadership. *The Journal of Applied Psychology*, *92*(5), 1357–67. doi:10.1037/0021-9010.92.5.1357

Brown, M. E., & Mitchell, M. S. (2010). Exploring new avenues for future research. *Business Ethics Quarterly*, *20*(4), 583–616.

Browning, B. W., & McNamee, L. G. (2012). Considering the temporary leader in temporary work arrangements: Sensemaking processes of internal interim leaders. *Human Relations*, *65*(6), 729–752. doi:10.1177/0018726711433615

Byrne, C. L., Shipman, A. S., & Mumford, M. D. (2010). The effects of forecasting on creative problem-solving: An experimental study. *Creativity Research Journal*, *22*(2), 119–138.

Calfee, R. (1993). Paper, pencil, potential, and performance. *Current Directions in Psychological Science*, *2*(1), 6–7. doi:10.1111/1467-8721.ep10770452

Carnevale, J. J., Inbar, Y., & Lerner, J. S. (2011). Individual differences in need for cognition and decision-making competence among leaders. *Personality and Individual Differences*, *51*(3), 274–278. doi:10.1016/j.paid.2010.07.002

Christian, M. S., Edwards, B. D., & Bradley, J. C. (2010). Situational judgment tests: Constructs assessed and a meta-analysis of their criterion-related validities. *Personnel Psychology*, *63*(1), 83–117. doi:10.1111/j.1744-6570.2009.01163.x

Clause, C. S., Mullins, M. E., Nee, M. T., Pulakos, E., & Schmitt, N. (1998). Parallel test form development: A procedure for alternate predictors and an example. *Personnel Psychology*, *51*(1), 193–208. doi:10.1111/j.1744-6570.1998.tb00722.x

Cooper, O. D., Arsenault, M. L., Mracek, D. L., Hardy, J. H., Day, E. A., & Connelly, S. (2012). *Developing a situational judgment test criterion measure for the Virtual Platoon Leadership Training for Training for Second Lieutenants*. Norman, OK.

Cooper, O. D., Day, E. A., Connelly, S., Arsenault, M. L., Hardy, J. H., & Mracek, D. L. (2013). Development of a construct-oriented situational judgment test of sensemaking skills. Poster presented at the 28th Annual Conference of the Society for Industrial and Organizational Psychology, Houston, TX.

De Meijer, L. A., Born, M. P., van Zielst, J., & van der Molen, H. T. (2010). Construct-driven development of a video-based situational judgment test for integrity. *European Psychologist*, *15*(3), 229–236. doi:10.1027/1016-9040/a000027

Diefendorff, J. M., & Richard, E. M. (2008). Not all emotional display rules are created equal: Distinguishing between prescriptive and contextual display rules. In N. M. Ashkanasy & C. L. Cooper (Eds.), *Research companion to emotion in organizations* (pp. 316–334). Northampton, MA: Edward Elgar Publishing.

Dutton, J. E., & Duncan, R. B. (1987). The influence of the strategic planning process on strategic change. *Strategic Management Journal*, *8*, 103–116.

Gentner, D., Loewenstein, J., & Thompson, L. (2003). Learning and transfer: A general role for analogical encoding. *Journal of Educational Psychology*, *95*(2), 393–408. doi:10.1037/0022-0663.95.2.393

George, J. M., & Zhou, J. (2007). Dual tuning in a supportive context: Joint contributions of positive mood, negative mood, and supervisory behaviors to employee creativity. *Academy Of Management Journal*, *50*(3), 605–622.

Gooty, J., Connelly, S., Griffith, J., & Gupta, A. (2010). Leadership, affect and emotions: A state of the science review. *The Leadership Quarterly*, *21*(6), 979–1004. doi:10.1016/j.leaqua.2010.10.005

Gosserand, R. H., & Diefendorff, J. M. (2005). Emotional display rules and emotional

    labor: The moderating role of commitment. *Journal of Applied Psychology*, *90*(6),

    1256–1264.

Gross, J. J. (1998). Antecedent- and response-focused emotion regulation: Divergent

    consequences for experience, expression, and physiology. *Journal of Personality*

    and Social Psychology, *74*(1), 224–37.

Gross, J. J. (2001). Emotion regulation in adulthood. *Current Directions in*

    Psychological Science, *10*(6), 214–219.

Hauck, E. L., Shipman, A. S., Arsenault, M. L., Cooper, O. D., Schuelke, M. J., Day, E.

    A., & Connelly, S. (2010). Developing prototype content for a virtual learning

    interface for the purpose of training, education, and developing leadership in the

    U.S. army: Virtual platoon leadership training for second lieutenants. Norman,

    OK.

Heilman, R. M., Crişan, L. G., Houser, D., Miclea, M., & Miu, A. C. (2010). Emotion

    regulation and decision making under risk and uncertainty. *Emotion*, *10*(2), 257–

    65. doi:10.1037/a0018489

Howard, A. (2008). Making assessment centers work the way they are supposed to.

    Industrial and Organizational Psychology, *1*(1), 98–104.

Hülsheger, U. R., Alberts, H. J. E. M., Feinholdt, A., & Lang, J. W. B. (2012). Benefits

    of mindfulness at work: The role of mindfulness in emotion regulation, emotional

    exhaustion, and job satisfaction. *The Journal of Applied Psychology*.

    doi:10.1037/a0031313

Isenberg, D. J. (1986). Thinking and managing: A verbal protocol analysis of

    managerial problem-solving. *Academy of Management Journal*, *29*, 775–789.

Jensen, A. R. (1993). Test validity: g versus "tacit knowledge". *Current Directions in*

    Psychological Science, *2*(1), 9–10. doi:10.1111/1467-8721.ep10770493

Jensen, E. (2009). Sensemaking in military planning: A methodological study of

    command teams. *Cognition, Technology & Work*, *11*(2), 103–118.

    doi:10.1007/s10111-007-0084-x

Jones, T. M. (1991). Ethical decision making by individuals in organizations: An issue-

    contigent model. *Academy of Management Review*, *16*(2), 366–395.

Klehe, U., König, C. J., Richter, G. M., Kleinmann, M., & Melchers, K. G. (2008).

    Transparency in structured interviews: Consequences for construct and criterion-

    related validity. *Human Performance*, *21*(2), 107–137.

    doi:10.1080/08959280801917636

Kleinmann, M. (1993). Are rating dimensions in assessment centers transparent for

    participants? Consequences for criterion and construct validity. *Journal of Applied*

    Psychology, *78*(6), 988–993.

Kleinmann, M., Kuptsch, C., & Koller, O. (1996). Transparency: A necessary

    requirement for the construct validity of assessment centres. *Applied Psychology:*

    An International Review, *45*(1), 67–84.

Kligyte, V., Marcy, R. T., Sevier, S. T., Godfrey, E. S., & Mumford, M. D. (2008). A

    qualitative approach to Responsible Conduct of Research (RCR) training

    development: Identification of metacognitive strategies. *Science and Engineering*

    Ethics, *14*(1), 3–31. doi:10.1007/s11948-007-9035-4

Kligyte, V., Marcy, R. T., Waples, E. P., Sevier, S. T., Godfrey, E. S., Mumford, M. D., & Hougen, D. F. (2008). Application of a sensemaking approach to ethics training in the physical sciences and engineering. *Science and Engineering Ethics*, *14*(2), 251–78. doi:10.1007/s11948-007-9048-z

Kolk, N. J., Born, M. P., & van der Flier, H. (2003). The transparent assessment centre: The effects of revealing dimensions to candidates. *Applied Psychology: An International Review*, *52*(4), 648–669.

Lance, C. E. (2008). Why assessment centers do not work the way they are supposed to. *Industrial and Organizational Psychology*, *1*(1), 84–97. doi:10.1111/j.1754-9434.2007.00017.x

Lievens, F., & Patterson, F. (2011). The validity and incremental validity of knowledge tests, low-fidelity simulations, and high-fidelity simulations for predicting job performance in advanced-level high-stakes selection. *The Journal of Applied Psychology*, *96*(5), 927–40. doi:10.1037/a0023496

Lonergan, D., Scott, G., & Mumford, M. (2004). Evaluative aspects of creative thought: Effects of appraisal and revision standards. *Creativity Research Journal*, *16*(2), 231–246. doi:10.1207/s15326934crj1602&3_7

Lopes, P. N., Nezlek, J. B., Extremera, N., Hertel, J., Fernández-Berrocal, P., Schütz, A., & Salovey, P. (2011). Emotion regulation and the quality of social interaction: Does the ability to evaluate emotional situations and identify effective responses matter? *Journal of Personality*, *79*(2), 429–67. doi:10.1111/j.1467-6494.2010.00689.x

Lundberg, C. G. (2000). Made sense and remembered sense: Sensemaking through

    abduction. *Journal of Economic Psychology*, *21*(6), 691–709. doi:10.1016/S0167-

    4870(00)00027-1

Marcy, R. T., & Mumford, M. D. (2007). Social innovation: Enhancing creative

    performance through causal analysis. *Creativity Research Journal*, *19*(2-3), 123–

    140. doi:10.1080/10400410701397180

Marcy, R. T., & Mumford, M. D. (2010). Leader cognition: Improving leader

    performance through causal analysis. *The Leadership Quarterly*, *21*(1), 1–19.

    doi:10.1016/j.leaqua.2009.10.001

Marta, S., Leritz, L. E., & Mumford, M. D. (2005). Leadership skills and the group

    performance: Situational demands, behavioral requirements, and planning. *The*

    Leadership Quarterly, *16*(1), 97–120. doi:10.1016/j.leaqua.2004.04.004

Maurer, T. J., Solamon, J. M., & Lippstreu, M. (2008). How does coaching interviewees

    affect the validity of a structured interview? *Journal of Organizational Behavior*,

    29, 355–371. doi:10.1002/job

McClelland, D. C. (1993). Intelligence is not the best predictor of job performance.

    Current Directions in Psychological Science, *2*(1), 5–6.

McDaniel, M. A., Hartman, N. S., Whetzel, D. L., & Grubb, W. L. (2007). Situational

    judgment tests, response instructions, and validity: A meta-analysis. *Personnel*

    Psychology, *60*, 63–91.

Mcdaniel, M. A., Morgeson, F. P., Finnegan, E. B., Campion, M. A., & Braverman, E.

    P. (2001). Use of situational judgment tests to predict job performance: A

clarification of the literature. *Journal of Applied Psychology*, *86*(4), 730–740. doi:10.1037//0021

McEntire, L. E., Dailey, L. R., Osburn, H. K., & Mumford, M. D. (2006). Innovations in job analysis: Development and application of metrics to analyze job data. *Human Resource Management Review*, *16*(3), 310–323. doi:10.1016/j.hrmr.2006.05.004

Memedovic, S., Grisham, J. R., Denson, T. F., & Moulds, M. L. (2010). The effects of trait reappraisal and suppression on anger and blood pressure in response to provocation. *Journal of Research in Personality*, *44*(4), 540–543. doi:10.1016/j.jrp.2010.05.002

Motowidlo, S. J., Crook, A. E., Kell, H. J., & Naemi, B. (2009). Measuring procedural knowledge more simply with a single-response situational judgment test. *Journal of Business and Psychology*, *24*(3), 281–288. doi:10.1007/s10869-009-9106-4

Motowidlo, S. J., Dunnette, M. D., & Carter, G. W. (1990). An alternative selection procedure: The low-fidelity simulation. *Journal of Applied Psychology*, *75*(6), 640–647. doi:10.1037//0021-9010.75.6.640

Mumford, M. D., Connelly, S., Brown, R. P., Murphy, S. T., Hill, J. H., Antes, A. L., … Devenport, L. D. (2008). A sensemaking approach to ethics training for scientists: Preliminary evidence of training effectiveness. *Ethics & Behavior*, *18*(4), 315–339. doi:10.1080/10508420802487815

Mumford, M. D., Devenport, L. D., Brown, R. P., Connelly, S., Murphy, S. T., Hill, J. H., & Antes, A. L. (2006). Validation of ethical decision making measures:

Evidence for a new set of measures. *Ethics & Behavior*, *16*(4), 319–345. doi:10.1207/s15327019eb1604_4

Mumford, M. D., Friedrich, T. L., Caughron, J. J., & Byrne, C. L. (2007). Leader cognition in real-world settings: How do leaders think about crises? *The Leadership Quarterly*, *18*(6), 515–543. doi:10.1016/j.leaqua.2007.09.002

Munro, D., Bore, M., & Powis, D. (2005). Personality factors in professional ethical behaviour: Studies of empathy and narcissism. *Australian Journal of Psychology*, *57*(1), 49–60. doi:10.1080/00049530412331283453

Osburn, H. K., & Mumford, M. D. (2006). Creativity and planning: Training interventions to develop creative problem-solving skills. *Creativity Research Journal*, *18*(2), 173–190. doi:10.1207/s15326934crj1802_4

Ployhart, R. E., & Ehrhart, M. G. (2003). Be careful what you ask for: Effects of response instructions on the construct validity and reliability of situational judgment tests. *International Journal of Selection and Assessment*, *11*(1), 1–16. doi:10.1111/1468-2389.00222

Ployhart, R. E., & MacKenzie, W. I. (2011). Situational judgment tests: A critical review and agenda for the future. In S. Zedeck (Ed.), *APA handbook of industrial and organizational psychology, Vol 2: Selecting and developing members for the organization* (pp. 237–252). Washington, DC: American Psychological Association. doi:10.1037/12170-008

Ployhart, R. E., & Weekley, J. A. (2006). Situational judgment: Some suggestions for future science and practice. *Situational Judgment Tests: Theory, Measurement and Application*, 345–350.

Pugh, S. D., Groth, M., & Hennig-Thurau, T. (2011). Willing and able to fake emotions: A closer examination of the link between emotional dissonance and employee well-being. *The Journal of Applied Psychology*, *96*(2), 377–90. doi:10.1037/a0021395

Rouleau, L., & Balogun, J. (2011). Middle managers, strategic sensemaking, and discursive competence. *Journal of Management Studies*, *48*(5), 953–983. doi:10.1111/j.1467-6486.2010.00941.x

Ruch, F., & Ruch, W. W. (1983). *Employee Aptitude Survey technical report*. Chicago, IL.

Sackett, P. R., & Lievens, F. (2008). Personnel selection. *Annual Review of Psychology*, *59*, 419–450. doi:10.1146/annurev.psych.59.103006.093716

Saffo, P. (2007). Six rules for effective forecasting. *Harvard Business Review*, *85*(7/8), 122–131.

Schmitt, N., & Chan, D. (2006). Situational judgment tests: Method or construct? In J. A. Weekley & R. E. Ployhart (Eds.), *Situational judgment tests: Theory, measurement, and application* (pp. 135–155). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Schuler, H. (1993). Social validity of selection situations: A concept and some empirical results. In H. Schuler, J. L. Farr, & M. Smith (Eds.), *Personnel Selection and Assessent: Individual and Organizational Perspectives* (pp. 11–26). Hillsdale, NJ: Lawrence Erlbaum & Associates, Inc.

Shipman, A. S., Byrne, C. L., & Mumford, M. D. (2010). Leader vision formation and forecasting: The effects of forecasting extent, resources, and timeframe. *The Leadership Quarterly*, *21*(3), 439–456. doi:10.1016/j.leaqua.2010.03.008

Smith-Jentsch, K. A. (2007). The impact of making targeted dimensions transparent on relations with typical performance predictors. *Human Performance*, *20*(3), 187–203.

Thornton, G. C., & Byham, W. C. (1982). *Assessment centers and managerial performance*. New York: Academic Press.

Van Der Linden, D., Sonnentag, S., Frese, M., & Van Dyck, C. (2001). Exploration strategies, performance, and error consequences when learning a complex computer task. *Behaviour and Information Technology*, *20*, 189–198. doi:10.1080/0144929011004799

van't Wout, M., Chang, L. J., & Sanfey, A. G. (2010). The influence of emotion regulation on social interactive decision-making. *Emotion*, *10*(6), 815–21. doi:10.1037/a0020069

Vroom, V. H., & Jago, A. G. (2007). The role of the situation in leadership. *The American Psychologist*, *62*(1), 17–24. doi:10.1037/0003-066X.62.1.17

Watley, L. D., & May, D. R. (2004). Enhancing moral intensity: The roles of personal and consequential information in ethical decision-making. *Journal of Business Ethics*, *50*(2), 105–126. doi:10.1023/B:BUSI.0000022147.41538.ba

Weick, K. E., Sutcliffe, K. M., & Obstfeld, D. (2005). Organizing and the process of sensemaking. *Organization Science*, *16*(4), 409–421. doi:10.1287/orsc.1050.0133

Whetzel, D. L., & McDaniel, M. a. (2009). Situational judgment tests: An overview of current research. *Human Resource Management Review*, *19*(3), 188–202. doi:10.1016/j.hrmr.2009.03.007

Yukl, G. (2006). *Leadership in Organizations*. Upper Saddle River; NJ.

Yurtsever, G. (2008). Negotiators' profit predicted by cognitive reappraisal, suppression of emotions, misrepresentation of information, and tolerance of ambiguity. *Perceptual and Motor Skills*, *106*(2), 590–608.

Table 1

Descriptive Statistics and Comparison of Means for Control and Transparency Conditions

SJT scale	Control			Transparency			t	d
	Range[a]	M	SD	Range[a]	M	SD		
1. Cognitive reappraisal	-1 to 16	11.94	3.35	2 to 16	12.68	3.04	-2.24*	-0.23
2. Considering other perspectives	-9 to 15	6.55	5.36	-9 to 16	8.15	5.13	-2.98**	-0.31
3. Identification of key causes	-2 to 16	10.97	2.98	4 to 16	11.46	2.84	-1.64	-0.17
4. Forecasting	-4 to 16	6.81	3.82	-8 to 15	7.57	3.38	-2.06*	-0.21
5. Ethical decision making	-6 to 15	8.97	3.94	-3 to 16	9.52	4.02	-1.37	-0.14
6. Expressive suppression	-9 to 16	7.77	6.71	-6 to 16	9.96	5.99	-3.37**	-0.35

Note. [a]Potential range for all scales = -16 to 16. For Control condition, n = 192. For Transparency condition, n = 191. df = 381. *$p < .05$. **$p < .01$.

Table 2

Correlations of SJT Scales by Condition

SJT scale	1	2	3	4	5	6
1. Cognitive reappraisal	--	.19**	.10	.14	.32**	.52**
2. Considering other perspectives	.21**	--	.17*	.15*	.28**	.24**
3. Identification of key causes	.08	.14	--	.17*	.26**	.22**
4. Forecasting	.23**	.30**	.29**	--	.37**	.14
5. Ethical decision making	.37**	.30**	.28**	.48**	--	.41**
6. Expressive suppression	.42**	.18*	.20**	.22**	.24**	--

Note. Correlations for the control condition (*n* = 192) are presented below the diagonal while correlations for the transparency condition (*n* = 191) are presented above the diagonal. **p* < .05. ***p* < .01.

46

Table 3

Reliability Estimates for Control and Transparency Conditions

| SJT scale | Control | | Transparency | |
	Odd-Even Split-half	95% CI	Odd-Even Split-half	95% CI
1. Cognitive reappraisal	.57	[.43, .68]	.53	[.38, .65]
2. Considering other perspectives	.64	[.52, .73]	.64	[.52, .73]
3. Identification of key causes	.28	[.04, .46]	.21	[.00, .41]
4. Forecasting	.37	[.16, .53]	.08	[.00, .31]
5. Ethical decision making	.39	[.19, .54]	.45	[.27, .59]
6. Expressive suppression	.86	[.81, .90]	.80	[.73, .85]

Table 4

Summary for Confirmatory Factor Analyses for Control Condition

Model	χ^2	df	BIC	CFI	RMSR	RMSEA [90% CI]	$\triangle\chi^2$
Full 48-item test							
1. One-factor	1741.69	1080	2246.41	.45	.08	.06 [.05, .06]	
2. Two-factor	1599.30	1079	2109.27	.57	.09	.05 [.05, .06]	
3. Six-factor[a]	1533.84	1065	2112.17	.61	.09	.05 [.04, .05]	
Difference between Model 1 and Model 2		1					142.39**
Difference between Model 1 and Model 3		15					207.85**
Modified 24-item test[b]							
4. One-factor	576.12	252	828.48	.57	.09	.08 [.07, .09]	
5. Three-factor	382.54	249	650.67	.81	.08	.05 [.04, .06]	
Difference between Model 4 and Model 5		3					193.58**

Note. [a]The six-factor model yielded a non-positive definite matrix. Therefore accuracy of the fit indices are questionable. [b]Excluding the items from the forecasting, identification of key causes, and ethical decision making scales. **$p < .01$.

48

Table 5

Summary for Confirmatory Factor Analyses for Transparency Condition

Model	χ^2	df	BIC	CFI	RMSR	RMSEA [90% CI]	$\triangle\chi^2$
Full 48-item test							
1. One-factor	1689.03	1080	2193.25	.52	.08	.05 [.05, .06]	
2. Two-factor	1595.60	1079	2105.08	.59	.08	.05 [.04, .06]	
3. Six-factor	1485.62	1065	2068.62	.67	.07	.05 [.04, .05]	
Difference between Model 1 and Model 2		1					93.43**
Difference between Model 1 and Model 3		15					203.41**
Modified 24-item test[a]							
4. One-factor	461.03	252	713.14	.71	.10	.07 [.06, .08]	
5. Three-factor	332.64	249	600.51	.89	.07	.04 [.03, .05]	
Difference between Model 4 and Model 5		3					128.39**

Note. [a]Excluding the items from the forecasting, identification of key causes, and ethical decision making scales. **$p < .01$.

Table 6

Correlations between SJT Scales and Cognitive Scales

SJT scale	General intelligence			Verbal reasoning			Military background[a]		
	Contr	Trans	z	Contr	Trans	z	Contr	Trans	z
1. Cognitive reappraisal	.17*	.13	0.40	.23**	.17*	0.61	.02	.06	-0.39
2. Considering other perspectives	.22**	.10	1.20	.25**	.18*	0.71	-.17*	.08	-2.45*
3. Identification of key causes	.18*	.24**	-0.61	.23**	.25**	-0.21	.04	-.03	0.68
4. Forecasting	.32**	.27**	0.53	.33**	.22**	1.16	.01	.02	-0.10
5. Ethical decision making	.25**	.28**	-0.31	.28**	.30**	-0.21	-.05	-.03	-0.19
6. Expressive suppression	.11	.15*	-0.40	.24**	.13	1.11	.09	.15*	-0.59

Note. [a]Military background was dummy coded such that a "0" represented no military background and a "1" represented having a military background. Contr = Control condition. Trans = Construct transparency condition. Z-scores represent comparison between correlations in the control (*n* = 192) and the construct transparency conditions (*n* = 191). **p* < .05. ***p* < .01.

50

Appendix B: Transparency Manipulation Presentation
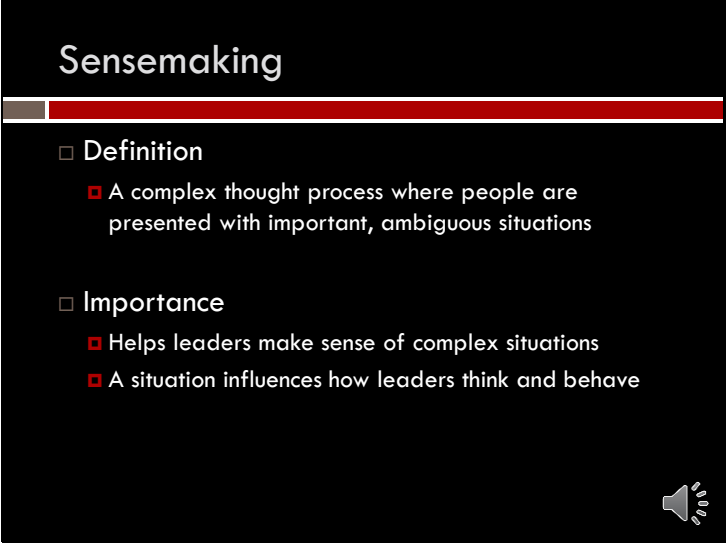
Slide 1

Slide audio:

This presentation will describe several skills important to leadership as measured in the current leadership and decision-making exercise. These skills are called "sensemaking leadership skills". Leaders are often presented with complex social problems that require the use of a certain set of sensemaking skills in order to find the best course of action. During this presentation you will learn about several skills essential to leadership in complex social situations.

Slide 2

Sensemaking

- Definition
  - A complex thought process where people are presented with important, ambiguous situations

- Importance
  - Helps leaders make sense of complex situations
  - A situation influences how leaders think and behave

Slide audio:

Sensemaking can be defined as a complex thought process that occurs when people are presented with important situations where it is unclear what needs to be done.

Sensemaking is important because during these kinds of situations, leaders use complex thinking frameworks that allow them to better handle and understand the situation.

The overarching idea of sensemaking is that multiple parts of situations influence how leaders think about the situations and influence the approaches leaders consider for how to deal with the situations.

Slide 3

Slide audio:

Sensemaking can be broken down into several skills. In today's presentation, we will

discuss 6. These 6 skills are the specific skills measured in the exercise you will soon be

doing. They are identification of key causes, cognitive reappraisal, expressive

suppression, forecasting, considering other perspectives, and ethical decision making.

Slide 4

Slide audio:

The first skill we will cover is identification of key causes. Identification of key causes can be defined as identifying causes or factors that influence important elements of a situation. This means that when in a complex situation, you look for aspects of the situation that are influencing or ultimately causing what is happening. It is important to keep in mind that key causes are those that (1) have a significant influence on the situation and (2) have an effect on the outcomes of the situation. Identification of key causes can also be thought of as critical thinking.

An effective example of identification of key causes would be critically thinking about and determining the underlying issues affecting a given situation. An ineffective example of identification of key causes would be focusing on a narrow, trivial, or irrelevant aspect of a situation and assuming that it is the cause of the situation.

Slide 5

Slide audio:

Cognitive reappraisal can be defined as a defensive cognitive strategy where one reexamines an emotional situation to reduce the experience and emotional impact so that goals can be successfully accomplished. Cognitive reappraisal occurs before an emotion is fully felt or experienced.

An effective example of cognitive reappraisal is when you turn an emotionally charged event into something that is less emotional. Doing this will allow you to better concentrate on your goal.

An ineffective example of cognitive reappraisal is thinking of an already emotional event in a way that makes it even more emotionally powerful, and consequently magnifying the emotion felt.

Slide 6

Slide audio:

Expressive suppression can be defined in two ways: (1) as preventing the expression of a felt emotion and (2) as behaving in such a way that others could not know what you are feeling. Expressive suppression occurs after an emotion is felt.

An effective example of expressive suppression is when you are confronted with an emotional situation, you hold back and do not express that emotion to other people so that you and others are able to stay on task and continue to accomplish goals.

An ineffective example of expressive suppression would be when in an emotional situation, you overreact and express the felt emotion even more intensely than what you feel.

Slide 7

Slide audio:

The next sensemaking skill that we will talk about is forecasting. Forecasting can be

defined as predicting likely outcomes or consequences of carrying out an idea within a

particular setting. When in complex situations, there are often a large number of

decision options to choose from. Each of these decision options has its own

consequence or outcome if the decision option is carried out.

An effective example of forecasting is thinking about a particular decision and the full

range of positive and negative outcomes that could result if that decision is

implemented.

An ineffective example of forecasting is thinking about a particular decision and

considering just one positive outcome that could result if that decision is implemented.

Slide 8

Slide audio:

Next we will discuss considering other perspectives. Considering other perspectives can

be defined as considering the perceptions, concerns, goals, and responsibilities of others

who are affected by a situation.

An effective example of considering other perspectives is when you attempt to look at

the situation from a range or diverse set of perspectives.

An ineffective example of considering other perspectives is thinking of only one

perspective, like an individual or a single group, or even thinking of only your own

perspective.

Slide 9

Slide audio:

Finally, we will talk about ethical decision making. Ethical decision making is used in situations where a "right" course of action is unclear as the situation can involve a conflict between what is morally right and what a set of rules dictates should be done. Ethical decision making can be defined as a process during which an individual takes into consideration the situation, potential responses to a problem, and then chooses a course of action that minimizes harm to others yet is consistent with ethical guidelines which includes norms in addition to rules, policies, and guidelines.

An effective example of ethical decision making is finding a balance between the rules and the concerns of individuals involved in the situation to reach a fair course of action. An ineffective example of ethical decision making is strictly applying rules when the situation may not fully apply or completely ignoring the rules when they should be followed.

Slide 10

Slide audio:

To review this presentation, we have discussed sensemaking which is defined as a complex thought process that occurs when people are presented with important situations where it is unclear what needs to be done. We then discussed multiple sensemaking skills including identification of key causes, cognitive reappraisal, expressive suppression, forecasting, considering other perspectives, and ethical decision making. Remember these are the skills that are measured in the leadership and decision-making exercise that you will be doing. The exercise consists of several scenarios each of which has 6 questions. Each question is designed to measure one of these 6 skills.