



CAPSIM|nbox®



Technical Synopsis



Introduction

This document provides a synopsis of the procedures followed in the design and development of an assessment and development platform called CapsimInbox®. The synopsis describes the conceptual approach, development procedures, and select product versions associated with this platform.

Overview of CapsimInbox




CapsimInbox is a simulation-based behavioral assessment that is primarily designed for developmental purposes. The development of CapsimInbox followed conventional test construction procedures, which began with reviewing the published behavioral science literature for the types of skills commonly measured during assessment centers, as well as the practitioner literature regarding highly desirable, but underdeveloped, workplace skills.

CapsimInbox is modeled after a managerial in-basket exercise that is typical to assessment centers.¹ Assessment centers have been widely used in industry settings for selection, promotion, and development. Substantial research has supported their validity and utility. For example, assessment centers have been shown to be among the best predictors of managerial potential and job performance.² Within an academic environment, assessment center results have been linked to early career salary growth and promotions, as well as intrinsic career success.³ Overall, assessment center exercises like the in-basket represent the state of the art in assessing key non-technical or “soft” skills relevant to the workplace. This unique value derives from the rare external assessment that is provided. Unlike most instruments that are often used to measure interpersonal competencies and only take the form of subjective “self-ratings” (e.g., MBTI, DiSC, Strengths Finder), CapsimInbox provides objective measurement of interpersonal and decision-making skills. This allows for external feedback that is critical for personal development.⁴

In general, an in-basket (or “inbox”) approach places participants in a simulated role in which they must respond to variety of situations, problems, and information related to effective management of oneself and others in the workplace. In CapsimInbox, these situations, problems, and information are presented within an online interface that simulates a typical communication platform, complete with emails, instant messages, and shared files.

CapsimInbox is comprised of two basic components. The first component is the objective skill assessment. The second component is a self-directed tool that follows this assessment experience in order to provide developmental feedback, evidence-based tactics for improvement, and an individual development plan (IDP) builder.

CapsimInbox is a platform-based application and thus multiple versions of the tool can be developed, with each assessing different skill sets and competencies. The current synopsis describes the development and validation of the three specific versions of CapsimInbox, which are briefly described in the table below.

 General Management	 People Management	 Ethical Decision-Making
Skills Assessed		
<ul style="list-style-type: none"> • Organizing • Leading • Problem Solving • Communicating • Initiating 	<ul style="list-style-type: none"> • Organizing • Leading • Problem Solving • Communicating • Initiating 	<ul style="list-style-type: none"> • Recognizing issues • Investigating facts • Identifying stakeholders • Generating solutions • Evaluating consequences
Decision Context		
General management with scenarios involving cross-functional situations and problems	People management with scenarios spanning structuring work, acquiring talent, management talent, and rewarding performance	Business situations with ethical issues spanning conflicts of interest; rights and responsibilities; information and confidentiality, and use of resources

Overall Assessment Design Philosophy

An evidence-based approach was taken to develop CapsimInbox that entailed conventional test construction and validation procedures, as well as deriving focal content from the relevant academic literature (e.g., essential workplace skills). As previously noted, CapsimInbox is modeled after a managerial in-basket exercise. When participants move through the CapsimInbox assessment, they must choose what they believe are the best responses to the various scenarios that are presented. The primary function of these assessment “stimuli” is to directly elicit responses that indicate a participant’s level of proficiency on the focal skills. The stimuli are therefore designed to ensure both generalizability and measurement precision. This means that the situations, problems, and information included in the stimuli are applicable across a wide range of jobs, careers, and organizations, yet are precise enough to elicit distinct skill proficiencies. For a participant, this also means that knowing the “correct answer” to a given stimulus (e.g., email) after CapsimInbox is completed is less important for development than the proficiency level the stimulus reveals. That is, how to “solve” each stimulus is not the point of the assessment; rather the goal is to provide participants’ an overall snapshot of their current skill levels.

CapsimInbox presents participants with four general kinds of stimuli:

1. “Just Read (FYI)” stimuli do not require a response and are intended to provide information that may or may not be relevant to other emails, messages, or files
2. “Read and Choose” stimuli are “self-contained” in that all of the information needed to choose among response options is contained within the email or message itself
3. “Research and Choose” stimuli require searching for a particular piece of information beyond the focal stimulus of the email or message to effectively choose among response options
4. “Synthesize and Choose” stimuli require searching for multiple pieces of information beyond the focal stimulus, or combining another piece of information with something in the stimulus, in order to effectively choose among response options

To reiterate, participants can only choose one response option they view as most appropriate or effective for addressing a given situation or problem. Most stimuli award points for a single response option (i.e., one correct answer). Others award points for more than one response option (i.e., two correct answers) such that participants earn points for choosing either of the appropriate options. Finally, there are some stimuli that present response options that detract points (i.e., negative answers). These are linked to responses that go beyond a simple lack of skill to instead indicate negative or deleterious behavior (e.g., engaging in a personal attack on a coworker, overlooking a policy violation, etc.).

An Illustrative Example

Below is an illustrative example of an assessment stimulus from CapsimInbox. This example is in the form of an email to the participant (Sandy) from an employee (Ted) that she knows well but works in a different department. This is an example of the “read and choose” type of stimulus discussed above.

 **Ted Merblestein**
Financial Analyst

To: Sandy Shelton
CC:
BCC:
From: Ted Merblestein
Sales Training Approval
2:15 PM

Dear Sandy,
I'm writing to ask if you'd approve my request to attend a company sponsored sales training program. I realize this might be catching you off-guard, but although I'm currently working in Finance, I'm planning on transitioning to sales with the company or elsewhere. Our policy states that I just need a manager to approve. You've always been very kind to me and you know what a hard worker I am, so I'm hoping you won't mind approving this great development opportunity for me. I'd prefer my boss doesn't find out yet as it might tip him off that I'm looking to leave, so please keep this confidential.

Thanks in advance,
Ted

Ted Merblestein
Financial Analyst

The possible response options follow the email's body text (at the end of the email) and are shown below.

Response Options

- Reply to Ted letting him know that you are happy to approve his request for the company training.
- Reply to Ted letting him know that you are delighted that he is seeking this training and to write a short letter of support to the training manager.
- Reply to Ted via email and let him know that you feel uncomfortable with the request unless he first tells his manager.
- Forward to your administrative assistant and ask him to send an approval on your behalf.
- Reply to Ted to set up a time to meet with him explore and why he feels the need to go around his manager before you commit to approving the training.

[Submit](#)

This particular email is designed to elicit choice behavior that depicts a participant's problem solving skills. While many of the response options appear completely reasonable, the best choice in this particular situation is option “e.” The reason this response is optimal is because it exemplifies the essential first step in any problem-solving context – ensuring that one has an adequate understanding of the situation by collecting pertinent information. The other responses fail to take this critical step in effective problem solving but instead rush to the solution stage (i.e., simply deciding to agree or disagree with the request).

Procedures for Assessment Development

Four PhD-level industrial-organizational psychologists, also university business school faculty, were a part of the CapsimInbox development process. All four subject matter experts are tenured faculty at their respective institutions and have previous experience in test development.

The first step in developing each version of CapsimInbox was to identify the targeted set of skills. Conventional criterion development procedures were followed to specify the domain of interest and identify more specific skills.⁵ This began by searching and reviewing the major theoretical and empirical studies pertaining to the focal domain (e.g., ethical decision making) from both academic sources (e.g., journal articles, dissertations, etc.) and non-academic sources (e.g., institutional research summaries, technical reports, etc.).

For the General Management and People Management versions of CapsimInbox, emphasis in this first step was on identifying a parsimonious set of key skills that research has suggested to be relevant but rare in the broader workforce. After a thorough review of the literature, five skills were consistently identified across numerous sources.⁶ These are defined below.

CAPSIM|inbox
General Management

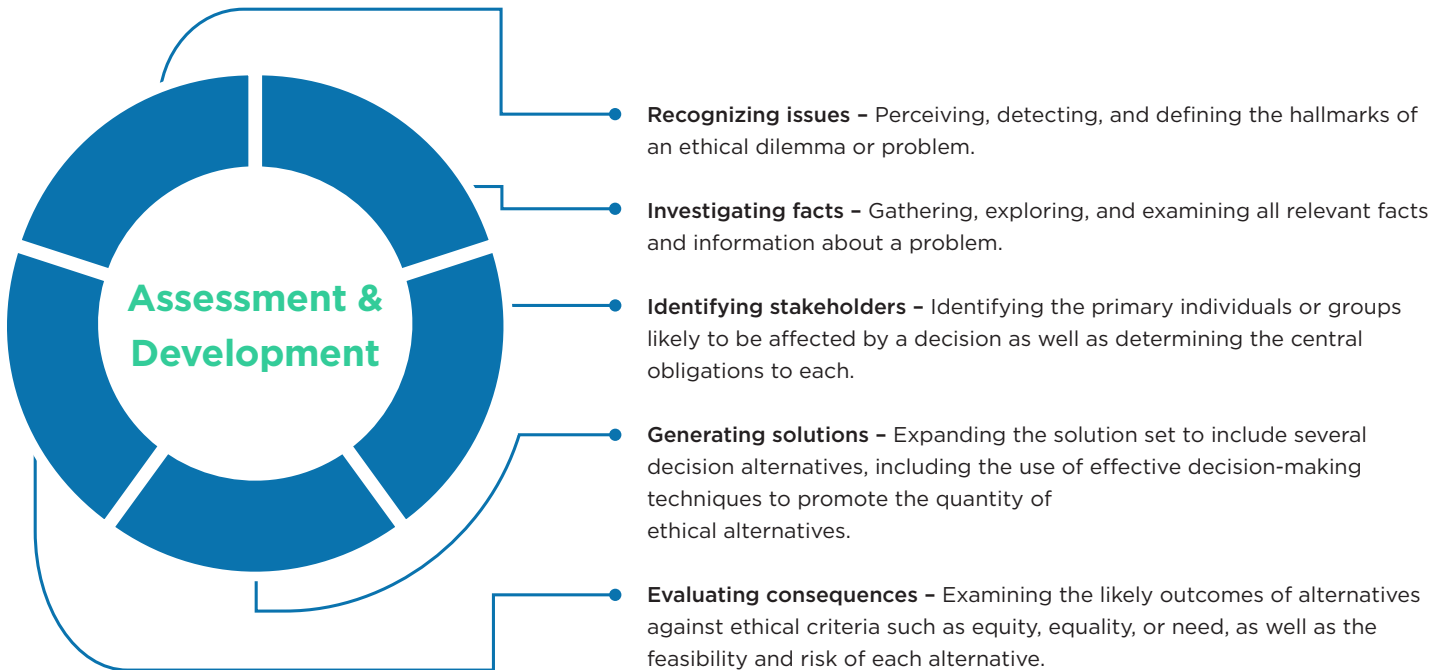
CAPSIM|inbox
People Management



For the Ethical Decision Making version of CapsimInbox, emphasis in this first step was on identifying the key skills that promote effective decision-making within common ethical situations such as conflicts of interest, individual rights, information confidentiality, and use of organizational resources. After a thorough review of the literature, five ethical decision-making skills were consistently identified across numerous sources.⁷ These are defined below.

CAPSIM|inbox

Ethical Decision-Making



Creating the Assessment Stimuli

The general approach to creating assessment stimuli across the three versions of CapsimInbox was similar to how situational judgment tests (SJTs) are commonly designed. Briefly, SJTs present a respondent with descriptions of work situations and multiple responses to each situation. In the case of CapsimInbox, these descriptions and situations are in the form of emails, instant messages, stored files, and/or other media (e.g., video). The respondent is instructed to endorse one choice only. SJTs are different than traditional multiple-choice tests. Unlike standardized multiple-choice tests, SJTs do not emphasize assessment of declarative knowledge (i.e., knowing about something). Instead, SJTs measure procedural knowledge, which reflects knowing how to do something. Research evidence has supported a strong relationship between SJTs and performance in a given role, primarily because procedural knowledge is a direct precursor to actual behavior and therefore scores from SJTs are valid indicators of people's levels of skill proficiency.⁸

Conventional SJT creation procedures were applied to develop CapsimInbox stimuli.⁹ These procedures entailed four general steps:

1. Generate multiple clusters of situations that align with the targeted skills chosen for the focal version of CapsimInbox.
2. Review situations for redundancy, consistency in length and complexity, and representativeness of each category.
3. Develop response options to each situation.
4. Revise initial stimuli based upon review from subject matter experts (SMEs), who assessed the stimuli for clarity and representativeness.

One concern with SJTs is the degree to which they may be susceptible to faking. Research suggests that although some small amount of faking is likely to occur with such items, response instructions and scoring formats are likely to substantially reduce the consequences of faking.¹⁰ Since "knowing" proceeds "doing," what matters most is that the respondent recognizes that given choice option represents the preferred response (action), not whether he or she could ultimately engage in the action in reality.

Content-related Validity

Content-related validity pertains to whether inferences drawn from an assessment are representative of the intended domain that is purportedly assessed by the device. Evidence of content-related validity supports conclusions that an assessment adequately reflects the key conceptual aspects of what it claims to measure. Ensuring favorable content-related validity of CapsimInbox began first with directly deriving the targeted skills for assessment from the academic and practitioner literatures; that is, relying on existing empirical scholarship to identify the focal content of each version of CapsimInbox as described in the previous section.

Additional evidence for the content-related validity of CapsimInbox came from SME reviews conducted by at least three of the four PhD-level industrial-organizational psychologists with experience in psychological testing and skill development noted earlier. The SME review process followed four general steps for each of the three versions of CapsimInbox.

1. SMEs reviewed descriptions of the focal CapsimInbox narrative and assessment domain, as well as information about the general participant experience
2. SMEs rated each stimulus (emails, instant messages, etc.) for its relevance to the focal skill (0 = not relevant, 1 = relevant, 2 = highly relevant)
3. SMEs rated each response option for each stimulus for its decision quality (-1 = inappropriate response, 0 = poor response, 1 = good response, 2 = excellent response); these ratings were also used to facilitate scoring
4. SMEs were asked to provide additional feedback and/or suggestions

Below are the results for the SME review for each CapsimInbox. Taken collectively, these results provide strong evidence for the content-related validity of CapsimInbox.



CapsimInbox – General Management:

99% of the stimuli items received skill relevance ratings of “1s” and “2s” (93% absolute interrater agreement). A single stimulus received a “0” rating by one of the SMEs. The average skill relevance rating across all stimuli was 1.80.



CapsimInbox – People Management:

100% of the stimuli items received skill relevance ratings of “1s” and “2s” (97% absolute interrater agreement). The average skill relevance rating across all stimuli was 1.72.



CapsimInbox – Ethical Decision Making:

97% of the stimuli items received skill relevance ratings of “1s” and “2s” (96% absolute interrater agreement). A single stimulus received a “0” rating by one of the SMEs. The average skill relevance rating across all stimuli was 1.66.

Reliability

Test-retest reliability correlations were .86, .88, and .93 for General Management (N = 61), People Management (N = 76), and Ethical Decision Making (N = 53) versions of CapsimInbox, respectively. The average time between administrations was 4.1 weeks.

Criterion-related Validity

Criterion-related validity pertains to whether scores from an assessment are associated with important outcomes. Evidence of criterion-related validity supports conclusions that an assessment predicts meaningful outcomes for individuals, teams, or organizations.

Data from the General Management and People Management versions of CapsimInbox were combined to ensure a larger sample size for analysis (each version measures the same five skills). Supplemental analysis did not show a significant difference in outcome correlations between the two versions, supporting cumulation of the data. Both overall CapsimInbox scores and more specific skill scores were examined for their associations with supervisory ratings of overall job performance and ratings of particular dimensions of job performance (i.e., organization, leadership, and communication effectiveness). Below are the results from this analysis (N = 398 job incumbents, average age = 35.2, average work experience = 12.6 years).

CapsimInbox overall scores were positively correlated with supervisory ratings of overall job performance effectiveness ($r = .43, p < .01$). By way of comparison, a meta-analysis of assessment center studies found the average correlation between assessment center scores and job performance to be .37 (Gaugler et al 1987).¹¹

CapsimInbox scores for all five specific skills were positively correlated with supervisory ratings of overall job performance effectiveness (average $r = .28$, min $r = .16$, max $r = .34$, all $p < .05$). A meta-analysis of specific skills measured by an assessment center showed similar criterion-related validity coefficients ranging from .25 to .39.¹²

CapsimInbox skill scores showed higher correlations with parallel performance dimension ratings by supervisors (average $r = .30, p < .01$) compared to non-parallel performance dimensions ($r = .12, p < .05$).

Overall, these results described above provide supportive evidence for the criterion-related validity of CapsimInbox.

For the Ethical Decision Making version of CapsimInbox, criterion-related validity was examined in a sample of 156 individuals enrolled in a graduate-level business program. Scores from CapsimInbox – Ethical Decision Making were investigated for their association with performance on a 10-item quiz that assessed knowledge of ethical theories, principles, and decision-making models relevant to business situations. The correlation between overall performance in CapsimInbox and the ethics quiz was .66 ($p < .01$). This finding provides supportive evidence for the criterion-related validity of this version of CapsimInbox.

Finally, relative weights analysis was conducted to examine the how each skill contributed to overall performance in CapsimInbox. For the General Management and People Management versions, skills accounted for comparable percentages of variance in the overall score: 17% for organizing; 18% for leading; 23% for problem solving; 17% for communicating; and 25% for initiating (model $R^2 = .88$). Similar results were found for the Ethical Decision Making version, where skills accounted for comparable percentages of variance in the overall score: 20% for recognizing issues; 20% for investigating facts; 19% for identifying stakeholders; 23% for generating solutions; and, 18% for evaluating consequences (model $R^2 = .94$). Taken collectively, these results indicate that each assessed skill contributes to overall assessment performance in an adequately balanced manner across CapsimInbox versions.

Summary

CapsimInbox is an evidence-based assessment tool and, as such, its design foundation was derived from the research literature on workplace skills and professional development, as well as assessment centers, situational judgment tests, and skill assessment. This is an important and distinguishing characteristic of CapsimInbox, not only because it ensures an empirically grounded approach to assessment, but also moves beyond popular assessment approaches that rely solely on self-ratings at the expense of delivering objective feedback for development.

The evidence-based approach to developing CapsimInbox also necessitated validity examinations to ensure the assessment measured what it was intended to measure and that scores were related to meaningful outcomes. The evidence presented in this technical synopsis provides strong empirical support for quality and utility of CapsimInbox-General Management, CapsimInbox-People Management, and CapsimInbox-Ethical Decision Making.

References

- 1 Spychalski, A. C., Quiñones, M. A., Gaugler, B. B., & Pohley, K. (1997). A survey of assessment center practices in organizations in the United States. *Personnel Psychology*, 50, 71-90. Thornton, G. C., III, & Byham, W. C. (1982). *Assessment centers and managerial performance*. New York, NY: Academic Press. Thornton, G. C., III, & Cleveland, J. N. (1990). Developing managerial talent through simulation. *American Psychologist*, 45, 190-199.
- 2 Arthur, W., Day, E.A., McNelly, T.L. & Edens, P.S. (2003) A meta-analysis of the criterion-related validity of assessment center dimensions. *Personnel Psychology*, 56, 125-154. Gaugler, B. B., Rosenthal, D. B., Thornton, G. C., III, & Bentson, C. (1987). Meta-analysis of assessment center validity. *Journal of Applied Psychology*, 72, 493-511. Bray, D. W., Campbell, R. J., & Grant, D. L. (1974). Formative years in business: A long-term AT&T study of managerial lives. New York: Wiley. Hinrichs, J. R. (1978). An eight-year follow-up of a management assessment center. *Journal of Applied Psychology*, 63, 596-601. Klimoski, R., & Brickner, M. (1987). Why do assessment centers work? The puzzle of assessment center validity. *Personnel Psychology*, 40, 243-260.
- 3 Jansen, P. G. W., & Stoop, B. A. M. (2001). The dynamics of assessment center validity: Results of a 7-year study. *Journal of Applied Psychology*, 86, 741-75. Waldman, D. A. & Korbar, T. (2003). Student assessment center performance in the prediction of early career success. *Academy of Management Learning & Education*, 3, 151-167.
- 4 Zell, E., & Krizan, Z. (2014). Do people have insight into their abilities? A metasynthesis. *Perspectives on Psychological Science*, 9, 111-125.
- 5 Cascio, W. F. & Aguinis, H. (2005). *Applied psychology in human resources management* (6th Ed). Upper Saddle River, NJ: Prentice Hall. Nagle, B. F. (1953). Criterion development. *Personnel Psychology*, 6, 271-289.
- 6 Association for Talent Development (2018). *Bridging the skills gap: Workforce development and the future of work*. Alexandria, VA: ATD Press. Casner-Lotto, J., & Barrington, L. (2006). Are they really ready to work? Employers' perspectives on the basic knowledge and applied skills of new entrants to the 21st Century U.S. workforce. *Partnership for 21st Century Skills*. Washington, DC. Dierdorff, E. C., & Ellington, J. K. (2019). O*NET and the nature of work. In F. Oswald, T. Behrend, L. Foster (eds). *Workforce Readiness and the Future of Work*. New York: Routledge. Hesse, F., Care, E., Buder, J., Sassenberg, K., & Griffin, P. (2015). A framework for teachable collaborative problem solving skills. In P. Griffin & E. Care (Eds.), *Assessment and teaching of 21st century skills: Methods and approach*. Dordrecht, The Netherlands: Springer. Manyika, J., Lund, S., Bughin, J., Robinson, K., Mischke, J., & Mahajan, D. (2016). *Independent work: Choice, necessity, and the gig economy*. McKinsey Global Institute. October. Oliveri, M. E., Lawless, R., & Molloy, H. (2017). A literature review on collaborative problem solving for college and workforce readiness. ETS Research Report Series.
- 7 Janis, I. L., & Mann, L. (1977). *Decision making: A psychological analysis of conflict, choice, and commitment*. London: Cassel & Collier Macmillan. Rest, M. (1985). *Moral development; advances in research and theory*. New York: Praeger. Trevino, L. K. (1986). Ethical decision making in organizations: A person-situation interactionist model. *Academy of management Review*, 601-617. Ferrell, O. C. & Gresham, L. G. (1985). A contingency framework for understanding ethical decision making in marketing. *Journal of Marketing*, 49, 87-96. Hunt, S. D., & Vitell, S. (1986). A general theory of marketing ethics. *Journal of Macromarketing*, 6(1), 5-16. Jones, T. M. (1991). Ethical decision making by individuals in organizations: An issue contingent model. *Academy of Management Review*, 15, 366-395. McDevitt, R., Giapponi, C., & Tromley, C. (2007). A model of ethical decision making: The integration of process and content. *Journal of Business Ethics*, 73(2), 219-229. Wilhelm, W. J. (2006). Ethical reasoning for the business classroom: A decision-making framework. *Journal of Applied Research for Business Education*, 4(2), 1-7. Johnson, C. E. (2007). Ethics in the workplace: Tools and tactics for organizational transformation. Hartman, L. P., & DesJardins, J. R. (2008). *Business ethics: Decision-making for personal integrity and social responsibility*. McGraw-Hill Irwin. Trevino, L. K., & Nelson, K. A. (2010). *Managing business ethics*. Wiley. Baldwin, T. T., Bommer, W. H., & Rubin, R. S. (2012). *Managing organizational behavior: What great managers know and do*. Boston, MA: McGraw-Hill/Irwin.
- 8 Chan, D. & Schmitt, N. (2002). Situational judgment and job performance. *Human Performance*, 15, 233-54. Lievens, F., Peeters, H., Schollaert, E. (2008). Situational tests: A review of recent research. *Personnel Review*, 37, 426-441. Lievens, F., & Sackett, P. R. (2012). The validity of interpersonal skills assessment via situational judgment tests for predicting academic success and job performance. *Journal of Applied Psychology*, 97, 460-468. 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, 730-740.
- 9 Weekley, J. A., Ployhart, R. E. (2006). *Situational judgment tests: Theory, management, and application*. Mahwah, NJ: Erlbaum

- 10** Hooper, A. C., Cullen, M. J., & Sackett, P. R. (2006). Operational threats to the use of SJTs: faking, coaching, and retesting issues. In Weekley, J.A. and Ployhart, R.E. (Eds), *Situational judgment tests: Theory, measurement and application*. Mahwah, NJ: Erlbaum, pp. 205-32.
- Nguyen, N. T., Biderman, M. D., & McDaniel, M. A. (2005). Effects of response instructions on faking a situational judgment test. *International Journal of Selection and Assessment*, 13, 250–260.
- 11** Gaugler B. B., Rosenthal D. B., Thornton G. C. III, & Bentson C. (1987). Meta-analysis of assessment center validity. *Journal of Applied Psychology*, 72, 493-511.
- 12** Arthur Jr, W., Day, E. A., McNelly, T. L., & Edens, P. S. (2003). A meta-analysis of the criterion-related validity of assessment center dimensions. *Personnel Psychology*, 56, 125-153.