The Unrealized Potential of Technology in Selection Assessment

Ann Marie Ryan

Michigan State University

Eva Derous

Ghent University

Author Note

Correspondence may be directed to Ann Marie Ryan, Department of Psychology,

Michigan State University, 316 Physics Drive, East Lansing, MI 48824. ryanan@msu.edu

Keywords: selection, assessment, technology

Abstract

Technological advances in assessment have radically changed the landscape of employee selection. This paper focuses on three areas where the promise of those technological changes remains undelivered. First, while new ways of measuring constructs are being implemented, new constructs are not being assessed, nor is it always clear what constructs the new ways are measuring. Second, while technology in assessment leads to much greater efficiency, there are also untested assumptions about effectiveness and fairness. There is little consideration of potential negative byproducts of contextual enhancement, removing human judges, and collecting more data. Third, there has been insufficient consideration of the changed nature of work due to technology when assessing candidates. Virtuality, contingent work arrangements, automation, transparency, and globalization should all be having greater impact on selection assessment design. A critique of the current state of affairs is offered and illustrations of future directions with regard to each aspect is provided.

Keywords: selection, assessment, technology

The Unrealized Potential of Technology in Selection Assessment

Over the past two decades, assessment for employee selection has undergone radical changes due to technological advances (see Tippins, 2015, for a review). Examples include changes in assessment delivery (computerized, online, mobile, use of adaptive testing), changes in assessment content (greater use of video and audio, graphics, gamification), changes in interactivity (recording of applicants, video interviewing), changes in scoring and reporting (quicker, internet delivered), and changes in many other aspects (e.g., use of algorithms and data mining). The purpose of this paper is not to provide another review of these advances as quality treatments exist (Scott, Bartram, & Reynolds, 2017; Tippins, 2015) but to discuss 3 ways incorporation and use of technological advances in assessment practice and research could be strengthened.

First, we discuss the need to think more creatively regarding what is measured. That is, despite the promise of technology as a means of measuring new KSAOs (i.e., knowledge, skills, abilities, and other constructs like motivation) and richer content, many of our assessments are just "transported" traditional tools. As an example, gamified assessments (Bhatia & Ryan, 2018) most often are measuring traditional constructs in environments with game elements layered on.

Second, we discuss the need to move our focus from efficiency to effectiveness and fairness in thinking about the use of technology (a point raised by Ryan & Derous, 2016). As examples, the use of adaptive testing, mobile testing, and video interviewing have all been undertaken because they make the hiring process more efficient (for both employer and applicant), but the investigation of factors that impact their effectiveness is still quite nascent. Similarly, the use of algorithms and social network information in hiring (e.g., Stoughton, Thompson, & Meade, 2015) has wide appeal as an efficient means of screening, but the evidence

suggests these methods are not always more effective than more traditional approaches to assessment of the same constructs.

Finally, the need to adapt assessments for how technology has changed the nature of work should be a consideration in our exploration of new approaches. For example, how are we thinking about assessment differently due to the changing nature of work arrangements to shorter engagements and contract work (Spreitzer, Cameron, & Garrett, 2017)? Are we assessing the skills needed for virtual workplaces (e.g., Schulze, Schultze, West, & Krumm, 2017)? How has the shift to online talent pools and talent badges changed the "business model" underlying the assessment industry (Chamorro-Premuzic, Winsborough, Sherman, & Hogan, 2016)? What are validation strategies that work with these new models? These changes are fueled by changes in the nature of work and employment relationships because of technology; our assessment research and practice also needs to consider how technology has changed work, not just how it has changed assessment.

New Construct Measurement

Much has been written about how technology has changed the way we assess individuals in hiring contexts (Reynolds & Dickter, 2017; Scott & Lezotte, 2012; Tippins, 2015). Some of the advances and advantages noted have been the efficiencies and cost savings that accompany web-based delivery of assessments and adaptive testing technologies (see Scott et al., 2017, for a book-length discussion). Technological advances are also credited with enhancing the candidate experience through providing more convenient and engaging assessments (e.g., use of video, gamification).

However, of all the promises of technology promoted in the past decade, we would argue the one which has been under-delivered is that of harnessing technology to assess new and

different attributes. The anticipation was that these technological changes in delivery and the use of greater context would open the door for measuring things that could not be easily measured before. We would argue that while *new ways of measuring* constructs are being implemented, *new constructs* are not being measured. Further in some cases, new ways of measuring are implemented but *what constructs* are being measured is unclear.

New ways of more easily assessing established content include using drag-and-drop matching items, image matching, and hotspot items (e.g., pointing to location of photo), using mobile functionality (e.g., swiping), recording video answers to interview or SJT questions, and scraping social media or other data sources (Dickter, Jockin, & Delany, 2017; Kantrowitz & Gutierrez, 2018). These new ways, however, are typically employed in service of assessing traditional constructs (e.g., verbal skills, quantitative reasoning). Adler, Boyce, and Caputo, (2018) note the majority of internet-delivered cognitive ability tests ask the traditional multiple choice questions on traditional content (e.g., chart and paragraph interpretations). Indeed, Chamorro-Premuzic et al. (2016) discuss that "new talent signals" such as social media and big data are often still looking at the same essential attributes of inter- and intra-personal competencies, abilities, and willingness to work hard. As another example, serious games or gamified assessments are often described as novel, but many are not assessing new constructs, just multiple traditional constructs simultaneously (with the concomitant concerns then about the ability to assess many things well in limited time periods). As Bhatia and Ryan (2018) noted, there is very little unpublished or published research on the validity evidence in support of games and gamified assessments in selection, particularly on construct validity, something that needs to be addressed.

Behaviors such as mouse-over hover times, response latencies, eye tracking, measurement of facial micro-expressions, and biometric sensors assessing emotions are all being implemented (Reynolds & Dickter, 2017). The question of the construct validity of these measures seems to receive a more superficial treatment. Have we thought about **what** they signify or **why** they might relate to valued criteria? That is, the constructs being assessed need to be more clearly specified.

What are some examples of a new "what" that could be assessed? Adler et al. (2018) discuss leveraging the interactive capabilities of technology to assess the rate of speed in acquiring new knowledge (i.e., learning agility), but note that this has not been done yet to efficient levels that would fit a hiring context. As a second example they note that natural language processing technologies and capabilities to interpret visual input could lead to new assessments of personality of a more projective nature (i.e., assessments that involve responses to ambiguous or unstructured stimuli), tapping into less conscious motivations and tendencies, but as yet these have not been adopted for wide use. An example of where steps are being taken to measure new constructs is in assessing emotions, such as through applications of databases of micro-expressions (Yan, Wang, Liu, Wu, & Fu, 2014) and speech patterns into digital interviewing. Yet, even here, the detection of micro-expression and speech mining are discussed as an advance, but that is a focus on the "how" (i.e., tools or methods to assess) and not the "what" (i.e., constructs to be assessed). Tying these tools or methods more clearly to job-relevant constructs (e.g., specific personality traits, emotional regulation concepts) is important for the purpose of validation, explainability, and acceptability.

Realizing the potential for selection assessment to add something new to what we already assess in hiring should start with the development of predictive hypotheses based on job analytic

information (Guion, 2011). That is, what KSAOs are important to work outcomes that we have not historically assessed because it was deemed too difficult and too time-consuming? This is where artificial intelligence (AI) might be leveraged in a less atheoretical fashion for assessing prior behavior from existing records. As an example, the field of learning analytics has led those in higher education to track all kinds of student behaviors (e.g., class attendance, course-taking patterns, performance on varied types of assessments). Could such data be used to go beyond GPA as an indicator of past learning ability and allow for assessment of more specific learning capabilities and motivational constructs not typically assessed in selection contexts? Another example would be to drill down into our job analytic results to specify what dynamic capabilities are required for a given job (e.g., adaptation to certain types of changes or events) and leverage the interactive nature of technology-enabled assessments to measure these more contextualized types of adaptive behavior.

As our next section discusses in more detail, are our technology-enhanced assessments **better** than what we had before or just different? Perhaps we are not assessing new things, but that the same old things are now assessed in more efficient (and effective) ways.

Enhanced Efficiency, Effectiveness, and Fairness

There is certainly ample evidence that technology has increased efficiency through increasing ease of access, shortening times of assessment and scoring, allowing greater use of multiple assessment types, reducing costs in delivery and scoring, and other process improvements (Adler et al., 2018). One early promise of technology was the ability to increase reliability and validity in assessment. A closer look as to whether this is being fully realized is warranted. We illustrate this with discussions of several assumptions that are unwarranted: 1) as context and richness increases, validity increases, 2) as efficiency in delivery increases,

effectiveness is unharmed, 3) as people have less of a role in the process of assessing, validity and fairness increases, and 4) as the quantity of data and information considered increases, validity and fairness increases.

Effects of Contextual Enhancement

One assumption associated with technologically-enhanced assessment is that creating more realistic items will more closely mimic what actually occurs at work and will therefore improve the accuracy of our measurement. However, when we transform our items from written descriptions to video items or other media rich depictions, we may actually be adding more noise and even systematic error (Hawkes, Cek, & Handler, 2018).

As an example, video-SJTs are viewed as an improvement in effectiveness over written SJTs because of the decreased reliance on reading comprehension as well as increased engagement of test takers (Chan & Schmitt, 1997; Jones & DeCotiis, 1986; Lievens & Sackett, 2006). However, video SJTs can have an additional source of non-construct relevant variance in scores, as actors in videos vary in gender, ethnicity, age and other factors. Individuals process demographic cues quickly and automatically (Ito & Urland, 2003), and this information can influence behavioral responses (e.g., Avery, McKay, & Wilson, 2008; Eagly & Crowley, 1986; Kunstman & Plant, 2008; Perkins, Thomas, & Taylor, 2000; Russell & Owens, 2001). Research on video SJTs has found that Black respondents perform better on an SJT when the videos include Black actors (Golubovich & Ryan, 2012) and White respondents have been found to react less favorably to a hypothetical organization after viewing its SJT videos featuring Black actors (Golubovich & Ryan, 2013).

These studies provide an example of how adding context or "situation" to our measures of "person" have effects. Besides how actors and avatars look, one can imagine effects for all

the nonverbal information that is conveyed in these scenarios. The question is whether the

technology advance over a written text for a question increases accuracy by assessing behavior in

context or adds information that affects measurement accuracy in negative ways. For example,

virtual reality (VR) can certainly transform assessments to feel more "real" for participants

(Reynolds & Dickter, 2017), but is a VR simulation a **better** measure of the targeted constructs?

Adler et al. (2018) suggest that a VR simulation of public speaking with a virtual audience might

be a better assessment than a more low-fidelity simulation of these skills; perhaps, but we do not

know. We would need to draw on existing research regarding audience presence to develop

clear hypotheses and test them.

  As another example, video-resumes can provide more personalized, job–relevant

information for use in screening, which should in theory result in less bias in information

processing. However, these methods also provide more social category information than

traditional resumes and more non-job related information, which can creates a cognitive

challenge for a rater (Apers & Derous, 2017). The category cues dominate our perceptual

systems and attract our attention and hence have a greater probability of being processed more

deeply (Kulik, Roberson, & Perry, 2007).

  As a final example, Arthur, Doverspike, Kinney, and O'Connell (2017) provide a strong

note of caution regarding game-thinking in selection contexts, pointing out that job candidates

are likely already highly motivated and so the enhanced engagement which game elements are

meant to deliver may not have appreciable effects. Further, game mechanics related to the value

of feedback may have negative consequences such as increased anxiety (Arthur et al., 2017).

Hawkes et al. (2018) go so far as to note that the reliability of game assessments may be affected

negatively by practice effects such as those seen in the gaming world or by adding noisy variance

associated with hand-eye coordination and mouse control. One direction for the future will be to show that all of this effort focused on enhancing fidelity and engagement adds something in terms of validity and that "something" added is not more error in measurement.

Effects of Efficiency

We should not assume that moving to a more "advanced" technology necessarily increases effectiveness even if it increases efficiency. This is clearly evident in the trend toward shorter assessments (Hardy, Gibson, Sloan, & Carr, 2017; Kruyen, Emons, & Sijtsma, 2012, 2013) where efficiency gains may result in reliability and validity decrements.

One stream of research focused on how more efficient technology advancements may distort what is measured is that on assessment mode equivalence. There is ample research available to indicate that some measures are equivalent across modes (non-cognitive tests; see review by Tippins, 2015) and others are not (e.g., speeded cognitive tests; King, Ryan, Kantrowitz, Grelle, & Dainis, 2015; Mead & Drasgow, 1993). For example, non-cognitive measures may be equivalent when moving from PC to mobile assessment but cognitive measures and SJTs might not, depending on features like scrolling etc… (see Arthur et al., 2017; King et al., 2015 for comparisons). Morelli, Potosky, Arthur, and Tippins (2017) note that "reactive equivalence" studies comparing assessment modes are not theoretically (or even practically) informative as they do not address why mode differences occur or the reasons for the construct non-equivalence. Frameworks such as Potosky (2008) and Arthur, Keiser, and Doverspike (2018) indicate ways in which different assessment delivery devices might differ; these need to be expanded and tested to more systematically understand whether and why more efficient delivery of assessments might be a more rather than less accurate way to measure (see Apers & Derous, 2017, for an example in the context of resume screening). Morelli et al. (2017) make a

strong case for better theory-based predictions that go beyond considering variance associated with technology use as "construct irrelevant." We echo their general call for elevating efforts regarding construct specification and evaluating technological feature influence in a more considered fashion.

As another example of this, while the potential for greater accuracy in measurement lies in use of adaptive testing, many computerized tests are not adaptive (Adler et al., 2018). Wider adoption will occur when we can scale up item banks more quickly and refresh them without onerous effort – those in educational testing are very much experimenting with ways to create item clones and variations more efficiently (e.g., using automatic item generation; see Drasgow & Olson-Buchanan, 2018; Kantrowitz & Gutierrez, 2018). Effectiveness of adaptive methods can be severely hampered by small item pools that lack adequate numbers of items at certain trait or ability levels, a not uncommon occurrence. As a final example, the use of an open badges system in Belgium (Derous, 2019) significantly shortened recruitment and selection procedures by allowing candidates with badges (for having passed assessments with another organization) to bypass retesting. Time of both the employer and applicant are saved, and in this case there is no observable negative impact on quality. Taken together, efficiency of assessment can mean more valid assessment, but it can also mean faster and cheaper but not better: it is important to recognize and address which is occurring.

Effects of Removing Judgment

Another broad assumption is that technology increases the accuracy of measurement as we pull humans out of the process. That is, by removing human administration, scoring, and judgment, a reduction in error in measurement is expected (e.g., structuring interviews can improve reliability and validity). However, this is not always the case. For example, recently

Facebook was questioned regarding bias against women as job ads were targeted toward certain demographic groups and not others (Scheiber, 2018; see also Datta, Tschantz, & Datta, 2015, and Sweeney, 2013, for similar examples). Similarly, news stories regarding Amazon's abandonment of AI screening of resumes because of biases (Meyer, 2018) as well as questions regarding video interview technology and bias continue (see Buolamwini, 2018; McIllvaine, 2018). Finally, a number of researchers have noted the potential biases that may emerge in using social media screening of job candidates, leading to both perceived unfairness (Stoughton et al., 2015) and actual discrimination (Van Iddekinge, Lanivich, Roth, & Junco, 2016). Because existing databases can capture historical biases, the need to consider how technological advances may tap into those biases needs to be at the forefront of assessment design considerations, rather than assuming that the computer is a better judge (e.g., by training the algorithm to make non-discriminatory decisions; Ajunwa, 2016).

As another example, Cascio and Montealegre (2016) noted that we might want to consider the changes to the role of the recruiter due to constant connectivity. Research focused on greater efficiency in reaching candidates has not always considered concurrent positive and negative effects on the internal organizational members such as recruiters, interviewers and selection system administrators. Rather than assuming that technology has made organizational members' lives easier, we might consider how their efficiency and effectiveness is both enhanced and burdened by technological innovations in assessment (e.g., candidates are processed more efficiently but the greater volume negates any gains in work time for recruiters).

The transfer of assessment tasks and roles to technology, as with any other automation in the workplace, theoretically removes the low skills, tedious parts of the job and frees people up to focus on the more creative tasks and the interpersonal elements of the work (Brynjolfsson &

McAfee, 2014). We need to attend to ways this is true for those involved in assessment as well as counterinfluences that decrease effectiveness in other aspects of assessment professionals' work.

Effects of More Data

The research on text analysis, natural language processing and social media scraping similarly raises questions regarding the effectiveness of efficient methods. While unobtrusive measurement has many advantages in terms of ease and efficiency (and many potential concerns regarding privacy and information control), a key for assessment experts is always validity of inferences. As Dickter et al. (2017) note, data science experts do not approach modeling considering the concept of construct validity.

Research on validity of social media data has been mixed. Connections to job performance have been found in several studies, e.g., based on Facebook ratings (e.g., Kluemper, Rosen, & Mossholder, 2012; Kluemper & Rosen, 2009) and – more recently- also based on LinkedIn ratings (e.g., Roulin & Levashina, 2018). Van Iddekinge et al. (2016), however, sounded a strong cautionary note, as they found that recruiter ratings of Facebook pages were unrelated to job performance or turnover and also had adverse impact (i.e., lower scores or selection rates for underrepresented groups; Guion, 2011). This line of research on validity and social media scraping is a good illustration of the work that needs to be done to build our understanding of when, where, why and how technology-enabled assessment tools are effective and when they do not fulfill their potential.

Indeed, as we alluded to earlier, the notion that bias is reduced and fairness perceptions increased when one relies on larger quantities of data has been shown to be a fallacy (see for example, Caliskan, Bryson, & Narayanan, 2017). Algorithms can and do build in existing biases when the data they are trained on may have been derived from a biased judgment process and/or

biased labor market. The lack of an ability to articulate to job candidates exactly what is being assessed and why makes big data use in selection decision-making a challenge from a fairness perceptions perspective as well.

Our examples in these last few sections should not be construed as indicating that technology-enabled assessment is bound to be less effective than traditional methods; indeed, we believe that validity and fairness can be enhanced through technology. Rather, we seek to emphasize the importance of not conflating efficiency and effectiveness when discussing the value of technology in assessment, and the need for more thoughtful, theory-driven examinations of when effectiveness is likely to be enhanced and when it is not.

Changing Nature of Work Requires Changing Assessments

Besides focusing on how technology changes assessments, we should focus on how technological changes to work itself should lead us to change what is assessed. To illustrate, we discuss how the virtuality, contingency, automation, transparency, and globalization of work (in some sense all byproducts of technological change) should be impacting selection assessment.

Virtuality

Discussions on the changing nature of work often focus on how individuals are much more likely to be working remotely and collaborating virtually (Brawley, 2017; Spreitzer et al., 2017). The literature on team virtuality has suggested that competencies for virtual collaborative work and using computer-mediated communication may differ in depth and complexity from those associated with face-to-face teamwork (see Schulze & Krumm, 2017, for a review; also Hassell & Cotton, 2017; Schulze et al., 2017). Beyond offering assessments in digital formats, are assessment designers considering more carefully what is being measured for virtual work?

KSAOs such as awareness of media capabilities, communication style adaptability, and other

"virtual skills" should be a greater focus of assessment developers.

As another example, in reviewing the effects of telecommuting, Allen, Golden, and

Shockley (2015) point out individual differences that relate to successful telecommuting (i.e.,

moderators of productivity as well as social isolation and satisfaction effects such as self-

management skills, personality characteristics, boundary management styles). Assessments

related to capability and satisfaction with greater levels of remote work may be valuable for

hiring contexts with greater remoteness.

Contingency

Today's workers are said to be more often employed in short-term engagements (gigs) or

contract work rather than long-term, traditional employee/employer arrangements. The

implications of this increased contingency of work (George & Chattopadhyay, 2017; Spreitzer et

al., 2017) should be given greater consideration. For example, does this mean administering

assessments more often as people move from contract to contract? Or does it imply less

assessing as organizations do not invest in evaluating KSAOs, relying on crowdsourced ratings

and rankings to evaluate talent (Aguinis & Lawal, 2013)? The use of online talent exchanges as

a means of securing employment should be forcing a consideration of what is assessed, as well

as when and how assessment for gigs and short term contracts may look different than that for

longer-term employment relationships.   While there is a need for validation work for

determining what works in predicting performance and other outcomes in these arrangements, as

Brawley (2017) notes, we first need better theory and empirical evidence regarding what relates

to the attitudes and behavior of "serious" gig workers.

In one effort at expanding theory regarding contingent workers, Petriglieri, Ashford, and Wrzesniewski, (2018) have provided a framework for understanding how some individuals are successful at managing the uncertainty of a freelance career. For example, they discuss how those who were able to cultivate established routines for their workdays are more effective than others in managing the greater ambiguity and lack of structure of these careers. Assessments to help individuals understand the predispositions and competencies required for success as a gig worker may be useful tools for platforms that seek to connect workers to work, even if they are not used as selection tools by employers but as self-selection tools used by gig workers.

Automation

As automation of jobs increases (Metz, 2018; Scheiber, 2018a; Wingfield, 2017), we need to consider what else should enter into the selection process. Automation has led to a decrease in lower skill jobs (Nedelkoska & Quintini, 2018), so perhaps some of the high volume selection assessments that are a big share of the assessment market (e.g., basic math skills) will have reduced demand, and the need for assessment of higher level skills (e.g., advanced math skills) will emerge as a larger focus. In general, we tend to focus our assessment development efforts on high volume entry level jobs that may be automated in the near future; as work changes, what is assessed may need to change.

As technology changes jobs, some have also asked what new skills should be assessed. As an example of one way change can impact assessment, research on the social acceptance of technology (Gaudiello, Zibetti, Lefort, Chetouani, & Ivaldi, 2016; Seo et al., 2018; Syrdal, Dautenhahn, Koay, & Walters, 2009) examines how people can collaborate with robots as coworkers (cobots). The questions as to how the form and level of teamwork skills such as

communication, collaboration, and conflict management might differ when coworkers are cobots, and what that means for assessment has not been explored.

One might also consider how robotic interfaces within the assessment process might affect tool validity. For example, in many online assessments, individuals interact with avatars. One can envision future assessment center role-players that are humanoid robots, freeing up assessor time and enhancing consistency across candidates. Rather than assuming that interaction with a "simulated other" is akin to an actual interpersonal interaction, understanding the acceptance of and skills for interacting with automation can aid us adopt technology enhancements to assessment processes that consider these issues of social acceptance.

Transparency

Another example of a workplace trend affecting assessments is the move of organizations toward greater transparency toward both external and internal stakeholders (Parris, Dapko, Arnold, & Arnold, 2016). In the selection context, this involves greater transparency regarding what is being assessed and why. Most of the research related to transparency in selection has suggested positive effects on validity and candidate experience (e.g., Klehe, König, Richter, Kleinmann, & Melchers, 2008; Kleinmann, Kuptsch, & Köller, 1996; Kolk, Born, & der Flier, 2003). However, Jacksch and Klehe (2016) demonstrated that transparency's positive effects are limited to nonthreatening performance dimensions; that is, transparency can benefit some candidates and harm others if the attribute being assessed is associated with a negative stereotype related to the social identity of those being assessed. Langer, König, and Fitili (2018) demonstrated that providing greater transparency regarding what is assessed in an online interview by an avatar (e.g., facial expression, gestures, voice pitch) had equivocal effects on organizational attractiveness, with individuals appreciating the organization's candor but

simultaneously decreasing their views of the organization overall. Langer et al. (2018) note that *what* information and *how much* information to provide regarding technology-enhanced assessments is deserving of greater research focus in this time of greater pressures for transparency.

Globalization

The globalization of business also has implications for assessment (Ryan & Ployhart, 2014). Not only does it mean that assessments must be delivered in multiple languages, and that all the ensuing efforts to ensure psychometric equivalence occur (see International Test Commission, 2005, for standards), but it means all the associated implementation issues related to cross-cultural implementation be considered (see Ryan & Tippins, 2012, for a comprehensive discussion; see Fell, König, & Kammerhoff, 2016, for a specific example of a cross-cultural differences in faking in interviews; see Ryan & Delany, 2017, for a discussion of recruiting globally). Also, the need to assess cultural competencies is amplified, like cultural intelligence (Ang et al., 2007), cultural values (Hofstede, 2001), cultural adjustment (Salgado & Bastida, 2017), and leadership and teamwork in multinational teams (Han & Beyerlein, 2016).

We also note that prior research has indicated there can be cross-cultural differences in acceptance of and skill in interacting with technology (Dinev, Goo, Hu, & Nam, 2009; Nistor, Lerche, Weinberger, Ceobanu, & Heymann, 2014). In the selection assessment space, we need to be more mindful of how technological change in the workplace more broadly is varied across cultures – in both adoption rates and in user acceptance – and how that might impact selection tool effectiveness.

Conclusion

The premise of this paper is that there is a lot of unrealized potential in the incorporation of technology into selection assessments. Figure 1 summarizes in a schematic where we believe that the (unrealized) potential of technology in selection assessment is currently situated and themes from this paper that both researchers and practitioners might focus on. First, as regards the predictor side of selection assessment, we have noted that advances in delivery, scoring, interactivity, and reporting have been implemented with a focus on greater efficiency as well as job applicant engagement and other stakeholder reactions. However, we see the potential for leveraging technology to assess new constructs as well as to increase validity as not yet fulfilled and we urge greater focus and energy in these directions. We also have admonished those who ignore the potential downsides of technology-enhanced assessments such as greater introduction of construct-irrelevant variance or instances of unfairness to disadvantaged groups.   Of course, technology can and should continue to serve as a means of enhancing assessment efficiency.

Second, we urge selection researchers, assessment developers, and other practitioners to take a closer look at the criterion side, i.e., how work and the workplace is changing, and seek ways to better align assessment content (i.e., KSAs measured) as well as practice (e.g., transparency) with those changes (see Figure 1). While we have highlighted some of the more prevalent trends in the changing nature of work (virtuality, automation, contingency, globalization and transparency), there are likely other emerging changes that also can serve as inspiration for developing new assessments, leveraging the advantages offered by technology.

The future of technology-enabled assessments is only limited by imagination; we anticipate continued changes in the way in which assessments are developed, delivered, and scored. More importantly, we hope that the next decade will have a greater focus on the areas we have outlined

in this paper as needing attention, as this would ensure the realization of the full potential of

technology-enabled assessments.

References

Adler, S., Boyce, A. S., & Caputo, P. M. (2018). Employment testing. In J. C. Scott, D. Bartram, & D. H. Reynolds (Eds.), *Next generation technology-enhanced assessment: Global perspectives on occupational and workplace testing* (pp. 3–35). Cambridge, UK: Cambridge University Press.

Aguinis, H., & Lawal, S. O. (2013). eLancing: A review and research agenda for bridging the science–practice gap. *Human Resource Management Review*, *23*(1), 6–17. doi:10.1016/j.hrmr.2012.06.003

Ajunwa, I. (2016). Hiring by algorithm. *SSRN Electronic Journal*. doi:10.2139/ssrn.2746078

Allen, T. D., Golden, T. D., & Shockley, K. M. (2015). How effective is telecommuting? Assessing the status of our scientific findings. *Psychological Science in the Public Interest*, *16*(2), 40–68. doi:10.1177/1529100615593273

Ang, S., Van Dyne, L., Koh, C., Ng, K., Templer, K., Tay, C., & Chandrasekar, N. (2007). Cultural intelligence: Its measurement and effects on cultural judgment and decision making, cultural adaptation, and task performance. *Management and Organization Review*, *3*, 335–371. doi:10.1111/j.1740-8784.2007.00082.x

Apers, C., & Derous, E. (2017). Are they accurate? Recruiters' personality judgments in paper versus video resumes. *Computers in Human Behavior*, *73*, 9–19. doi:10.1016/j.chb.2017.02.063

Arthur, W. J., Doverspike, D., Kinney, T. B., & O'Connell, M. (2017). The impact of emerging technologies on selection models and research: Mobile devices and gamification as exemplars. In J. L. Farr & N. T. Tippins (Eds.), *Handbook of employee selection* (2nd ed., pp. 967–986). New York: Routledge.

Arthur, W., Keiser, N. L., & Doverspike, D. (2018). An information-processing-based

    conceptual framework of the effects of unproctored internet-based testing devices on scores

    on employment-related assessments and tests. *Human Performance*, *31*(1), 1–32.

    doi:10.1080/08959285.2017.1403441

Avery, D. R., McKay, P. F., & Wilson, D. C. (2008). What are the odds? How demographic

    similarity affects the prevalence of perceived employment discrimination. *Journal of

    Applied Psychology*, *93*(2), 235–249. doi:10.1037/0021-9010.93.2.235

Bhatia, S., & Ryan, A. M. (2018). Hiring for the win: Game-based assessment in employee

    selection. In J. H. Dulebohn & D. L. Stone (Eds.), *The brave new world of eHRM 2.0* (pp.

    81–110). Charlotte, NC: Information Age Publishing.

Brawley, A. M. (2017). The Big, Gig Picture: We Can't Assume the Same Constructs Matter.

    Industrial and Organizational Psychology, *10*(04), 687–696. doi:10.1017/iop.2017.77

Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and

    prosperity in a time of brilliant technologies*. New York: W.W. Norton.

Buolamwini, J. (2018). When the robot doesn't see dark skin. *NY Times*. Retrieved from

    https://www.nytimes.com/2018/06/21/opinion/facial-analysis-technology-bias.html

Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from

    language corpora contain human-like biases. *Science*, *356*(6334), 183–186.

    doi:10.1126/science.aal4230

Cascio, W. F., & Montealegre, R. (2016). How technology is changing work and organizations.

    Annual Review of Organizational Psychology and Organizational Behavior, *3*(1), 349–375.

    doi:10.1146/annurev-orgpsych-041015-062352

Chamorro-Premuzic, T., Winsborough, D., Sherman, R. A., & Hogan, R. (2016). New Talent

Signals: Shiny New Objects or a Brave New World? *Industrial and Organizational Psychology*, *9*(03), 621–640. doi:10.1017/iop.2016.6

Chan, D., & Schmitt, N. (1997). Video-based versus paper-and-pencil method of assessment in situational judgment tests: Subgroup differences in test performance and face validity perceptions. *Journal of Applied Psychology*, *82*(1), 143–159. doi:10.1037/0021-9010.82.1.143

Datta, A., Tschantz, M. C., & Datta, A. (2015). Automated experiments on ad privacy settings. *Proceedings on Privacy Enhancing Technologies*, *2015*(1). doi:10.1515/popets-2015-0007

Derous, E. (2019). Van boe-boe machine tot sociale media: Evidence-based werven en selecteren [From reaction time tests to social media: Evidence-based recruitment and selection]. In J. Valk & L. Lopes de Leao Laguna (Eds.), *HRM heden en morgen: Evidence-based practice & Practice-based evidence* (pp. 1–25). Amsterdam: Vakmedianet Management.

Dickter, D. N., Jockin, V., & Delany, T. (2017). The evolution of E-selection. In *The Wiley Blackwell Handbook of the Psychology of the Internet at Work* (pp. 257–283). Chichester, UK: John Wiley & Sons, Ltd. doi:10.1002/9781119256151.ch13

Dinev, T., Goo, J., Hu, Q., & Nam, K. (2009). User behaviour towards protective information technologies: The role of national cultural differences. *Information Systems Journal*, *19*(4), 391–412. doi:10.1111/j.1365-2575.2007.00289.x

Drasgow, F., & Olson-Buchanan, J. B. (2018). Technology-driven developments in psychometrics. In J. C. Scott, D. Bartram, & D. H. Reynolds (Eds.), *Next generation technology-enhanced assessment: Global perspectives on occupational and workplace testing.* (pp. 239–264). Cambridge, UK: Cambridge University Press.

Eagly, A. H., & Crowley, M. (1986). Gender and helping behavior: A meta-analytic review of

the social psychological literature. *Psychological Bulletin*, *100*(3), 283–308. doi:10.1037/0033-2909.100.3.283

Fell, C. B., König, C. J., & Kammerhoff, J. (2016). Cross-cultural differences in the attitude toward applicants' faking in job interviews. *Journal of Business and Psychology*, *31*(1), 65– 85. doi:10.1007/s10869-015-9407-8

Gaudiello, I., Zibetti, E., Lefort, S., Chetouani, M., & Ivaldi, S. (2016). Trust as indicator of robot functional and social acceptance. An experimental study on user conformation to iCub answers. *Computers in Human Behavior*, *61*, 633–655. doi:10.1016/j.chb.2016.03.057

George, E., & Chattopadhyay, P. (2017). *Understanding nonstandard work arrangements: Using research to inform practice. SHRM-SIOP Science of HR Series*. Retrieved from http://www.siop.org/SIOP-SHRM/2017_03_SHRM-SIOP_Nonstandard_Workers.pdf

Golubovich, J., & Ryan, A. M. (2012). *Demographic cues in video-based situational judgment items. Symposium presented at the Annual Meeting of the Society for Industrial and Organizational Psychology*. San Diego, CA.

Golubovich, J., & Ryan, A. M. (2013). *Demographic cues in video-based situational judgment items: An extension. Poster presented at the annual meeting of the Society for Industrial and Organizational Psychology*. Houston, TX.

Guion, R. M. (2011). *Assessment, measurement, and prediction for personnel decisions*. New York: Routledge.

Han, S. J., & Beyerlein, M. (2016). Framing the effects of nultinational cultural diversity on virtual team processes. *Small Group Research*, *47*(4), 351–383. doi:10.1177/1046496416653480

Hardy, J. H., Gibson, C., Sloan, M., & Carr, A. (2017). Are applicants more likely to quit longer

assessments? Examining the effect of assessment length on applicant attrition behavior. *Journal of Applied Psychology*, *102*(7), 1148–1158. doi:10.1037/apl0000213

Hassell, M. D., & Cotton, J. L. (2017). Some things are better left unseen: Toward more effective communication and team performance in video-mediated interactions. *Computers in Human Behavior*, *73*, 200–208. doi:10.1016/j.chb.2017.03.039

Hawkes, B., Cek, I., & Handler, C. (2018). The gamification of employee selection tools: An exploration of viability, utility, and future directions. In J. C. Scott, D. Bartram, & D. H. Reynolds (Eds.), *Next generation technology-enhanced assessment: Global perspectives on occupational and workplace testing.* (pp. 288–316). Cambridge, UK: Cambridge University Press.

Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors, institutions, and organizations across nations*. Thousand Oaks, CA: Sage Publications.

International Test Commission. (2005). *Guidelines on computer-based and internet delivered testing*. Retrieved from www.intestcom.org

Ito, T. A., & Urland, G. R. (2003). Race and gender on the brain: Electrocortical measures of attention to the race and gender of multiply categorizable individuals. *Journal of Personality and Social Psychology*, *85*(4), 616–626. doi:10.1037/0022-3514.85.4.616

Jacksch, V., & Klehe, U.-C. (2016). Unintended consequences of transparency during personnel selection: Benefitting some candidates, but harming others? *International Journal of Selection and Assessment*, *24*(1), 4–13. doi:10.1111/ijsa.12124

Jones, C., & DeCotiis, T. A. (1986). Video-Assisted Selection of Hospitality Employees. *Cornell Hotel and Restaurant Administration Quarterly*, *27*(2), 67–73. doi:10.1177/001088048602700222

Kantrowitz, T. M., & Gutierrez, S. L. (2018). The changing landscape of technology-enhanced

   test administration. In J. C. Scott, D. Bartram, & D. H. Reynolds (Eds.), *Next generation*

   technology-enhanced assessment: Global perspectives on occupational and workplace

   testing (pp. 193–2015). Cambridge, UK: Cambridge University Press.

King, D. D., Ryan, A. M., Kantrowitz, T., Grelle, D., & Dainis, A. (2015). Mobile internet

   testing: An analysis of equivalence, individual differences, and reactions. *International*

   Journal of Selection and Assessment, *23*(4), 382–394. doi:10.1111/ijsa.12122

Klehe, U.-C., 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., Kuptsch, C., & Köller, O. (1996). Transparency: A necessary requirement for

   the construct validity of assessment centres. *Applied Psychology*, *45*(1), 67–84.

   doi:10.1111/j.1464-0597.1996.tb00849.x

Kluemper, D. H., & Rosen, P. A. (2009). Future employment selection methods: evaluating

   social networking web sites. *Journal of Managerial Psychology*, *24*(6), 567–580.

   doi:10.1108/02683940910974134

Kluemper, D. H., Rosen, P. A., & Mossholder, K. W. (2012). Social networking websites,

   personality ratings, and the organizational context: More than meets the eye? *Journal of*

   Applied Social Psychology, *42*(5), 1143–1172. doi:10.1111/j.1559-1816.2011.00881.x

Kolk, N. J., Born, M. P., & der Flier, H. van. (2003). The transparent assessment centre: The

   effects of revealing dimensions to candidates. *Applied Psychology*, *52*(4), 648–668.

   doi:10.1111/1464-0597.00156

Kruyen, P. M., Emons, W. H. M., & Sijtsma, K. (2012). Test length and decision quality in

personnel selection: When is short too short? *International Journal of Testing*, *12*(4), 321–344. doi:10.1080/15305058.2011.643517

Kruyen, P. M., Emons, W. H. M., & Sijtsma, K. (2013). On the shortcomings of shortened tests: A literature review. *International Journal of Testing*, *13*(3), 223–248. doi:10.1080/15305058.2012.703734

Kulik, C. T., Roberson, L., & Perry, E. L. (2007). The multiple-category problem: Category activation and inhibition in the hiring process. *Academy of Management Review*, *32*(2), 529–548. doi:10.5465/AMR.2007.24351855

Kunstman, J. W., & Plant, E. A. (2008). Racing to help: Racial bias in high emergency helping situations. *Journal of Personality and Social Psychology*, *95*(6), 1499–1510. doi:10.1037/a0012822

Langer, M., König, C. J., & Fitili, A. (2018). Information as a double-edged sword: The role of computer experience and information on applicant reactions towards novel technologies for personnel selection. *Computers in Human Behavior*, *81*, 19–30. doi:10.1016/j.chb.2017.11.036

Lievens, F., & Sackett, P. R. (2006). Video-based versus written situational judgment tests: A comparison in terms of predictive validity. *Journal of Applied Psychology*, *91*(5), 1181–1188. doi:10.1037/0021-9010.91.5.1181

McIllvaine, A. R. (2018). In the fight against bias, AI faces backlash. Retrieved from www/hrexecutive.com

Mead, A. D., & Drasgow, F. (1993). Equivalence of computerized and paper-and-pencil cognitive ability tests: A meta-analysis. *Psychological Bulletin*, *114*(3), 449–458. doi:10.1037/0033-2909.114.3.449

Metz, C. (2018). FedEx follows Amazon into the robotic future. *New York Times*. Retrieved

    from https://www.nytimes.com/2018/03/18/technology/fedex-robots.html

Meyer, D. (2018). Amazon reportedly killed an AI recruitment system because it couldn't stop

    the tool from discriminating against women. Retrieved from

    http://forutne.com/2018/10/10/amazon-ai-recruitment-bais-women-sexist/

Morelli, N., Potosky, D., Arthur, W., & Tippins, N. (2017). A call for conceptual models of

    technology in I-O Psychology: An example from technology-based talent assessment.

    Industrial and Organizational Psychology, *10*(04), 634–653. doi:10.1017/iop.2017.70

Nedelkoska, L., & Quintini, G. (2018). *Automation, skills use and training.* (No. No. 202). Paris.

    doi:10.1787/2e2f4eea-en

Nistor, N., Lerche, T., Weinberger, A., Ceobanu, C., & Heymann, O. (2014). Towards the

    integration of culture into the Unified Theory of Acceptance and Use of Technology. *British

    Journal of Educational Technology*, *45*(1), 36–55. doi:10.1111/j.1467-8535.2012.01383.x

Parris, D. L., Dapko, J. L., Arnold, R. W., & Arnold, D. (2016). Exploring transparency: A new

    framework for responsible business management. *Management Decision*, *54*(1), 222–247.

    doi:10.1108/MD-07-2015-0279

Perkins, L. A., Thomas, K. M., & Taylor, G. A. (2000). Advertising and recruitment: Marketing

    to minorities. *Psychology and Marketing*, *17*(3), 235–255.

Petriglieri, G., Ashford, S. J., & Wrzesniewski, A. (2018). Thriving in the gig economy. *Harvard

    Business Review*, 140–143. Retrieved from https://hbr.org/2018/03/thriving-in-the-gig-

    economy

Potosky, D. (2008). A conceptual framework for the role of the administration medium in the

    personnel assessment process. *Academy of Management Review*, *33*(3), 629–648.

doi:10.5465/amr.2008.32465704

Reynolds, D., & Dickter, D. N. (2017). Technology and employee selection: An overview. In J.

L. Farr & N. T. Tippins (Eds.), *Handbook of employee selection* (2nd ed., pp. 855–873).

New York, NY: Routledge.

Roulin, N., & Levashina, J. (2018). LinkedIn as a new selection method: Psychometric

properties and assessment approach. *Personnel Psychology*. doi:10.1111/peps.12296

Russell, A., & Owens, L. (2001). Peer estimates of school-aged boys' and girls' aggression to

same- and cross-sex targets. *Social Development*, *8*(3), 364–379. doi:10.1111/1467-

9507.00101

Ryan, A. M., & Delany, T. (2017). Attracting job candidates to organizations. In J. Farr & N.

Tippins (Eds.), *Handbook of employee selection* (2nd ed., pp. 165–181). Routledge.

Ryan, A. M., & Derous, E. (2016). Highlighting tensions in recruitment and selection research

and practice. *International Journal of Selection and Assessment*, *24*(1), 54–62.

doi:10.1111/ijsa.12129

Ryan, A. M., & Ployhart, R. E. (2014). A Century of selection. *Annual Review of Psychology*,

65(1), 693–717. doi:10.1146/annurev-psych-010213-115134

Ryan, A. M., & Tippins, N. (2012). *Designing and implementing global selection systems*.

Oxford, UK: Wiley-Blackwell.

Salgado, J. F., & Bastida, M. (2017). Predicting expatriate effectiveness: The role of personality,

cross-cultural adjustment, and organizational support. *International Journal of Selection

and Assessment*, *25*(3), 267–275. doi:10.1111/ijsa.12178

Scheiber, N. (2018a). High-skilled white-collar work? Machines can do that, too. *New York

Times*. Retrieved from https://www.nytimes.com/2018/07/07/business/economy/algorithm-

fashion-jobs.html

Scheiber, N. (2018b, September 18). Facebook accused of allowing bias against women in job

ads. *NY Times*. Retrieved from

https://www.nytimes.com/2018/09/18/business/economy/facebook-job-ads.html

Schulze, J., & Krumm, S. (2017). The "virtual team player." *Organizational Psychology Review*,

7(1), 66–95. doi:10.1177/2041386616675522

Schulze, J., Schultze, M., West, S. G., & Krumm, S. (2017). The knowledge, skills, abilities, and

other characteristics required for face-to-face versus computer-mediated communication:

Similar or distinct constructs? *Journal of Business and Psychology*, *32*(3), 283–300.

doi:10.1007/s10869-016-9465-6

Scott, J. C., Bartram, D., & Reynolds, D. H. (2018). *Next Generation Technology-Enhanced

Assessment*. (J. C. Scott, D. Bartram, & D. H. Reynolds, Eds.). Cambridge, UK: Cambridge

University Press.

Scott, J. C., & Lezotte, D. V. (2012). *Web-Based Assessments*. Oxford University Press.

doi:10.1093/oxfordhb/9780199732579.013.0021

Seo, S. H., Griffin, K., Young, J. E., Bunt, A., Prentice, S., & Loureiro-Rodríguez, V. (2018).

Investigating people's rapport Building and hindering behaviors when working with a

collaborative robot. *International Journal of Social Robotics*, *10*(1), 147–161.

doi:10.1007/s12369-017-0441-8

Spreitzer, G. M., Cameron, L., & Garrett, L. (2017). Alternative work arrangements: Two

images of the new world of work. *Annual Review of Organizational Psychology and

Organizational Behavior*, *4*(1), 473–499. doi:10.1146/annurev-orgpsych-032516-113332

Stoughton, J. W., Thompson, L. F., & Meade, A. W. (2015). Examining Applicant Reactions to

the Use of Social Networking Websites in Pre-Employment Screening. *Journal of Business and Psychology*, *30*(1), 73–88. doi:10.1007/s10869-013-9333-6

Sweeney, L. (2013). Discrimination in online ad delivery. *Queue*, *11*(3), 10. doi:10.1145/2460276.2460278

Syrdal, D. S., Dautenhahn, K., Koay, K. L., & Walters, M. L. (2009). The Negative Attitudes towards Robots Scale and reactions to robot behaviour in a live human-robot interaction study. In N. Taylor (Ed.), *Adaptive and Emergent Behaviour and Complex Systems: Proceedings of the 23rd Convention of the Society for the Study of Artificial Intelligence and Simulation of Behaviour* (pp. 109–115). Edinburgh, UK: Society for the Study of Artificial Intelligence and Simulation for Behavior (AISB). Retrieved from http://www.scopus.com/inward/record.url?partnerID=yv4JPVwI&eid=2-s2.0-84859046918&md5=0901ed76f558c614f66619620d76878d

Tippins, N. T. (2015). Technology and Assessment in Selection. *Annual Review of Organizational Psychology and Organizational Behavior*, *2*(1), 551–582. doi:10.1146/annurev-orgpsych-031413-091317

Van Iddekinge, C. H., Lanivich, S. E., Roth, P. L., & Junco, E. (2016). Social media for selection? Validity and adverse impact potential of a facebook-based assessment. *Journal of Management*, *42*(7), 1811–1835. doi:10.1177/0149206313515524

Wingfield, N. (2017, September 10). As Amazon pushes forward with robots, workers find new roles. *New York Times*. Retrieved from https://www.nytimes.com/2017/09/10/technology/amazon-robots-workers.html

Yan, W.-J., Wang, S.-J., Liu, Y.-J., Wu, Q., & Fu, X. (2014). For micro-expression recognition: Database and suggestions. *Neurocomputing*, *136*, 82–87 doi:10.1016/j.neucom.2014.01.029

Efficiency
- Big data / algorithms
- Fast assessment
- Reduced assessment cycles
- Worldwide assessment

Effectiveness
- Theory-driven assessment
- Reliability of assessment
- Validity of assessment
- Utility of assessment

Fairness
- Perceived procedural fairness (e.g., face validity)
- Actual, distributive fairness (e.g., adverse impact)

Potential of Technology in Selection Assessment

PREDICTOR SIDE

New ways of assessing KSAOs
- Physiological vs. Psychological vs. behavioral performance indicators
- Unobtrusive vs. obtrusive assessment
- Unproctored vs. proctored assessment
- Virtual vs. non-virtual assessment

CRITERION SIDE

New types of jobs and work arrangements
- Importance of 'traditional' KSAOs
- 'New' types of KSAOs
- Technological developments in the work environment

Virtuality
- Telecommunication
- Virtual work teams

Contingency
- Gig workers
- Freelancing

Automation
- Lower vs. higher level skills to assess
- Tech acceptance

Transparency
- What is assessed
- Why is it assessed

Globalization
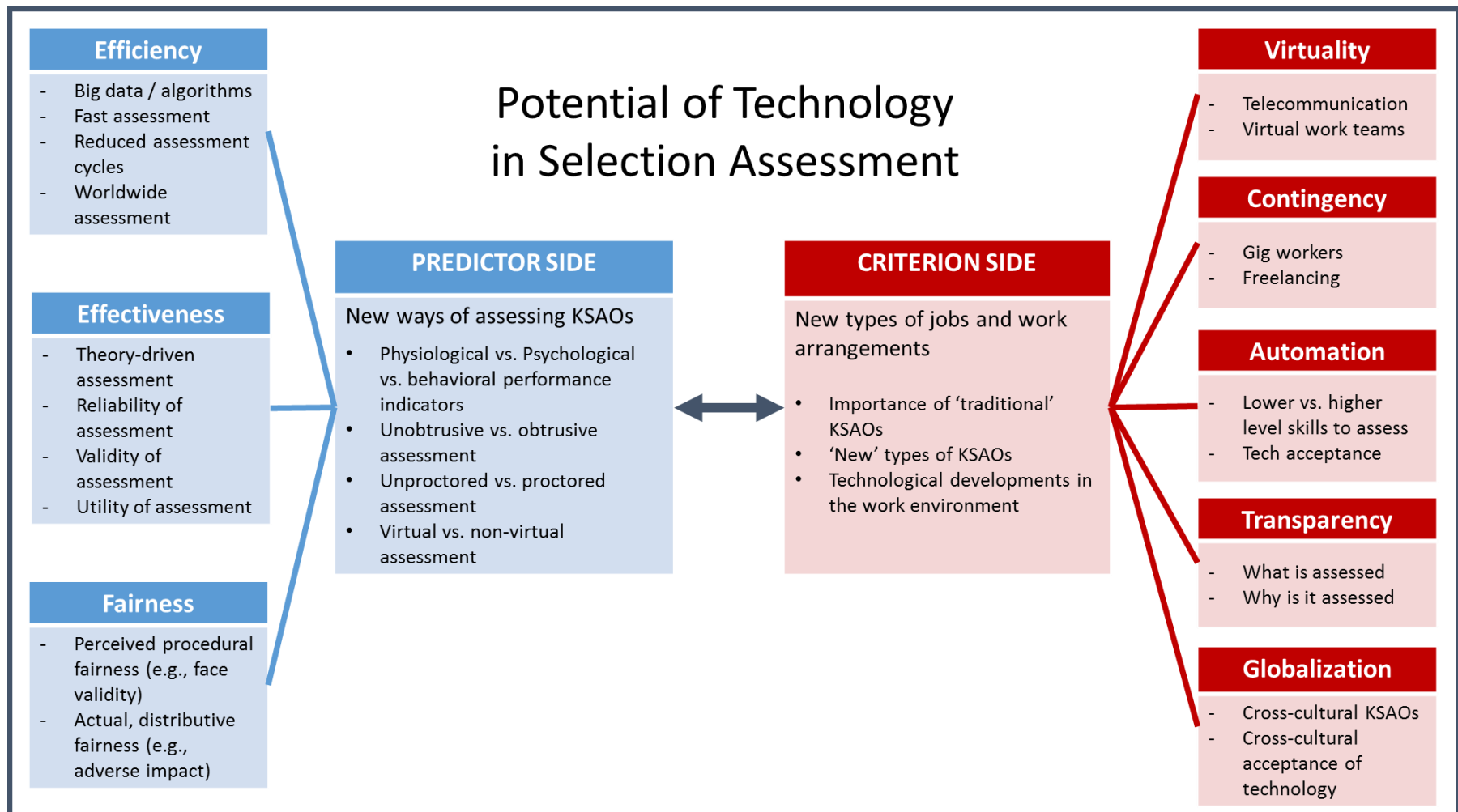- Cross-cultural KSAOs
- Cross-cultural acceptance of technology

Figure 1. The Unrealized Potential of Technology in Selection Assessment