

COMMENTARY

## Going beyond a validity focus to accommodate megatrends in selection system design

John W. Jones<sup>1</sup>  and Michael R. Cunningham<sup>2</sup>

<sup>1</sup>R&D and Consulting, FifthTheory, LLC, Chicago, Illinois, USA and <sup>2</sup>Department of Communication, University of Louisville, Louisville, Kentucky, USA

**Corresponding author:** John W. Jones; Email: [Jack.Jones@FifthTheory.Com](mailto:Jack.Jones@FifthTheory.Com)

Sackett, Zhang, Berry, and Lievens (2023) are to be commended for correcting the validity estimates of widely used predictors, many of which turned out to have less validity than prior studies led us to believe. Yet, we should recognize that psychologists and their clients were misled for many years about the utility of some mainstream assessments and selection system design surely suffered. Although Sackett et al. (2023) offered useful recommendations for researchers, they never really addressed selection system design from a practitioner perspective. This response aims to address that omission, emphasizing a multidimensional approach to design science (Casillas et al., 2019).

### Shifting client expectations for selection system design

Our main issue with Sackett et al. (2023) is that they provided an extremely narrow “design” perspective, apparently assuming that selection systems should be constructed using corrected validity statistics as the primary guide. In our view, this approach is out of step with the fast-moving, technocentric marketplace. Although the utilization of the most valid and reliable assessments drives the most convincing utility analyses and impact studies, validity is a necessary but not a sufficient condition for designing successful selection systems (Casillas et al., 2019).

For the past several years, many companies have had difficulty hiring qualified job candidates (Billings & Jones, 2019). To address this challenge, employers are not exclusively focusing on utilizing assessments with the highest validity. Instead, from a practitioner perspective, the building blocks for designing and implementing an assessment-based intervention require accommodating the three emerging megatrends described below. These megatrends lead to multidimensional considerations when designing contemporary selection systems and should extend the Sackett et al. (2023) analysis.

#### **Megatrend 1: Enhance test-taker experience in ultra-tight labor markets**

In today’s labor market, applicants have high expectations (Weiner et al., 2022). Although organizations are seeking qualified employees, job candidates now expect to have satisfying test-taker experiences. Test takers do not want to deal with organizations that make them jump through what seem like unnecessary hoops. Creating a satisfying test-taker experience and a valid selection system with less friction in the test-taker journey is part of the new design process.

In brief, a frictionless test-taker experience is all about creating a seamless job application, assessment, and onboarding process. The design requirement is to identify where friction exists

throughout the entire test-taker journey and eliminate it, to make the encounter as convenient and pleasant as possible. Gamification is an assessment procedure that leans into this design element.

#### *Ultra-brief assessment systems*

Lengthy and demoralizing assessments cannot be part of the frictionless design strategy. Once upon a time, employers could be advised to deploy a selection battery consisting of several instruments with hundreds of items, taking hours for applicants to complete. Sackett et al. (2023) alluded to the advantages of using multiple assessments, but without acknowledging the issue of administration time.

Certainly, lengthy batteries offer high reliability coefficients, job-related validity statistics, and often extended narrative accounts of the attributes of each job candidate and even follow-up interview questions. Over the past decades, however, the world has changed. Applicants became reluctant to devote half a day to responding to challenging items, especially for entry-level positions. Recognizing that the applicant experience was an essential element in the selection process, human resource managers created a demand for tests that are ultra-short and targeted, even if that involves some sacrifice in reliability and validity (Jones, Orban, Dages, & McHenry, 2016). It is not uncommon for large retailers today to request 15- to 20-minute express measures as part of the selection system design. Practitioners must accommodate those requests with valid and reliable tests in the allocated time.

#### *DEI-friendly measures*

Companies are increasingly willing to consider a selection system with moderate levels of validity over an assessment with higher levels of validity, if the tool with moderate validity has less adverse impact on specific populations. This design step ensures that protected groups have more confidence in the fairness of the assessment system. Sackett et al. (2023) touched on this point and reviewed a variety of strategies that can be deployed with mental ability tests, for example, to improve the overall design quality of the assessment.

Additional strategies include having items reviewed and critiqued by diverse groups who are members of the design team. Such reviews ensure that a measure does not make assumptions or include inappropriate content that creates unpleasant reactions in diverse and disadvantaged test-takers (cf. Dages & Jones, 2022). In addition, valid assessments that might be challenging to specific groups covered by an organization's DEI policy, such as the neurodiverse with disabilities (e.g., job candidates with attention-deficit/hyperactivity disorders or autism spectrum disorders), could have alternate forms developed and validated that are more DEI-friendly (cf. Dages & Jones, 2023). Accommodations for the differently abled should be considered during the design phase. From a design science perspective, this research is as important as that on recomputed validity coefficients.

#### *Utilization of extreme selection ratios*

Finally, there is a trend that is related to the test-taker experience dealing with the increasing utilization of extreme selection ratios. This means dropping the assessments altogether, or setting the cut-scores so only a few applicants will be screened out. Some organizations feel that screening out a job applicant will lead to disgruntlement and the loss of a current or potential customer. Sackett et al. (2023) wrote about lowering the cut-scores for mental ability tests to reduce adverse impact, but some organizations can barely tolerate screening out any job applicants in an ultra-tight labor market. The use of potentially compromised decision rules is not one we endorse but is an issue that needs to be a design topic.

***Megatrend 2: Use new procedures, constructs, and criteria in rapidly changing workplaces***

A second megatrend is that many organizations desire tailored assessments that address their specific strategic initiatives and align with their emerging operations. And they want those assessments faster than ever before, which creates conflicts with established paradigms. The strength of meta-analysis is its provision of an estimate of the validity of a predictor based on the weighted-average integration of the corrected results of multiple studies. The weakness of meta-analysis is that producing and accumulating enough studies to provide a stable and plausible estimate can take years, especially if focusing on one new instrument rather than a broad family of traditional instruments.

*Speed to assessment implementation*

Many organizations want to stand up their selection system within months and do not have time for lengthy, replicated validation studies that might quickly become obsolete due to high velocity shifts in their marketplace. In a time of rapid change due to COVID-19, financial disruptions in traditionally stable industries, an ultra-tight labor market, and other causes like remote work preference, the assumption that reliable knowledge will slowly and steadily accumulate can be self-defeating. Instead, current clients increasingly expect agility, speed, and just-in-time creativity from I-O psychologists. They want validity for their criteria for their new employees in their locations, rather than broadly generalizable validity statistics.

*New criteria for upskilling worker competencies*

The current marketplace is demanding new criteria for upskilling worker competencies, such as employees who can integrate easily into a diverse team or work remotely without compromising their productivity or loyalty. A related post pandemic need, especially in healthcare, military, retail, and transportation sectors, is for trauma-informed leaders, who create psychologically safe workplaces and exhibit empathy and people skills to help co-workers who are still in the process of healing from serious stressors (e.g., lethal viruses, war and injury, mass shootings, and customer attacks; Dreschler, Behrens, Cunningham, & Jones, 2023). Traditional teamwork and emotional intelligence assessments provide some guidance, but existing instruments often must be customized to function effectively in a trauma-informed organizational culture. Hence, these unique, niched instruments must be configured during the selection system design process.

*Rapid prototyping of novel assessments*

Regardless of client pressure, the deployment of unvalidated selection tests is unacceptable. Fortunately, new procedures have become available to allow for the rapid prototyping and validation of novel assessment constructs. It is now possible to recognize the need for a new assessment dimension, such as the Coronavirus Behavioral Health Mindset (a measure of biosafety in the workplace during the pandemic; Cunningham et al., 2021), develop the items, obtain samples of relevant online respondents, and analyze the data for reliability and concurrent validity within four to six weeks, or even faster. This type of high-speed prototyping and development must be included in the design process to satisfy today's clients.

***Megatrend 3: Leverage disruptive testing technologies and radical innovations****Leverage radical innovations*

A radical innovation does not need to be technological in nature, but it can be enabled by technology. Unproctored preemployment testing with responses made on a phone was once a radical innovation. Research showed that for occupational personality tests, there were no significant differences between taking an unproctored test remotely and online versus taking a

proctored test in an HR office. Due to the greater risk of cheating on knowledge tests, and as another example of radical innovation, however, both certification and licensure exams are now increasingly using proctored home-based testing with surveillance cameras and analytics instead of only offering high stakes exams in a secure physical test center. When designing an assessment, decisions need to be made about the level of innovation that an end user not only desires but can accommodate with their current or affordably expanded infrastructure.

#### *Embrace disruptive, deep technologies*

Sackett et al. (2023) documented the high validity of structured interviews. But the labor-intensive nature of interviewing has led to efforts to digitally automate them, despite the loss of the high-touch interaction element. By incorporating automatic data recording and more standardization, including AI-assisted follow-up questions, a digital structured interview begins to resemble a standardized test. Therefore, it is important that Sackett et al. (2023) mentioned innovative digital interviewing methodology, but there is a wide variety of new, disruptive technology busting out in the marketplace. The next wave of personnel selections systems hitting the marketplace employ artificial intelligence, virtual/extended/augmented realities, Internet-of-things, biotechnology, and quantum computing, among other “deep” technological advances. All of these technological advances are becoming major design considerations. The design of deep technology solutions must, however, consider digital equity and inclusion, that is, ensuring equitable access to all high-end technologies and information that are part of the assessment value chain.

#### *Ensure legal and professional compliance*

A primary requirement, even for disruptive testing technologies and innovations, is that they comply with legal and professional requirements. The marketplace is not solely concerned with the sizzle and awe associated with the new offerings, but with the need for assurance that they will not cause harm. For example, ethical artificial intelligence algorithms must not exclude a disproportionate percentage of any protected group (Behrens et al., 2021).

Although local validation studies, fairness studies, and privacy compliance audits are needed with these tools, the market will not wait for meta-analytic results. Intelligent choices will be made, and there will be winners and losers with such deep-tech risk-taking, but the meta-analytic support will probably be narrower and more assessment-specific rather than broad and all-encompassing. The degree of disruptive technology that will be included in the design plan needs to be determined based on the client’s appetite for risk and the availability of the technology for real-world application. In addition, any credible design initiatives must address new technology-focused guidelines released by organizations like SIOP and the Association of Test Publishers, in addition to the EEOC, to ensure proper risk management with end user clients.

## **Conclusion**

Both Schmidt and Hunter (1998) and Sackett et al. (2023) documented useful levels of validity for a wide variety of selection approaches. In designing a selection system in a post-pandemic economy, however, practitioners must embrace multifaceted design science and consider more than just validity. A narrow focus on the prediction of overall work performance with an emphasis on using those tools that have the highest validity is probably a relic of the past.

**Competing interest.** The authors have no known conflict of interest and published a wide range of employment tests and interviews.

## References

- Behrens, G., Cunningham, M. R., Jones, J. W., Becker, P., & Thiemann, A.** (2021, Sept.). Practicing ethical algorithms in assessment applications: Scientific, operational, and legal perspectives. European Association of Test Publishers Annual Conference, Virtual. ATP Media.
- Billings, W. W., & Jones, J. W.** (2019). *Strategies for the continued use of personnel assessments given the current ultra-tight labor market*. Fifth Theory Technical Report.
- Casillas, A., Forte, E., & Lai, E.** (2019, March). Shifting to a design science paradigm to develop desirable, feasible, and viable tests. Workshop presented at the Annual Conference of the Association of Test Publishers, Orlando, Florida.
- Cunningham, M. R., Druen, P. B., Barbee, A. P., Jones, J. W., & Dreschler, B. W.** (2021). Causes and consequences of a coronavirus behavioral health mindset: Demographics, personality, occupational interest and COVID-19 prevention behaviors. *Basic and Applied Social Psychology*, *43*(2), 120–140.
- Dages, K., & Jones, J.W.** (2022). *Addressing test mastery mindset and deep psychological underminers of testing performance among disadvantaged test takers*. Fifth Theory White Paper.
- Dages, K., & Jones, J.W.** (2023). *Development and validation of an analogue measure of neurodiversity among high-stakes test takers: Implications for certification and licensure exams*. Technical Report.
- Dreschler, B.W., Behrens, G., Cunningham, M.R., & Jones, J.W.** (2023). *Development and validation of a trauma-informed leadership scale for the Campbell Leadership Index-90 (CLI-90TIL)*. Fifth Theory, LLC.
- Jones, J. W., Orban, J., Dages, K., & McHenry, R.** (2016, March). Design considerations for express scales in a digital world. Symposium presented at the Annual Conference of the Association of Test Publishers, Orlando, Florida.
- Sackett, P. R., Zhang, C., Berry, C. M., & Lievens, F.** (2023). Revisiting the design of selection systems in light of new findings regarding the validity of widely used predictors. *Industrial and Organizational Psychology: Perspectives on Science and Practice* *16*(3), 283–300. [https://ink.library.smu.edu.sg/lkcsb\\_research/7188](https://ink.library.smu.edu.sg/lkcsb_research/7188)
- Schmidt, F. L., & Hunter, J. E.** (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, *124*(2), 262–274.
- Weiner, J., Morin, M., Wiley, A., & Muckle, T.** (2022, October). Candidate experience in online testing: Measurement issues, research, and implications for practice. Institute for Credentialing Excellence's EXG 2022 Conference, Savannah, Georgia.