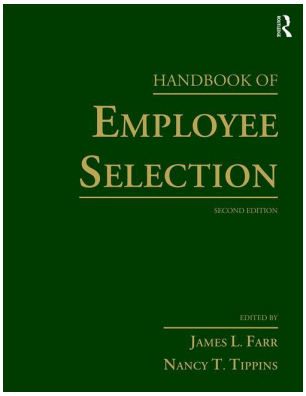

Handbook of Employee Selection

James L. Farr, Nancy T. Tippins, Walter C. Borman, David Chan, Michael D. Coovert, Rick Jacobs, P. Richard Jeanneret, Jerard F. Kehoe, Filip Lievens, S. Morton McPhail, Kevin R. Murphy, Robert E. Ployhart, Elaine D. Pulakos, Douglas H. Reynolds, Ann Marie Ryan, Neal Schmitt, Benjamin Schneider

The Impact of Emerging Technologies on Selection Models and Research

Publication details
https://test.routledgehandbooks.com/doi/10.4324/9781315690193-44
Winfred Arthur, Dennis Doverspike, Ted B. Kinney, Matthew O'Connell
Published online on: 22 Mar 2017

44

The Impact of Emerging Technologies on Selection Models and Research

Mobile Devices and Gamification as Exemplars

WINFRED ARTHUR JR., DENNIS DOVERSPIKE, TED B. KINNEY, AND MATTHEW O'CONNELL

INTRODUCTION

To stay current, assessment professionals must track developments in a plethora of emerging technologies including mobile assessment, gamification, serious games, simulations, social media, artificial intelligence, avatars, and Big Data. The introduction of each new technology seems to result in a similar cycle of calls for research and validation efforts, studies on equivalence, and the publication of findings in journals or presentations at conferences. This is accompanied by much brow beating regarding the lack of impact of research on practice, the lag between the adoption of the technologies and scientific publications, and the lack of impact of the academy on practice.

Of course, there will *always* be emerging technologies, some of which may have implications for employment-related testing, assessment, and research, and others which will not; a cynic might argue that this has always been true and that once upon a time we worried about OpScan sheets or computer anxiety. In writing this chapter, we are cognizant of the likelihood that any discussion of the topic of "emerging" technologies could very well be outdated soon after it appeared in print. Therefore, instead of a discussion of the whole gamut of emerging technologies, this chapter focuses on mobile devices and gamification as exemplars of the interplay between the emergence of new technologies and the practices and methods of the fields of personnel assessment and industrial-organizational (I-O) psychology. Specifically, we examine and explore the extent to which emerging technologies may lead to disruptive innovations in the way the field conceptualizes and implements the traditional methods and approaches to test development and validation; that is, will any of these technologies alter the basic psychometric tools of our profession?

Three themes characterize discussions about emerging and new technologies with organizational stakeholders: (1) Many organizations want *all* of these trending topics applied to their selection program (e.g., Hypothetical client question: "Are you able to provide my organization with an avatar-based simulation with artificial intelligence and game-like features that can

be administered on any Internet device . . . Oh, and it would be helpful if you can also farm big data sets to get reliable measures of traits from Facebook posts. Can you do that?"); (2) There is very little agreement on how any of these trends are defined or how effectively they can be applied in a selection context (e.g., Hypothetical client request: "I don't really know what gamification is, but I know that Nike and Walmart do it, so I need to have gamification incorporated into all of my talent strategies too"; and (3) Empirical research from I-O psychology lags practice on each of these topics. In fact, to date, only six empirical investigations of assessments delivered via mobile assessment have been published in typical I-O or other related applied journals (Arthur, Keiser, & Doverspike, 2017), and we were unable to locate any published empirical investigations of the use of gamification in employment-related testing and assessment. Despite this absence of empirical research, Dale (2014) had projected that organizations will allocate more than $2.8 billion in spending on gamification by 2015. So, with such a high level of organizational interest in the use of emerging technology in talent acquisition and interventions, there is clearly both value and need for personnel psychology to devote some research attention to these topics in an effort to not only keep up with but also get ahead of these trends.

Subsequent sections of this chapter first present a review of the literature on the selected exemplars, specifically mobile devices and gamification, including games and simulations, as it pertains to employment-related testing and assessment. Next, we present a discussion of the traditional test development and validation model and its intersection with said emerging technologies. Finally, the chapter concludes with a discussion of recommendations, the need for research, other emerging technologies, and some future-oriented speculation.

MOBILE DEVICES AND GAMES: REVIEW OF THE LITEATURE AND IMPLICATIONS FOR EMPLOYMENT-RELATED TESTING AND ASSESSMENT

Delivering Assessments on Mobile Devices

For years there have been serious concerns in the I-O community about unproctored Internet testing (UIT). There were legitimate concerns about test security, equivalence, and cheating (Pearlman, 2009; Tippins et al., 2006). Although some of these concerns remain, especially in high-stakes testing situations such as certification tests, research has consistently failed to show practical or meaningful differences between UIT and non-UIT tests and assessments in reference to psychometric properties, test score validity, or candidate reactions (Davies & Waddlington, 2006; Do, Shepherd, & Drasgow, 2005; O'Connell, Delgado, & Kung, 2012).

It is acknowledged that UIT is here to stay (O'Connell, Arthur, & Doverspike, 2015), and the advent of mobile devices has made it even easier for candidates to take tests anywhere, anytime, and almost exclusively in unproctored environments. Usage data suggest that test taking on mobile devices continues to increase substantially (e.g., see Illingworth, Morelli, Scott, & Boyd, 2015; McClure Johnson & Boyce, 2015). This increase in usage gives more people than ever the opportunity to apply for jobs, reduces testing-related costs for organizations, and also increases the size of the applicant pool, thereby resulting in smaller selection ratios, which favor the hiring organization. A pivotal issue in this research domain is "What is and is not a *mobile* device and what theories or constructs would even lead us to expect that this differentiation should affect outcomes that are psychologically interesting and meaningful?"

In an effort to provide theoretical guidance to inform why Internet-based testing (IT) device-type (i.e., "mobile" vs. non-mobile) should or should not have an effect on test scores, Arthur et al. (2017) presented a framework for conceptualizing device types in terms of the construct-irrelevant information processing demands placed on the test taker while taking the assessment. Said information processing demands translate into additional, construct-irrelevant cognitive load, which interacts with the device type, resulting in differential outcomes as a function of the construct assessed. So, instead of differentiating mobile and non-mobile devices simply in terms of whether said devices are tethered to the wall or not (i.e., wireless vs. wired connection

to the Internet), Arthur et al. (2017) identified four information processing variables—working memory, perceptual speed and visual acuity, psychomotor ability, and selective attention—that correspond to four structural characteristics of IT assessment devices, specifically screen size, screen clutter, response interface, and permissibility (i.e., distractibility). Arthur et al.'s Structural Characteristics/Information Processing (SCIP) model permits the classification of current IT devices on a continuum that ranges from desktops at one end (i.e., large screen, low clutter, easy response interface, and low permissibility, which translates into lower construct-irrelevant cognitive load) to smartphones at the other (small screen, relatively high clutter, difficult response interface, and high permissibility, which translates into high construct-irrelevant cognitive load). Thus, when the literature uses the label *mobile device*, in terms of Arthur et al.'s SCIP model, this refers to devices at the high end of the information processing continuum, which definitely includes smartphones but may also include tablets as well, which are lower than smartphones on Arthur et al.'s continuum.

Measurement Equivalence

The issue of interest here is whether the psychometric properties of a test administered on a mobile device are similar to those administered on a non-mobile device such as a desktop computer. A large-scale, high-stakes study comparing mobile to non-mobile devices with a sample of 2.8 million applicants (approximately 49,000 of whom used mobile devices) found comparable reliabilities, factor loadings, and intercorrelations for cognitive and non-cognitive measures (Arthur, Doverspike, Muñoz, Taylor, & Carr, 2014). Similar findings are reported in other studies as well (Lawrence, Wasko, Delgado, Kinney, & Wolf, 2013; Morelli, Mahan, & Illingsworth, 2014; Parker & Meade, 2015). In summary, the vast majority of research, using large sample sizes, suggests that the psychometric properties, including factor structure and reliability, are similar for mobile and non-mobile devices when assessments are intentionally designed to be administered across devices. These findings are the case for the measurement of *both* cognitive and non-cognitive constructs.

Mean Differences

Mean score differences between mobile and non-mobile devices appears to be a function of the constructs assessed. Thus, a robust finding that characterizes this literature is that whereas there are no mean differences on non-cognitive assessements (e.g., personality) taken on mobile and non-mobile devices (e.g., Arthur et al., 2014; Dages & Jones, 2015; Morelli et al., 2014; Wood, Stephens, & Sliter, 2015), there are pronounced differences for cognitive constructs, with scores on mobile devices being consistently and substantially lower. For instance, Arthur et al. (2014) reported a *d* of .90. Impelman (2013) found similar performance decrements on cognitive measures across four organizational samples. Wood et al. (2015) report *d*s of .46 and .35 for two cognitive ability tests and .93 and .26 for two mechanical aptitude tests. Finally, to the extent that UIT devices are used to take assessments in the form of complex interactive simulations and situational judgment tests (SJTs), assessments that generally engender higher construct-irrelevant cognitive load, one would expect lower scores on mobile devices compared to non-mobile devices.

The preceding pattern of findings are in accord with the percepts of Arthur et al.'s (2017) model in that to the extent that the four information processing variables (i.e., working memory, perceptual speed and visual acuity, psychomotor ability, and selective attention) that correspond to four structural characteristics of UIT assessment devices (i.e., screen size, screen clutter, response interface, and permissibility [distractibility]) play a role in using the UIT device, they then result in additional construct-irrelevant cognitive load that is likely to influence performance on the test when said cognitive demands are not the focal construct of interest.

Criterion-Related Validity

Research on the criterion-related validity of mobile device assessments, and more importantly, compared to non-mobile assessments, is almost non-existent. However, a limited number of studies have examined the comparative criterion-related validity of proctored versus unproctored assessments, and their findings indicate little if any differences (Beaty et al., 2011; Wasko, Lawrence, & O'Connell, 2015; Weiner & Morrison, 2009). So, although the volume of research is quite small, the preceding lends credence to the proposition that there is little theoretical or conceptual basis to expect differential criterion-related validity in the comparisons of mobile versus vs. non-mobile Internet devices (Kinney, Chang, Lawrence, & Moretti, 2015; O'Connell et al., 2015).

Demographic Differences in Usage

The research to date suggests that African Americans, Hispanics, and females are more likely than white males to take a test on a mobile device (Arthur et al., 2014; Illingworth et al., 2015; McClure Johnson & Boyce, 2015). If taking cognitive tests on mobile devices results in lower scores, and the tendency to take assessments on mobile devices covaries with specified protected group status, then this raises the specter of observed subgroup differences and higher adverse impact potential resulting from the use of mobile devices in employment-related assessments. However, there is no research that we are aware of that has examined this issue for cognitive constructs. That being said, a detailed look at Arthur et al.'s (2014) data suggests that the mobile device effect appears to be a main effect; that it does not appear to interact with demography to result in larger subgroup differences. This pattern of results appears to be similar to those reported by Arthur, Edwards, and Barrett (2002), and Edwards and Arthur (2007) in their comparisons of constructed-response and multiple-choice tests. Finally, it should be noted that for non-cognitive constructs, the absence of meaningful subgroup differences reported in the general personnel selection and assessment literature is observed for mobile device assessments as well (e.g., Golubovich & Boyce, 2013; Kinney, Lawrence, & Chang, 2014; McClure Johnson & Boyce, 2015).

Applicant Reactions

Assessment professionals and organizations generally consider providing applicants with the opportunity to take tests on mobile devices to be a positive attribute (Fursman & Tuzinski, 2015; Gutierrez, Meyer, & Fursman, 2015). However, it is unclear whether applicants experientially actually prefer to take assessments on mobile devices over desktops or personal computers (PCs). So, for instance, although Kinney et al. (2014) found no difference in applicant satisfaction based on mode of delivery, other researchers have found applicants to have much more favorable reactions to PCs than mobile devices (Fursman & Tuzinski, 2015; Gutierrez & Meyer, 2013; Landers, Reddock, Cavanaugh, & Proaps, 2014). Hence, whereas applicants generally indicated that test takers should be given the opportunity to complete assessments on mobile devices, they also generally had more negative reactions to using mobile devices for assessments and consistently expressed a preference for PCs over mobile devices in taking personnel selection tests and assessments.

In summary, a number of conclusions and recommendations can be made concerning the use of mobile devices in personnel selection and assessment. First, the growth in unproctored mobile device testing continues to display an accelerating upward trend. Second, it is important to improve the experience for applicants by designing mobile tests in an optimized manner (e.g., maximizing the use of screen space, limiting unnecessary buttons, etc.). Third, the permissibility of mobile devices (i.e., the ability to use them in a variety of locations and conditions) means that they also potentially engender high levels of distractibility, which may be a contributory

factor in the lower scores observed for cognitive constructs (Arthur et al., 2017). Consequently, organizations and testing and assessment professionals should consider instructing and encouraging candidates to take control of their test environment and make sure they are free of distractions during the assessment. Fourth, because the use of mobile devices generally results in substantially lower scores on cognitive constructs, research that directly investigates differential subgroup differences on mobile versus vs. non-mobile assessments of cognitive constructs is needed. Fifth, comparative criterion-related validity studies are woefully absent in the literature—even conference presentations.

Finally, with very few exceptions (e.g., Arthur et al., 2017), at present most research uses a simple classification of mobile versus non-mobile device, with a very small number of recent studies recognizing distinctions between PCs versus tablets versus smartphones. Consequently, future research needs to pay closer attention to finer device-type designations. So, for instance, on the basis of their two dimensions (structural characteristics and information processing variables), Arthur et al. (2017) currently place UIT devices on the following continuum: desktops→laptops→tablets→phablets→smartphones, with desktops engendering the lowest levels of construct-irrelevant cognitive load and smartphones engendering the highest levels of construct-irrelevant cognitive load. In conclusion, the growth of mobile device testing poses a number of challenges but at the same time opens up a wide range of exciting opportunities for reaching non-traditional candidates, expanding the applicant pool, and also increasing the potential to reduce testing-related costs for organizations, especially those pertaining to test administration.

Gamification (and Serious Games and Simulations)

Prevalence of Gamification and Game-Thinking in Organizations

Game-thinking is a term that has been used to broadly present the concepts of gamification and serious games (Amstrong, Collmus, & Landers, 2015). Gamification has been embraced as a common technique to facilitate change in organizations by making traditional interventions more engaging. Many uses of game-thinking are regularly applied in organizations for a host of purposes, including recruitment, training, sales prospecting, professional development, and performance reviews (Oprescu, Jones, & Katsikitis, 2014). Starting with e-learning systems, there has been an exponential growth in the interest in gamification (Dale, 2014), with gamification appearing on *Google Trends* in 2010 (DuVernet & Popp, 2014). In 2011, and the *Oxford Dictionary* added *gamification* to its word-of-the-year shortlist, with a definition referring to the application of game features to non-game applications. The recent and projected growth in game-thinking is due to the convergence of cheaper technology and the prevalence of games in society in general (Deterding, Sicart, Nacke, O'Hara, & Dixon, 2011).

Gamification versus Serious Games versus Simulations

With the rapid growth of game-thinking in organizations and with the relative lag of scientific inquiry on these approaches, several definitions have emerged for the seemingly related concepts of *gamification*, *serious games*, and *simulations*. Most researchers broadly define *gamification* as the application of game mechanics, elements, and features to non-game environments (Attali & Arieli-Attali, 2015; Dale, 2014; Deterding et al., 2011; Gartner, 2011). For instance, Figure 44.1 presents a screenshot of a gamified assessment designed to measure attention to detail and critical thinking. Some traditional activities (e.g., assessments, surveys) in organizations are built by leveraging technology that is not particularly eye-catching or engaging, whereas games are designed to be fun. The basic concept of gamification is to apply the elements that make games interesting to non-game contexts to make them more entertaining than they would otherwise be in their traditional form (Attali & Arieli-Attali, 2015). In fact, Dale (2014) reported that the

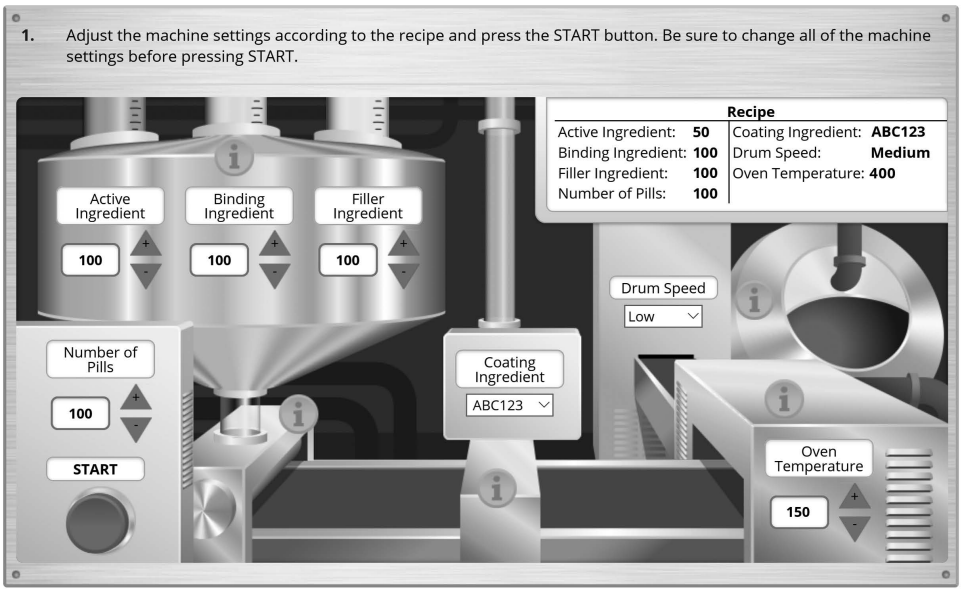

FIGURE 44.1 Screenshot of a Gamified Assessment Designed to Measure Attention to Detail and Critical Thinking Developed by Select International Inc.

Reprinted with permission

primary purpose of gamification was to *engage* participants and motivate future and lasting participation.

Interventions involving gamification are not necessarily games. Huizinga (2000) defined a *game* as a non-serious but intensely engaging voluntary activity structured by rules. Kapp (2014) explained that games are self-contained units with discrete starting points, game play sections, and clearly defined endings, with winning being a possibility. Thus, participants are aware they are playing a game.

Unlike gamification, serious games are a type of game. The concept of a serious game simply refers to the application of a game for non-trivial or non-entertainment purposes (Petridis, Baines, Lightfoot, & Shi, 2014; Simpson & Jenkins, 2015). Participants in a serious game are aware that they are in a game, there is a defined start and end point; however, the elements of fun or engagement discussed in descriptions of gamification are less relevant. Consequently, serious games and gamification have several features in common but also important differences. Gamification is a technique or collection of techniques applied to programs, assessments, or other content. A serious game is a discrete unit developed for a particular purpose. Figure 44.2 presents a screenshot of a vehicle assembly task designed as a serious game.

How do simulations relate to game-thinking? Just as serious games and gamification are not the same concept, simulations are not necessarily the same as gamification or serious games. Gamification, again, is the process of making a tool more "game-like"; a serious game is a game developed for a serious purpose. That being said, serious games can be considered a "type" of simulation. Not all simulations are serious games, but all serious games are simulations (e.g., a flight simulator with scores, levels, and objectives used to train pilots). Likewise, the process of gamification can certainly be applied (and often is) to simulations of all types. Today's organizations routinely build simulations that feature certain common gamification elements such as progress bars, timed sections, narrative stories, and challenges. Simulations also often have features common to games (as opposed to simply applying game features to an existing assessment, as is the case with gamification) in that there is typically a defined start and end. Simulations are also often created to capture either a work sample or other clearly job-related behavior in a high-fidelity, engaging user experience. A differentiator between game-thinking and simulations

FIGURE 44.2 Screenshot of a Vehicle Assembly Task Designed as a Serious Game Developed by Select International Inc.

Reprinted with permission

is that there are several core concepts to games and gamification that are typically not built into simulations, such as leaderboards, "win states," levels, and community sharing.

In sum, game-thinking is a term used to describe gamification and serious games. Gamification and serious games are different concepts used for different purposes (gamification is a process; serious games are a type of simulation). Simulations are related to both, and the use of simulations in selection contexts is increasing (see Fetzer & Tuzinski, 2013, for a comprehensive review). However, questions remain about the value of gamification in selection contexts. In particular, what key features of games are appropriate to leverage in a selection context? To consider whether or not gamification is appropriate in most selection contexts, further explanation of what gamification typically entails is needed.

Gamification Concepts

Gamification is about more than merit badges, it is about understanding, influencing, and rewarding desired behaviors (Dale, 2014). Like any effective applied psychological intervention, a gamification manipulation should focus on a specified outcome. Dale (2014) explained that good gamification design is user-centric and not mechanism-centric. Gamification is not just about adding attractive technology to an existing measure; instead, gamification manipulations should add features and elements to existing content to increase interest, engagement, and participation.

The typical desired outcome of gamification is engagement; however, whether this is of particular value in high-stakes employment testing is debatable. Nevertheless, engagement is

created by what Gartner (2011) referred to as the three Ms: Motivation (rewards—both extrinsic and intrinsic), Momentum (sustained participation often determined by the match between task difficulty and participant skill), and Meaning (which is the extent to which the outcome of the experience is desired).

Deterding et al. (2011) and Attali and Arieli-Attali (2015) explained that points, leaderboards, and badges are among the most basic elements of games and that these lead to engagement. Game characteristics include rules, tools, mechanics, and players. Rules and tools are specific to the particular game. Players, obviously, are the participants. Mechanics that are employed in a game vary, but there are consistent features and common elements such as achievements (points, levels, bonuses), exercises (challenges or quests), synchronization with the community (leaderboards), results transparency (experience bars, continuous feedback), time (countdowns, speed), and luck (lottery, random achievement). These game mechanics make up the typical "toolset" applied in gamification (Dale, 2014).

Kapp (2014) described two classes of gamification—structural and content. Structural gamification refers to the application of game elements to encourage participation through content with no actual content changes. Content gamification is the application of game elements, game mechanics, and game-thinking to alter content to make it more game-like (e.g., through the use of stories, challenges, and quests). Bailey, Pritchard, and Kernohan (2015) applied these concepts in a study on survey research in a marketing context and found that applying game elements to marketing surveys increased survey completion rates. Nevertheless, the motivational dynamics of participating in a market survey are quite different from those of completing an assessment in a high-stakes employment-related context as a job applicant.

Challenges to Applying Game-Thinking in Selection Contexts

The common theme in the gamification literature is that applying game principles leads to increased engagement. The idea is that increased engagement leads to desirable outcomes such as increased completion rates, sustained participation, and competency development or behavior change. When considering how gamification can be leveraged in selection contexts, it is important to consider whether or not these outcomes provide value to the organization. In a selection context, job candidates are typically highly motivated to engage in the selection component and pay close attention to the assessment content because the outcome of high performance (e.g., progression to the next stage in the selection process) is highly valued.

Consequently, it would seem that the primary value of gamification—engagement—is not a major or particularly important outcome in a selection context. Candidates do not need game-like interventions to motivate them to try hard; repeated participation is limited and delayed, and typically, selection assessments measure individual characteristics that are not expected to change (over and beyond measurement error) across multiple administrations for any individual candidate. As DuVernet and Popp (2014, p. 41) point out: "in an assessment application the goal is to measure a skill or characteristic rather than to train or motivate, thus repeated exposure to content or feedback may not be desirable."

The most common game mechanic used in many gamification initiatives is adding scoring, badges, rewards, or providing some other form of feedback to the participant. However, as Geimer, Sanderson, and Popp (2015) note, providing feedback to job candidates can have negative unintended consequences. Consequently, Geimer et al. (2015) warned that when gamification introduces a feedback component, negative performance information may increase anxiety, hinder concentration, and reduce the perception of having an opportunity to perform.

The Case for Using Gamification in Personnel Selection

Although there may be several arguments against using gamification in selection contexts, nevertheless some characteristics of this approach translate into some promising possibilities

for further exploration of the use of gamification in personnel selection and assessment. For instance, gamification can shift the frame of reference of the candidate to a job-relevant context by applying game mechanics, such as a work-related quest. This approach could assist the candidates in drawing on work-related past behavioral examples when responding to items. Similarly, Armstrong and Landers (2015) suggested that game-thinking may be desirable in assessment contexts if the game-thinking makes the desired behavior less transparent and susceptible to social desirability responding. Bailey et al. (2015) noted the difference between "hard" and "soft" gamification. Hard gamification refers to embedding items into gamified solutions such that participants are not aware of the items, whereas soft gamification refers to simpler interventions to "frame" items and encourage participation with item presentation features. To the extent that a hard gamification approach can be accommodated, there may be some value in gamification's ability to reduce certain socially desirable response patterns.

Other possible positive outcomes from gamification include an improved candidate experience and face validity. Armstrong et al. (2015) suggested that gamification can increase a sense of job relevance but also that applicant reactions may only be improved by gamification under certain conditions. In summary, whereas there may be some value to creating interactive and attractive assessments, simulations, and serious games, simply building a great-looking assessment is not in and of itself a gamification intervention.

What Does the Future Hold for Gamification in Personnel Selection?

The future of game-thinking in selection is difficult to foretell. To date, there has been very little empirical research on how game-thinking can add value to the selection process. In a broad review of the general gamification literature, Hamari, Koivisto, and Sarsa (2014) identified only 24 empirical studies across multiple disciplines. These studies all generally addressed whether or not gamification "works." They found that the most common application of gamification is in education and learning contexts. No empirical studies on the use of gamification in selection were identified. Geimer et al. (2015) reported that to date there is no known empirical research on gamified assessments.

As technology advances, selection tools and assessment devices will continue to become more attractive. For example, detailed high-fidelity simulations are becoming more commonplace, with a corresponding emergence of SJTs incorporating "stories" and images (Tippins, 2014; Weekley, Hawkes, Guenole, & Ployhart, 2015). These features, which enhance the look and feel of assessments, will certainly continue to be used, but these enhancements are not at the core of gamification.

Even if future research fails to indicate that gamification for assessment purposes provides a return on investment to organizations, there may be value to adding gaming elements to other human resources processes that are related to selection, such as recruiting, onboarding, and training. As such, there does appear to be a place for gamification in organizations, even if the impact on personnel assessment and selection is not an easy and natural fit.

INTERSECTION OF EMERGING TECHNOLOGIES AND THE TRADITIONAL TEST DEVELOPMENT AND VALIDATION MODEL

In the preceding section, we provided a brief overview of several exemplars of current emerging technologies. In the present section, we ask the question: "What implications do the emerging technologies have for the traditional model used in the development and validation of tests in personnel selection and assessment?" The traditional model in employee selection and placement follows a well-established sequence of steps, as illustrated in Figure 44.3. Each step in this sequence and the extent to which it is impacted by and can readily incorporate the exemplar emerging technologies that are the focus of this chapter are discussed. Table 44.1 presents a brief summary of the key features and characteristics of each step, along with the role and influence or lack thereof, of the emerging technologies of interest here.

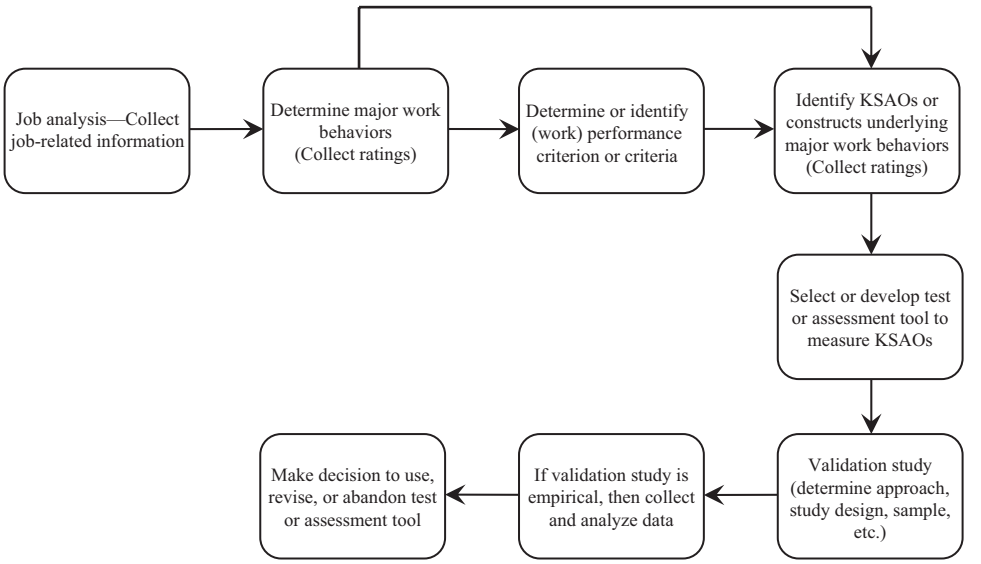

FIGURE 44.3 Prototypical Test Development and Validation Sequence. KSAOs = Knowledge, Skills, Abilities, and Other Characteristics

TABLE 44.1

Summary of Key Features and Characteristics of the Test Development and Validation Process and the Role of Simulations, Games, and Mobile Devices

Test Development and Validation Step	*Key Features and Characteristics*	*Impact of Emerging Technologies*
Work/Job Analysis		
Information gathering method	Wide range of methods available. Choice of methods is determined by practical constraints and other factors. Based on the extant research, the recommendation is to use multiple methods.	Influence is in the form of technological aids in the collection of data such as online surveys for job analysis questionnaires, video recording of performance episodes instead of live observation, and remote focus groups and interviews. Simulations and games are unlikely to impact this step. However, job analysis questionnaires could be completed on mobile devices.
Collecting ratings—rating scales	Rating data/scales pertaining to factors such as importance, time spent, frequency, consequences of error, time to proficiency, difficulty, and needed-upon-entry. Choice of scales is determined by the purpose of the job analysis and professional judgement.	Potential ease of online data collection might facilitate the collection of data and thus the number of factors assessed. Simulations and games are unlikely to impact this step. However, ratings could be completed on mobile devices. Limitations of mobile devices may impact the use of large matrices and also the length of questionnaires.

Collecting ratings—raters	Decisions about the source of ratings pertain to (a) the level of expertise [and experience], (b) the level in the organizational hierarchy [incumbents, subordinates, supervisors], and (c) the number of raters. The extant research supports the use of experts, sampling across the organizational hierarchy, and including as large a number of ratees as possible.	Potential ease of online data collection, especially if extended to mobile devices, vastly increases the ease with which larger numbers of raters can be sampled and correspondingly a wider range of rater types. However, simulations and games, are unlikely to impact this step.
Identifying KSAOs/constructs and work performance criteria	Arrived at on the basis of the statistical analysis of the job analysis ratings and also informed by the expertise, experience, and judgment of the job analyst, assessment researcher, or professional.	Gamification, may increase the need for precision and detail in the collection of KSAOs. The need for high levels of fidelity will put additional strains on the work/job analysis system. Other types of games and simulations put less emphasis on specific KSAOs, and instead use a broader, more work sample, behaviorally based approach.
Selecting, or Developing the Test or Assessment Tool	Primarily entails determining the specific method(s) or approach (e.g., interviews, SJT, work sample) to measuring the specified constructs.	Emerging technologies offer not only new methods (e.g., simulations and games) to assessing specified constructs but also different delivery platforms (e.g., mobile devices) as well.
Validation Study	Decisions pertain to choice of validation approach or source of validity evidence. If empirical (e.g., criterion-related or construct-related), then one would design and implement a research study. Some design choices include type of correlational design (e.g., predictive vs. concurrent), and sample size and type (i.e., applicants vs. incumbents).	To the extent that they are the test or assessment tool (i.e., simulations and games) or the platform via which the test is administered (i.e., mobile devices), then emerging technologies play an important role in this step since they are the source of the scores being validated.
Conclusion or Decision to Use, Revise, or Abandon Test or Assessment Tool	Primarily informed by the results of the validation study; evidence that speaks to the job-relatedness of the test or assessment scores or lack thereof.	Decisions should be carefully made by informed experts, including those trained in psychometrics and I-O psychology. Regardless of the fidelity of the assessment, we need to know whether we can make appropriate inferences regarding work performance from the obtained scores.

Work/Job Analysis

Gathering Information

Job analysis is recognized in both the scientific and professional literatures (e.g., Society for Industrial and Organizational Psychology [SIOP], 2003]) and legal guidelines (Equal Employment Opportunity Commission [EEOC], 1978) as a pivotal initial step in test development and validation. Work (job) analysis (recently broadened to include competency modeling efforts as well) is a process via which information is gathered about work and jobs with the objective of identifying and describing what incumbents do, how they do it, and the knowledge, skills, abilities, and other characteristics (KSAOs) or competencies that are required to successfully

perform said job tasks and activities. (See Chapter 6 in this volume for more details concerning work analysis.) In the implementation of a work analysis, several decisions and choices must be made. For instance, one of the first decisions is the choice of information gathering *method*. A wide range of methods is available to researchers and practitioners, ranging from interviews, observations, questionnaires, and the job analyst performing the tasks/activities to the use of materials and sources such as training materials, task inventories and checklists, employee log books and diaries, previous/old job descriptions, and the O*NET. With limited research demonstrating the superiority of one method over others, the general recommendation is to use multiple methods to permit a more complete information-gathering effort that balances the tradeoffs between the strengths and weaknesses of the various methods.

The role or influence of the emerging technologies of interest is primarily in the form of facilitating the information-gathering process. Thus, for instance, whereas gamification is unlikely to influence or play a role in the information-gathering process, mobile devices (e.g., smartphones, tablets, and even laptops) can broaden the scope of the online administration of job analysis surveys and questionnaires. Furthermore, recognizing that they might pose their own set of challenges, the video capabilities of mobile devices have the potential to permit remote job analysis interviews, focus groups, and "video job analysis" involving the recording of activities as they occur in the workplace.

Collecting Ratings

The next step in the work analysis process is typically to obtain ratings on the major work behaviors and tasks that have been identified in the preceding information-gathering step. To this end, ratings on factors such as importance, time spent, frequency, consequences of error, difficulty, and task interdependence are collected in an effort to further elucidate and refine the list of major work behaviors and tasks. Next, the KSAOs that underlie the successful performance of the major work behaviors and tasks are identified. Once again, ratings of the KSAOs (on factors such as importance and needed-upon-entry) will be obtained. Finally, as an additional step to developing the test specification plan, linkages between the KSAOs and the major work behaviors will be made, again by means of a questionnaire, resulting in a task by KSAO matrix.

Another decision in the collection of ratings pertains to the source of the ratings (i.e., the individuals who will provide the ratings). Using raters from multiple levels of the organizational hierarchy is encouraged because it permits the triangulation of the data, and thus in the aggregate, higher levels of completeness and accuracy. It should be noted that regardless of their position in the organizational hierarchy, all raters should be fairly knowledgeable about the job to provide informed ratings. Finally, because it has the additional advantage of giving every employee a voice and fosters a sense of participation, unless the sample sizes are too large to make it unmanageable, the recommendation is to sample all eligible responders (Doverspike & Arthur, 2012).

As with the information-gathering phase, gamification is unlikely to influence or play a role in the rating process. However, in contrast, once again mobile devices can broaden the scope of the rating process. Specifically, the extension of online data collection to mobile devices vastly increases the ease with which larger numbers of raters can be sampled and, correspondingly, a wider range of rater types as well. However, the small screen size of some mobile devices may limit the ability to use large linkage matrices of the type possible with traditional presentations on paper. Job analysis surveys may also have to be shortened to limit the amount of time respondents have to spend completing surveys on mobile devices.

Identifying KSAOs/Constructs to Be Assessed (and Work Performance Criteria)

In the context of personnel selection, the primary objective of the test development and validation process is the development of an assessment tool or predictor whose scores can be used to make inferences about future job/work performance. The demonstration of this then puts one

in a position to use the scores from the test or assessment tool for employment decision-making purposes. Consequently, this requires that one has a criterion against which the test is validated (i.e., the work performance criterion or criteria that one is trying to "forecast" with the use of the predictor scores). As such, one goal of the work analysis process is to determine or identify the specified work performance criterion or criteria. As illustrated in Figure 44.3, these criteria are typically the outcomes associated with the successful performance of the specified major work behaviors and tasks.

As previously noted, as part of this process, the KSAOs that underlie the successful performance of the major work behaviors and tasks are identified. Deciding on the final list of KSAOs is based on the statistical analysis of the job analysis ratings and is also informed by the expertise, experience, and judgment of the job analyst, assessment researcher, or professional (Doverspike & Arthur, 2012). For instance, on the basis of the ratings, KSAOs that are linked to low importance, and low-frequency major work behaviors and tasks, and are not needed-upon-entry would typically not be assessed. In addition, despite what the ratings may indicate, professional decisions have to be made about the psychometric and practical feasibility of measuring the specified constructs because some constructs may be more amenable to measurement (e.g., GMA) than others (e.g., "vision," "inspiration").

Gamification, especially the use of simulations and games, may increase the need for precision and detail in the collection of KSAOs, especially as developers attempt to achieve 100% physical and psychological fidelity with the actual work environment. The need for such high levels of fidelity between the assessment and the job will put additional strains on the work analysis system in terms of depth and detail, making the process more similar to those carried out in human factors investigations. At the same time, other types of games and simulations seem to put less emphasis on specific KSAOs, using a broader, more work sample and behaviorally based approach (Wernimont & Campbell, 1968).

Selecting or Developing the Test or Assessment Tool

This step entails determining the specific method(s) or approach(es) (e.g., interviews, assessment centers, SJTs, paper-and-pencil tests) to measuring the specified KSAOs or constructs. As noted by Doverspike and Arthur (2012), this may entail either selecting a previously developed assessment tool that measures the constructs/KSAOs of interest or developing an assessment tool from scratch. An important issue associated with this step of the test development and validation process is the pivotal distinction between constructs (*what* is being measured) and methods (*how* the construct is being measured; Arthur & Villado, 2008). It is important to recognize the distinction between methods, modes, and delivery platforms. So, using SJTs as an example, it is a method that can be administered in different modes (e.g., text vs. video) on different delivery platforms (e.g., desktop computer vs. smartphone). The preceding distinctions are important because they clearly highlight the fact that emerging technologies such as simulations and games offer not only new methods of assessing specified constructs but also new and different platforms via which said methods can be implemented or delivered. Interestingly, a review of the simulation and gaming literature indicates that the issues noted by Arthur and Villado (2008) characterize and are present with these methods as well. Specifically, there is an absence of attention to the specific constructs measured by these assessments with an almost exclusive focus on the methods (i.e., the simulations or games). Of course, within the context of the sign versus sample distinction (Wernimont & Campbell, 1968), an argument could be made that simulations and games are more aligned with a sample instead of sign approach to assessment, but such a position is not explicitly or clearly articulated by the developers of these assessment tools.

On a related note, the traditional test development and validation model usually entails an item writing phase where on the basis of a test specification plan, items are generated, reviewed, revised, and finalized (Doverspike & Arthur, 2012). Indeed, this focus on items serves as the basis for psychometric item analysis procedures. However, in the context of emerging technologies such as simulations and games, it is unclear what constitutes an "item" since there is not a single item, query, or problem to which the test taker provides an answer or response. For instance, in Arthur

et al.'s (2015) *Crisis in the Kodiak*, an oil-rig disaster search-and-rescue simulation, the participant's tasks are to (1) shut off four burning oil valves (50 points each), (2) locate and heal the 20 survivors on the burning oil rig (10 points for each survivor healed), and rescue the healed survivors (10 points for each survivor successfully evacuated off the oil rig). Hence, there are no traditional "items" (i.e., queries to which the participant provides an answer or response), but instead a series of tasks that must be and are completed in a fluid and dynamic fashion in a limited amount of time.

In summary, even if the overall test development and validation model remains the same, emerging technologies greatly increase the number of options for methods, modes, and delivery platforms. There is an associated increase in the possible item types and scoring methods. Unfortunately, although a large body of knowledge and guidelines now exists on developing traditional multiple-choice tests (e.g., see Haladyna & Rodriguez, 2013), we know far less about the important features impacting the design of assessments based on emerging technologies. Nevertheless, it is important that the design choices be informed by psychological theories and constructs, and not be left to the information technologists.

Validation Study[1]

Consonant with the prevailing unitarian view of validity (American Educational Research Association [AERA], American Psychological Association [APA], and National Council on Measurement in Education [NCME], 2014; SIOP, 2003), the "validation of personnel selection decisions is merely a special case of the more general validation process" (Binning & Barrett, 1989, p. 480). Hence, content-, criterion-, and construct-related validity are simply different strategies for demonstrating the construct validity of a test or measure—that is, what a test measures, how well it does so, and the accumulation of evidence that speaks to the extent to which the inferences drawn from the test scores are appropriate (Binning & Barrett, 1989). Consequently, to the extent that within the unitarian framework of validity, content-related, criterion-related, and construct-related validity are considered to be three of several evidential bases for demonstrating the construct validity of a test or measure (e.g., see AERA et al., 2014; SIOP, 2003), a decision must be made as to the most appropriate validation approach for the specified circumstances. Furthermore, if this decision results in the use of an empirical validation approach that generates a specified validity coefficient (e.g., criterion-related or construct-related), then one would design and implement an empirical correlational research study. In the subsequent implementation of such a study, some design choices will include the type of correlational design (e.g., predictive vs. concurrent) and the sample size and type (i.e., applicants vs. incumbents).

Concerning the role of emerging technologies, to the extent that they are the test or assessment tool (i.e., simulations and games) or the platform via which the test is administered (i.e., mobile devices, and maybe even Google Glass, Apple Watch, and Samsung Gear S2 in the future), then said technologies play an important albeit indirect role in this step since they are the source of the scores being validated. In the case of simulations, especially those with close to 100% physical and psychological fidelity, there may be an argument as to whether the simulated work task should serve as the predictor, criterion, or both (i.e., "perfect" overlap between the predictor and the criterion as in high-end commercial aircraft simulators). In addition, as is the case with SJTs, developers may argue that validation is not needed for simulations or games, as the development process itself guarantees job-relatedness.

We would certainly argue that our traditional validation models apply to assessments based on emerging technologies and that the input of I-O psychologists is critical in designing appropriate validation studies. Even if we found ourselves in disagreement with such a viewpoint, we would still argue for the necessity of validation based on the likely viewpoint of regulatory agencies in the United States. In particular, the existence of an *algorithm* does not preclude the need for professional involvement in the development process leading up to the creation of the algorithm, nor does it eliminate the need for the validation of inferences made based on machine-generated scores. In summary, it is our view that in terms of scientific, professional, and legal standards, the use of emerging technologies in employment-related decision making

needs to be validated and held to the same psychometric standards as any other assessment tool (AERA et al., 2014; SIOP, 2003).

Conclusion or Decision to Use, Revise, or Abandon the Test or Assessment Tool

The final step in the test development and validation model entails making a decision to use the assessment tool for employment decision-making purposes, revise or further refine the test, or abandon it. This decision is primarily informed by the results of the validation study in terms of its ability to furnish evidence that speaks to the job-relatedness of the test or assessment scores or lack thereof, but also requires an analysis of practicality, utility, the impact on protected classes, and user acceptance and reactions. As argued above, these decisions should be carefully made by informed experts, including those trained in psychometrics and I-O psychology.

The use of mobile devices has led to a variety of practical and ethical questions. In particular, consideration must be paid to security issues, including the verification of the identity of the test taker. Furthermore, because mobile devices allow for a more diverse and geographically distributed applicant pool, there are also issues of global distribution, translation, and accommodation of disabilities.

A reliance on gamification, including games and simulations, leads to its own set of potential practical concerns. This is especially true when job candidates must complete the assessments on mobile devices; some of the concerns listed above, such as translation and global distribution, may be magnified. In addition, gamification may increase the information processing load, require greater effort and time from the candidate, and increase costs to the organization.

FINAL THOUGHTS, DISCUSSION, AND CONCLUSION

This chapter started with an acknowledgment that there have always been and will always be emerging technologies. So, rather than attempt to review a laundry list of emerging technologies, we sought to examine the extent to which new technologies, as exemplified by mobile devices and gamification, are compatible with traditional approaches to developing and validating tests. Asked another way: "Do emerging technologies lead to disruptive innovations in the way we think about the traditional methods we use to develop and validate tests?" Although we may be biased by our professional affiliation, our conclusion is that existing approaches to test development and validation are still relevant and appropriate in the context of evaluating emerging technologies to assessment. That is, regardless of the fidelity of the assessment, the amount of fun created, or the technologies involved, the fundamental question remains one of whether we can make accurate inferences regarding future work performance from the scores obtained from the assessment. Emerging technologies, rather than reducing the need for validation, *increase* the importance of ensuring that decision making regarding the use of these devices and technologies for employee assessment and testing includes input from assessment experts with backgrounds in I-O psychology. Thus, the good news is that our methods, approaches, and expertise are probably more needed and relevant than ever, but we need to address the way we do, share, and communicate research. For instance, we may need to expedite our research initiation and communication cycles to keep up with the pace of technological changes and innovations. Hence, it is not surprising that most of our "emerging technology" research is more likely to be found in conference presentations (which have a short initiation-to-communication cycle) than peer-reviewed publications (e.g., see Arthur et al.'s [2017] review).

The Need for Research

One of the current oddities is that it is often easier to do field research on emerging technologies, in the case of mobile devices collecting millions of cases a year, than it is to do laboratory research. Nevertheless, we feel strongly that there is a need for cooperation among testing

companies, consultants, and academics in order to carry out well-designed laboratory research. Although we would be among the first to argue for the increased publication of practice-oriented articles, there also is a need for theoretical development, as well as theory-guided research. For example, in the case of mobile devices, one theory (Arthur et al., 2017) is that certain types of devices increase information processing demands, which then leads to differences across device types in test scores. For gamification and games, a basic hypothesis is that gamification increases engagement, which in turn leads to improved user reactions, increased motivation, and potentially greater effort. Such mediated models can be tested in the field but are probably easier to first test in the laboratory.

Another concern is the interaction between the introduction of new technologies and the demographic characteristics of users. For example, concerning the role of mobile devices as a delivery platform, a detailed review of the literature as reported in Arthur et al. (2017) indicated that there were differences in the extent to which different demographic groups use mobile devices to complete employment-related assessments, with African Americans, Hispanics, and women displaying higher mobile device usage. A resultant question then is: "What implications does this have for the diversity of the candidates selected for employment?" A similar issue emerges regarding the reactions of various cultural and gender groups to the gamification of assessments.

It is not enough to conduct research; it must also be shared and communicated. This may be our greatest challenge, as the traditional journal publication model does not always allow for a particularly rapid dissemination of results. In order to remain relevant and to contribute to the conversation on new technologies, we will have to find ways to expedite the process of peer review and professional publication.

Other Emerging Technologies

Admittedly, mobile devices and gamification represent only two potential technologies. Other chapters in this book address additional technologies. Some technologies worth noting include:

- Big data
- Mining Facebook and other social media to extract personality and other data
- Automated scoring of essays and written material
- Applications of machine learning
- The use of avatars (discussed in more detail in the next section)

Although we have restricted our attention to the direct impact on assessment, technologies affect selection in other ways as well. Technology leads to the creation of new jobs, as well as the elimination of some occupations. Organizations are also changed through technology, although we have yet to see the widespread emergence of virtual organizations, accompanied by the elimination of all jobs, as was predicted in the 1990s.

The Future

One emerging technology that we believe will impact assessment significantly in the near future is the use of artificial intelligence (AI)-enhanced avatars. The combination of AI, natural language processing, and realistic avatars is being used in assessment applications to enhance the applicant experience, increase realism, and deliver tailored feedback. Computer-generated avatars in one form or another have been used in assessments for almost a decade (see Fetzer & Tuzinski, 2013, for a review). Typically, they have been used to enhance the look and feel of SJTs and other simulations. However, in recent years, avatars, both human and computer-generated, have become more intelligent.

Applications of such AI-enhanced avatars include guiding candidates through the hiring process from initial application, or resume submission, through testing and final interviews by

answering questions, explaining the human resources hiring process, introducing company culture, providing functional position details, scheduling tests and interviews, and keeping applicants informed regarding their status in the process. In these applications, the goal is to improve the applicant experience, increase the likelihood that top candidates remain in the selection process, and improve their perceptions of the organization.

Another area where AI-enhanced avatar technology is being deployed is in providing tailored feedback on test results to individuals, typically in developmental, as opposed to selection, situations. These applications strive to marry the richness of a professional coach with cost effectiveness and 24/7 access. Avatar-based coaches understand the individual's profile, based on the assessment results, can go over their results, answer questions, recommend a course of action, keep people on track with reminders, set up a personalized dashboard for tracking progress, and even link to other individuals, trusted others, and learning resources.

While these applications are in their infancy, it is clear that AI-based avatars will take on more significant roles in the assessment, application, and feedback processes in the not-too-distant future. Interestingly, in some healthcare applications that use a human avatar, patients are more likely to provide detailed feedback in responding to the avatar than they are to a live nurse or even a person over the phone. The same thing has been found in retail applications where customers provide product feedback. It is likely that we will see similar findings when avatars are used as test administrators, as actual components in the assessment, and also in providing developmental feedback regarding test performance.

If futurists and science fiction novelists are correct, someday soon technology will eliminate jobs since all decisions will be made by robots or machines (Autor, 2015; Frey & Osborne, 2013). Hopefully, and optimistically, I-O psychologists may be some of the last individuals working, matching the last few job applicants to the few remaining roles performed by people.

NOTE

1. It is recognized that in some instances, the implementation of a selection procedure may occur concurrently with the validation process.

REFERENCES

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

Armstrong, M. B., Collmus, A. B., & Landers, R., N. (2015). *Game-thinking in human resource management.* Poster presented at the 30th Annual Conference of the Society for Industrial and Organizational Psychology, Philadelphia, PA.

Armstrong, M. B., & Landers, R. N. (2015). *Game-thinking in Assessment: Applications of gamification and serious games.* Symposium presented at the 30th Annual Conference of the Society for Industrial and Organizational Psychology, Philadelphia, PA.

Arthur, W., Jr., Doverspike, D., Muñoz, G. J., Taylor, J. E., & Carr, A. E. (2014). The use of mobile devices in high-stakes remotely delivered assessments and testing. *International Journal of Assessment and Selection, 22*, 113–123.

Arthur, W., Jr., Edwards, B. D., & Barrett, G. V. (2002). Multiple-choice and constructed-response tests of ability: Race-based subgroup performance differences on alternative paper-and-pencil test formats. *Personnel Psychology, 55*, 985–1008.

Arthur, W., Jr., Keiser, N. L., & Doverspike, D. (2017). *An information processing-based conceptual model of the effects of the use of internet-based testing devices on scores in employment-related assessments and tests.* Manuscript submitted for publication.

Arthur, W., Jr., Naber, A. N., Muñoz, G. J., McDonald, J. N., Atoba, O. A., Cho, I., Keiser, N. L., White. C. D., Glaze, R. M., Jarrett, S. M., Schurig, I., & Bennett, W., Jr. (2015). *An investigation of skill decay and reacquisition of individual- and team-based skills in a synthetic training environment.* Toronto, ON, Canada: American Psychological Association Division 19 Suite presentation at the 123rd Annual Convention of the American Psychological Association.

Arthur, W., Jr., & Villado, A. J. (2008). The importance of distinguishing between constructs and methods when comparing predictors in personnel selection research and practice. *Journal of Applied Psychology, 93*, 435–442.

Autor, D. H. (2015). Why are there still so many jobs? The history and future of workplace automation. *The Journal of Economic Perspectives, 29*, 3–30.

Attali, Y., & Arieli-Attali, M. (2015). Gamification in assessment: Do points affect test performance? *Computers and Education, 83*, 57–63.

Bailey, P., Pritchard, G., & Kernohan, H. (2015). Gamification in market research: Increasing enjoyment, participation engagement and richness of data, but what about data validity? *International Journal of Market Research, 57*, 17–28.

Beaty, J. C., Nye, C. D., Borneman, M. J., Kantrowitz, T. M., Drasgow, F., & Grauer, E. (2011). Proctored versus unproctored internet tests: Are unproctored noncognitive tests as predictive of job performance? *International Journal of Assessment and Selection, 19*, 1–10.

Binning, J. F., & Barrett, G. V. (1989). Validity of personnel decisions: A conceptual analysis of the inferential and evidential bases. *Journal of Applied Psychology, 74*, 478–494.

Dages, K., & Jones, J. (2015). Mobile device Administration: Does length or level of assessment matter? In N. A. Morelli (Chair), *Mobile devices in talent assessment: The next chapter*. Symposium presented at the 30th Annual Conference of the Society for Industrial and Organizational Psychology, Philadelphia, PA.

Dale, S. (2014). Gamification: Making work fun, or making fun of work? *Business Information Review, 31*, 82–90.

Davies, S. A., & Wadlington, P. L. (2006). *Factor and parameter invariance of a five factor personality test across proctored/unproctored computerized administration*. Paper presented at the 21st Annual Conference of the Society for Industrial and Organizational Psychology, Dallas, TX.

Deterding, T. C., Sicart, M., Nacke, L., O'Hara, K., & Dixon, D. (2011). Gamification: Toward a definition. *Proceedings of the CHI2011 Gamification Workshop*, Vancouver, BC, Canada.

Do, B. R., Shepherd, W. J., & Drasgow, F. (2005). *Measurement equivalence across proctored and unproctored administration modes of web-based measures*. Paper presented at the 20th Annual Conference of the Society for Industrial and Organizational Psychology, San Francisco, CA.

Doverspike, D., & Arthur, W., Jr. (2012). The role of job analysis in test selection and development. In M. A. Wilson, W. Bennett, Jr., S. G. Gibson, & G. M. Alliger (Eds.), *The handbook of work analysis in organizations: Methods, systems, applications, and science of work measurement in organizations* (pp. 381–399). New York, NY: Routledge/Psychology Press.

DuVernet, A. M., & Popp, E. (2014). Gamification of workplace practices. *The Industrial Organizational Psychologist, 52*(1), 39–44.

Edwards, B. D., & Arthur, W., Jr. (2007). An examination of factors contributing to a reduction of subgroup differences on a constructed-response paper-and-pencil test of scholastic achievement. *Journal of Applied Psychology, 92*, 794–801.

Equal Employment Opportunity Commission, Civil Service Commission, Department of Labor, & Department of Justice. (1978). Adoption by four agencies of uniform guidelines on employee selection procedures. *Federal Register, 43*, 38290–38315.

Fetzer, M., & Tuzinski, K. (Eds.) (2013). *Simulations for personnel selection*. New York, NY: Springer.

Frey, C. B., & Osborne, M. A. (2013). *The future of employment: How susceptible are jobs to computerisation?* Oxford, UK: The Oxford Martin Programme on the Impacts of Future Technology. Retrieved from www.futuretech.ox.ac.uk/

Fursman, P. M., & Tuzinski, K. A. (2015). *Reactions to mobile testing from the perspective of job Applicants*. Paper presented at the 30th Annual Conference of the Society for Industrial and Organizational Psychology, Philadelphia, PA.

Gartner, Inc. (November 9 2011). *Garner predicts over 70 percent of Global 2000 organisations will have at least one gamified application by 2014*. Retrieved from http://www.gartner.com/newsroom/id/1844115 [6/17/2015]

Geimer, J., Sanderson, K., & Popp, E. (2015). *Effects of gamification on test performance and test taker reactions*. Symposium presented at the 30th Annual Conference of the Society for Industrial and Organizational Psychology, Philadelphia, PA.

Golubovich, J., & Boyce, A. S. (2013). *Hiring tests: Trends in mobile device usage*. Paper presented at the 28th Annual Conference of the Society for Industrial and Organizational Psychology, Houston, TX.

Gutierrez, S. L., & Meyer, J. M. (2013). Assessment on the go: Applicant reactions to mobile testing. In J. C. Scott & N. Morelli (Chairs), *Mobile devices in talent assessment: Where are we now?* Symposium presented at the 28th Annual Conference of the Society for Industrial and Organizational Psychology, Houston, TX.

Gutierrez, S. L., Meyer, J. M., & Fursman, P. (2015). *What exactly drives positive reactions to mobile device administration?* Paper presented at the 30th Annual Conference of the Society for Industrial and Organizational Psychology, Philadelphia, PA.

Haladyna, T. M., & Rodriguez, M. C. (2013). *Developing and validating test items*. New York, NY: Routledge.

Hamari, J., Koivisto, J., & Sarsa, H. (2014). *Does gamification work? A literature review of empirical studies on gamification*. Paper presented at the 47th Hawaii International Conference on System Science.

Huizinga, J. (2000). *Home ludens: A study of the play-element in culture*. London, UK: Routledge.

Illingworth, A. J., Morelli, N. A., Scott, J. C., & Boyd, S. L. (2015). Internet-based, unproctored assessments on mobile and non-mobile devices: Usage, measurement equivalence, and outcomes. *Journal of Business and Psychology, 30*, 325–343.

Impelman, K. (2013). *Mobile assessment: Who's doing it and how it impacts selection*. Paper presented at the 28th Annual Conference of the Society for Industrial and Organizational Psychology, Houston, TX.

Kapp, K. M. (Spring 2014). What L&D professionals need to know about gamification. *Training Industry Magazine*, pp. 17–19.

Kinney, T. B., Chang, L., Lawrence, A. D., & Moretti, D. M. (2015). *Demonstrating criterion-related validity equivalence with PC, phone, and tablet test-takers*. Paper submitted to the 31st Annual Conference of the Society for Industrial and Organizational Psychology, Anaheim, CA.

Kinney, T. B., Lawrence, A. D., & Chang, L. (2014). *Understanding the mobile candidate experience: Reactions across device and industry*. Paper presented at the 29th Annual Conference of the Society for Industrial and Organizational Psychology, Honolulu, HI.

Landers, R. N., Reddock, C. M., Cavanaugh, K. J., & Proaps, A. B. (2014). Talent assessment using mobile devices. In T. Kantrowitz & C. M. Reddock (Chairs), *Shaping the future of mobile assessment: Research and practice update*. Symposium presented at the 29th Annual Conference of the Society for Industrial and Organizational Psychology, Honolulu, HI.

Lawrence, A. D., Wasko, L., Delgado, K., Kinney, T., & Wolf, D. (2013). *Does mobile assessment Administration impact psychological measurement?* Paper presented at the 28th Annual Conference of the Society for Industrial and Organizational Psychology, Houston, TX.

McClure Johnson, T. K., & Boyce, A. S. (2015). Selection testing: An updated look at trends in mobile device usage. In N. Morelli (Chair), *Mobile devices in talent assessment: The next chapter*. Symposium presented at the 30th Annual Conference of the Society for Industrial and Organizational Psychology, Philadelphia, PA.

Morelli, N. A., Mahan, R. P., & Illingworth, A. J. (2014). Establishing the measurement equivalence of online selection assessments delivered on mobile versus nonmobile devices. *International Journal of Selection and Assessment, 22*, 124–138.

O'Connell, M. S., Arthur, W., Jr., & Doverspike, D. (2015). *Mobile assessment: The horses have left the barn . . . now what?* Pre-conference workshop presented at the 29th Annual Conference of the Society for Industrial and Organizational Psychology, Philadelphia, PA.

O'Connell, M. S., Delgado, K., & Kung, M. C. (2012). *Does proctoring impact measurement methods differently? An evaluation in a high stakes testing environment*. Paper presented at the 26th Annual Conference of the Society for Industrial and Organizational Psychology, San Diego, CA.

Oprescu, F., Jones, C., & Katsikitis, M. (2014). I PLAY AT WORK: Ten principles for transforming work processes through gamification. *Frontiers in Psychology, 5*, 1–5.

Parker, B., & Meade, A. (2015). Smartphones in selection: Exploring measurement invariance using item response theory. In N. Morelli (Chair), *Mobile devices in talent assessment: The next chapter*. Symposium presented at the 30th Annual Conference of the Society for Industrial and Organizational Psychology, Philadelphia, PA.

Pearlman, K. (2009). Unproctored internet testing: Practical, legal, and ethical concerns. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 2*, 14–19.

Petridis, P., Baines, T., Lightfoot, H., & Shi, V. G. (2014). *Gamification: Using gaming mechanics to promote a business*. Available at http://www.researchgate.net/profile/Panagiotis_Petridis/publication/263236374_Gamification__using_gaming_mechanics_to_promote_a_business/links/53e4b36c0cf2fb748710dbe4.pdf [6/17/15]

Simpson, P., & Jenkins, P. (2015). *Gamification and Human Resources: an overview*. Retrieved from https://www.brighton.ac.uk/_pdf/research/crome/gamification-and-hr-overview-january-2015.pdf [6/12/15]

Society for Industrial and Organizational Psychology. (2003). *Principles for the validation and use of personnel selection procedures* (4th ed.). Bowling Green, OH: Author.

Tippins, N. T. (2014). Technology and assessment in selection. *Annual Review of Organizational Psychology and Organizational Behavior, 2*, 5.1–5.32.

Tippins, N. T., Beaty, J., Drasgow, F., Gibson, W., Pearlman, K., Segall, D., & Shepherd, W. (2006). Unproctored internet testing in employment settings. *Personnel Psychology, 59*, 189–225.

Wasko, L. E., Lawrence, A. D., & O'Connell, M. S. (2015). *What matters in the test environment?* Paper presented at the 29th Annual Conference of the Society for Industrial and Organizational Psychology, Philadelphia, PA.

Weekley, J. A., Hawkes, B., Guenole, N., & Ployhart, R. E. (2015). Low-fidelity simulations. *Annual Review of Organizational Psychology and Organizational Behavior, 2*, 18.1–18.28.

Weiner, J. A., & Morrison, J. D. (2009). Unproctored online testing: Environmental conditions and validity. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 2*, 27–30.

Wernimont, P. F., & Campbell, J. P. (1968). Signs, samples, and criteria. *Journal of Applied Psychology, 52*, 372–376.

Wood, E., Stephens, K., & Sliter, K. (2015). *Applies to oranges? Use and comparative scores for mobile and nonmobile selection assessments*. Paper presented at the 2015 Annual Conference of the International Personnel Assessment Council, Atlanta, GA.