APPLICANT REACTIONS TO GAME-BASED ASSESSMENTS: THE EFFECT OF FLOW, FAIRNESS, AND FIT

By

Sarena Bhatia

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Psychology – Doctor of Philosophy

2018

ABSTRACT

APPLICANT REACTIONS TO GAME-BASED ASSESSMENTS:
THE EFFECT OF FLOW, FAIRNESS, AND FIT

By

Sarena Bhatia

This study examines the effect of game-based assessments (GBAs), assessments that incorporate game elements to evoke and measure the relevant ability or skill, on applicant reactions. The author introduces and finds support for the idea that a flow state explains applicants' engagement and immersion with GBAs, and as well as functions as a mediator to other positive reactions. This study utilizes a 2 (GBA, traditional selection assessment) by 2 (made an offer, not made an offer) experimental time-lagged design and a diverse sample to examine the effect of assessment method on a number of reactions measures. Participants were asked to imagine they were applying to a sales role at a fictional global corporation, and randomly assigned to an assessment condition. The next day, they were emailed either an acceptance or rejection letter. Those in the GBA condition were significantly more likely to experience higher rates of flow than those in the traditional assessment condition. Technology self-efficacy moderated the relationship between assessment method condition and flow: those in the GBA condition and high in efficacy experienced significantly higher rates of flow. Taking the GBA was positively related to perceived job-relatedness (when mediated by flow), justice perceptions, perceived person-organization fit, organizational attractiveness and finally, positive intentions to accept the offer. Selection decision moderated the relationship between perceived job-relatedness and justice perceptions: those who were made an offer and perceived the assessment to be job relevant had the highest perceptions of justice. The implications for GBA research, game-based hiring, and the role of individual differences in understanding reactions to game-based tools are discussed.

ACKNOWLEDGEMENTS

I want to start by thanking my advisor, Dr. Ann Marie Ryan, for her guidance and unwavering support during my graduate school journey. Ann Marie's open door, willingness to talk through any issue, and remarkably quick editing time made the process so much easier and more productive. I want to thank both Ann Marie and Dr. Kevin Ford for showing me what a true scientist-practitioner can be, and for making Michigan State a place that encourages and prepares students for either route they choose to take. I could not have asked for a better place to be trained.

Thank you as well to my other two committee members, Drs. Chris Nye and Robby Ratan, for their constructive and insightful feedback. Chris' deep statistical knowledge base and patience allowed me to learn a lot, and Robby's fresh perspective and reminder to have a little fun while studying games made the process a lot more enjoyable.

Thanks to my team at Aon, who was not only enormously supportive while I was finishing my dissertation but who challenges, teaches, and encourages me every day on the job.

A huge thank you (and sigh of relief!) to my graduate school friends, without whom I would not have made it to the finish line. Collectively, you are some of the brightest, most compassionate, and most entertaining people I have ever met and I feel very lucky to have been by your side through the cold days, long nights, and wonderful times spent together.

Lastly and most importantly, thank you to my family. To my parents, who are my foundation and without whom I would not and could not have pursued this degree. To my brother and grandparents, who always offered their love. And to Adnan, who outshines the PhD as my favorite find at Michigan State.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

The area of personnel assessments has experienced many changes in its one hundred plus year history, many of which have coincided with global business trends. For example, experts overwhelmingly agree that technology is one of the biggest influences on our work and so unsurprisingly, organizations are using different technology systems to recruit, assess, and track applicants (Ryan & Ployhart, 2014; Silverthorne, 2011; Tippins, 2015). At the same time, engagement has become a big priority for companies (Schwartz & Porath, 2014). Reflecting this, there has been a surge of research and interest looking at how applicants respond to and engage with organizations during the hiring process (Kantrowitz, 2014; Ryan & Ployhart, 2000). Lastly, gamification, or the inclusion of game concepts to make traditional processes more fun, has spread quickly in the working world (Altman, 2015; Chou, 2015). Examples of gamification are ubiquitous such as using narratives and increasingly harder levels in training programs, or leaderboards and points systems to make data entry less tedious (Shergill, n.d.). Somewhat predictably, game-based thinking is slowly percolating into the assessment space with several vendors offering tools for purchase (i.e. Artic Shores, Knack, Persona Labs, Pymetrics, Revelian). Despite this movement, and the intersection of game-based assessments with the big business trends of technology and engagement, little research attention has been dedicated specifically to it.

One reason game-based hiring tools lack research attention is that a number of areas of expertise are needed to create these tools. Organizational psychologists are involved with identifying the constructs to be measured, and providing background on how other similar selection tools have been used in the past. Then they, or psychometricians, provide insight to how the psychometrics of game-based tools compare with previously validated tools. Game

design experts provide input with how to best structure the assessment, and which game characteristics are suited for selection. Then information technology (IT) specialists recommend how the assessment fits into the existing IT structure. Game-based assessments are not solely in the domain of psychologists and partially as a result, they are not a dominant area of research in organizational psychology literature.

Despite this lack of research, there are many ways gamed assessments can contribute to researchers' objectives. There have been calls to isolate and study predictor methods - or the specific process by which domain-related information is elicited, assessed, and used - in order to gain more meaningful findings (Arthur & Villado, 2008; Ryan & Huth, 2008). Game-based hiring tools offer a novel way of tapping multiple constructs simultaneously, a growing trend in the assessment space (Ryan & Ployhart, 2014). They can act as *stealth assessments* by evaluating skills and abilities while engaging the applicant, allowing for measurement without the confounding effects of cognitive load or faking (Fetzer, 2013; 2015; Shute, 2011). A systematic study of gamified tools can answer questions about the psychometrics of these kinds of assessments and the circumstances under which they are appropriate to be used (McCarthy et al., in press). There is ample opportunity in this space to contribute to very timely conversations.

For this study, I focus on game-based assessments (GBAs), defined as a tool or method that incorporates game elements to evoke and measure one's ability or skill. These are tools that incorporate game elements - such as feedback, narratives, or graphics - so that the elements are a fundamental part of the way the assessment elicits the knowledge, skills, and abilities it is designed to measure. An example is an assessment that evaluates persistence by presenting the applicant with an impossible challenges and the option to exit the module, and seeing how long the applicant tries to accomplish the goal of the challenge before making the decision to exit.

I focus on GBAs because there is fresh research insight to be gained and there is a practical need to study them. GBAs are gaining popularity in large part because they are perceived as being attractive to the applicant. Vendors who sell GBAs market them as enjoyable, immersive, and preferable (Revelian, n.d.). The assumption is that as long as the assessments are as reliable and valid as traditional measures, applicants will opt to play games. While there is some data to support this (Anderson & Rainie, 2012; Shergill,, n.d.), there is other work that says the mediums through which GBAs would normally be administered, such as a smartphone, leave applicants feeling they did not have an adequate chance to perform (Kantrowitz, 2014). Adding fuel to this fire, some work shows that common procedures are viewed as more fair (Ployhart, Ehrhart, & Hayes, 2005; Steiner & Gilliland, 1996) meaning the novelty of GBAs could put them at a disadvantage. There is also a broad base of work showing assessments that are not contextualized to work (like some game-based tools) are viewed less favorably (Macan, Avedon, Paese, & Smith, 1994; Pulakos, 2005; Shaffer & Postlethwaite, 2012; Truxillo, Steiner, & Gilliland, 2004). Clearly there are competing hypotheses about how game-based tools will be viewed, yet there is no work that is grounded in existing theory. This is despite researchers stating that "method comparison studies need to be well mapped back onto theories of applicant reactions" in order to develop "taxonomies of reactions that selection system designers can use" (Ryan & Huth, 2008, p. 121).

Because game-based tools are new to the hiring context, there is also very little theoretical work on how assessments that are designed to be fun or engaging affect later outcomes. Organizational psychologists have definitively shown that applicant reactions during the selection process affect later dealings with the organization, such as recommendation intentions or willingness to buy the company's products (Hausknecht, Day, & Thomas, 2004).

While there is theory that links motivation, anxiety, and fairness during the hiring process to these later reactions (Ryan & Ployhart, 2000), this theorizing is limited and a product of the type of assessments that have been studied so far. GBAs are some of the first tools designed to make the assessment process enjoyable so there needs to be a fresh conceptualization of how intrinsic interest and its correlates fit into the applicant's experience.

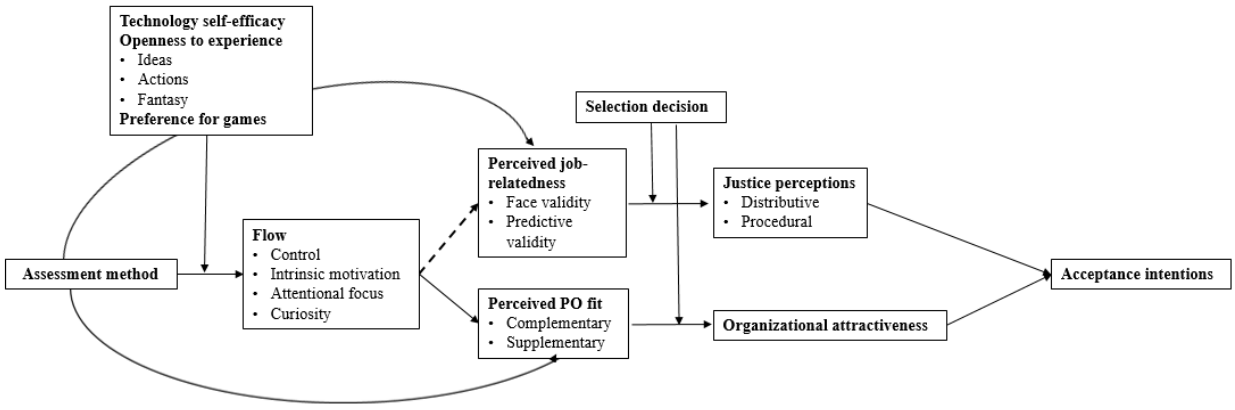
Lastly, there is little work on individual differences that are pertinent to gaming and technology despite calls to examine these further in the applicant reactions arena (Anderson, 2003). While selection and gaming researchers have studied the role of personality and self-efficacy more generally (Hartmann & Klimmt, 2006; Maertz, Bauer, Mosley, Posthuma, & Campion, 2005; Ostberg, Truxillo, & Bauer, 2001; Teng, 2008; Truxillo, Bauer, Campion, & Paronto, 2006; Zibarra & Patterson, 2015), very little of this work has been applied to a game-based hiring situation. Personal characteristics such as technology self-efficacy and openness to experience play a pivotal role in applicants' basic willingness and ability to engage with gaming tools, so it is important to know if there are specific segments of the workforce that are negatively affected by their rise in popularity.

This study contributes to the literature in three ways. The first contribution is understanding whether GBAs as an assessment method have a positive effect on applicant reactions. I take two well-grounded theories, justice and person-organization fit, to understand how GBAs affect an applicant's intentions to accept an offer at a company. The second contribution is to the applicant reactions literature more generally; I posit the concept of flow to understand how high levels of engagement and absorption during the assessment process affect later reactions. To do this, I manipulate whether participants see a GBA, or traditional computerized measures (i.e. cognitive ability and Likert items) that measure the same constructs

as the GBA. This is much needed as assessments become increasingly high-fidelity and simulation-like; if we have a deeper understanding of how flow affects applicants, we can extrapolate to similarly captivating assessments. Lastly, this study evaluates the role of three relevant individual differences - technology self-efficacy, openness to experience, and preference for games - in reactions to GBAs (please see Figure 1 for the proposed model).

In the following sections, I first review game elements and the game-based assessment into which game elements can be incorporated. In doing this I situate game-based assessments among similar tools so that the reader understands how they relate and differ. Next I review Csikszentmihalyi's original definition of a flow state (Csikszentmihalyi, 1990), and explicate how the concept of flow has been altered to better understand the interactions between humans and computers. I discuss the moderating effects of key individual differences. I then review the literature on applicant reactions, specifically from a justice perceptive, and incorporate literature on perceptions of fit with the organization. Throughout these sections, I explain the role of selection decisions in affecting reactions and finally, how an applicant's reactions to the assessment affect his or her intentions of joining an organization.

Figure 1. *Model of proposed relationships for study*



Games in assessment

Game researchers have distinguished *gaming* from *playing*- the latter captures a more "free-form, expressive, improvisional" set of behaviors while gaming stands as a structured and goal-directed activity, dominated by competition and rules (Deterding, Dixon, Khaled, & Nacke, 2011, p. 3). While gaming is more structured, it has been hard to neatly define- Crawford described games as a representation of reality that requires: interaction between a user and a system, conflict, and safety through simulation (Crawford, 1984). To Huizinga, games are "non-serious but intensely engaging voluntary activities structured by rules and secret social boundaries" (Huizinga, 2000; Seaborn & Fels, 2015, p. 16). Prensky talked about them as organized play that provides enjoyment (2001). He also discussed how computer games can be identified through six key structural elements that engage the player. These definitions show that games are described by how they make the player feel, what characteristics they include, or both.

Despite the fact that there is no list of characteristics that are common to all games (Wittgenstein, 1953), in order to understand games one must understand game elements. In the following section, I review the meaning of game elements as well as taxonomies that have been created in the educational and technological arenas. The purpose of this section is to describe the building blocks of game-based assessments. I then spend time discussing the different kinds of gamified tools that are used in business and education to give greater insight to how games are being conceptualized and used in similar spaces.

Game elements

Before embarking on a journey through the definitions of the game-based space, it is important to clarify what is meant by game elements. Game elements are defined by Deterding and colleagues as "a set of building blocks or features shared by games (rather than a necessary

set of conditions for a game)" and said to be "*characteristic* to games… readily associated with games, and found to play a significant role in gameplay" (2011, p. 4). An example of a game element is a leaderboard in which users are tracked and ranked for their accomplishments in a game.

Game elements are applicable to, and therefore have been discussed by, diverse literatures from computer science to learning and development. Game element taxonomies vary widely as well, a result of the different disciplines from and purposes for which they were created. Before presenting individual elements, I will review two "organizers" that have been posited to sift through the array of game components (please see Table 1 for summaries of these organizers in blue). These organizers are high-level, and can each contain many specific game elements. After presenting the organizers, I will review taxonomies which provide more narrow and specific representations of game elements (also reviewed in Table 1).

Table 1. *Game element taxonomies*

Author and Year	**Discipline**	**Elements**
Deterding et al. (2011)	Media research	Game interface design patterns, Game design patterns and mechanics, Game design principles and heuristics, Game models, Game design methods
Dickey (2005)	Educational technology research and development	Focused goals, Challenging tasks, Clear and compelling standards, Protection from adverse consequences for initial failures, Affirmation of performance, Affiliation with others, Novelty and variety, Choice
Floryan (2009)	Computer science	Goals, Content and user tasks, Simulation fidelity, User freedom
Garris, Ahlers, & Driskell (2002)	Gaming	Fantasy, Rules/goals, Sensory stimuli, Challenge, Mystery, Control

Table 1 *(cont'd)*

Author and Year	Discipline	Elements
Gee (2009)	Game-based learning	Underlying rule system and game goal to which player is attached, Micro-control that creates sense of intimacy and power, Experiences that offer good learning opportunities, Match between affordance and effectivity, Modeling to make learning more general and abstract, Encouragement to players to enact their own unique trajectory in game
King, Delfabbro, & Griffiths (2010)	Clinical psychology	Social features, Manipulation and control features, Narrative and identity features, Reward and punishment features, Presentation features
Prensky (2001)	Game-based learning	Rules, Goals and objectives, Outcomes, Feedback, Conflict (competition, challenge, opposition), Interaction, Representation (story)
Reeves & Read (2009)	Business technology	Self-representation with avatars; 3D environments; Narrative context; Feedback; Reputation, ranks, and levels; Marketplaces and economies; Competition under rules that are explicit and enforced; Parallel communication systems that can be easily configured; Time pressure
Shute & Ke (2012)	Game-based learning	Interactive problem solving, Specific goals/rules, Adaptive challenge, Control, Ongoing feedback, Uncertainty, Sensory stimuli
Sweetser & Wyeth (2005)	Computer science	Concentration, Challenge, Player skills, Control, Clear goals, Feedback, Immersion, Social interaction
Wilson et al. (2009)	Gaming	Adaptation, Assessment, Challenge, Conflict, Control, Fantasy, Interaction (equipment), Interaction (interpersonal), Interaction (social), Language/communication, Location, Mystery, Pieces or players, Progress and surprise, Representation, Rules/goals, Safety, Sensory stimuli
Wood, Griffiths, Chappell, & Davies (2004)	Cyber psychology	Sound, Graphics, Background and setting, Duration of game, Rate of play, Advancement rate, Use of humor, Control options, Game dynamics, Winning and losing features, Character development, Brand assurance, Multiplayer features

Note: Organizers are in blue.

The first organizer is by Deterding and colleagues, created for the 2011 Mindtrek conference which brings companies, associations, and students together to discuss the way technology is and will be affecting business (Deterding et al., 2011). The organizer by Deterding and colleagues uses five levels to sort game elements. The levels range from concrete to abstract. The first and the most concrete is *game interface design patterns*. This refers to what the player will see on the screen, such as a leaderboard or a badge, and how he or she will interact with these elements. The second level is *game design patterns and mechanics* and are reoccurring elements that the player interacts with such as time constraints or limited resources. These are separated from the first level in that they are not visual, but rather arrangements with which the player meets several times throughout the game. The third level is *game design principles and heuristics* which serve as guidelines within which players must work to advance or solve problems. An example of this would be a set of rules a player has to obey to get from Point A to Point B. Fourth are *game models* which are used to conceptually understand a player's experience while gaming. Lastly, there is the most abstract level, *game design methods*, or strategies used to design games like playcentric design where game designers keep the player's experience at the forefront during the development process (Morford, Witts, Killingsworth, & Alavosius, 2014). Many of the individual element taxonomies we review later focus on elements in the first three levels.

A second organizer comes from clinical psychology, and was created to understand what elements in video games can encourage excessive and even addictive playing behaviors (King, Delfabbro, & Griffiths, 2010). The authors expanded on a 2004 taxonomy by Wood and colleagues by reorganizing and supplementing using recent research (Wood, Griffiths, Chappell, & Davies, 2004). Their new organizer contains five categories. The first is *social features*, or

features that allow for affiliation and communication. An example of this is chat functionality within a video game. The second category contains *manipulation and control features*, and these allow the player to interact with and control the game such as a save button or a display that tells the player how many resources or pieces of inventory he or she has available. Interestingly, the authors include elements external to the game such as keyboards or hand controllers under this category. They make the argument that these tactile materials affect how the player inputs and controls the game, and therefore must be taken into account as part of the experience. The third category has *narrative and identity features* and speaks both to the storyline of the game, and the various identities a player can take on through avatars and characters. Fourth are *reward and punishment features*. These refer to the reinforcements players get in the game, both negative like decreased health statistics and positive like purchasing weapons with coins that have accumulated over the level. Lastly, *presentation features* refer to aesthetic features like graphics and sound effects. In the article, King and colleagues were interested in how these features can contribute to excessive use of video games. The organizer can also inform us of the elements that often go into games.

These two organizers provide a high-level framework for other taxonomies of game elements. By this, I mean that the organizers are abstracted, and each category encompasses a number of specific elements. However, just as the two organizers come from divergent areas so too do game element taxonomies. This means that game elements do not always cleanly fit into one organizer category. For example, one taxonomy from the educational technology domain includes the game element of *novelty and variety* (Dickey, 2005). *Novelty and variety* could be tentatively placed in the third or fourth level of Deterding and colleagues' organizer, but does not naturally fit into King and colleagues'. This further illustrates the point that game elements are

used and considered in a number of fields, and there is no widely accepted framework in which to study them.

I now review specific game element taxonomies, with an eye toward the diverse disciplines from which they derive. The Dickey taxonomy, referenced above, contains seven other elements in addition to novelty and variety: focused goals; challenging tasks; clear and compelling standards; protection from adverse consequences for initial failures; affirmation of performance; affiliation with others; and choice (2005). Most of these elements have sub-elements that add clarity to the taxonomy.

Another taxonomy, derived in the game-based learning space, looks at interactive problem solving, specific goals/rules, adaptive challenge, control, ongoing feedback, uncertainty, and sensory stimuli (Shute & Ke, 2012). This taxonomy focuses on seven elements that are conducive to a learning environment; rather than focusing on player chat functions, it examines interactive problem solving. Instead of mentioning narratives or identity, it simply requires adaptive challenges and some measure of control. Shute and Ke's taxonomy sits at a higher level of abstraction than some other game element taxonomies, but it serves the intended purpose for its audience: to guide developers in creating games in which students grow their skills and abilities in an engaging arena.

Reeves and Read provide a third taxonomy I review here to better illustrate *game interface design patterns* discussed by Deterding and colleagues (Reeves & Read, 2009). This taxonomy includes self-representation with avatars; 3D environments; narrative context; feedback; reputation, rank, and levels; marketplaces and economies; competition under rules that are explicit and enforced; parallel communication systems that can be easily configured; and

time pressure. Many of these elements could be layered on top of existing assessments or trainings to make them gamified, and are also essential for game-based tools.

This section provides an overview of game elements, with the purpose of describing the types of game elements that can be used, and illustrating examples of elements that can be incorporated into games. The taxonomies referenced here are by no means an exhaustive list. There are older ones that have been expanded upon (Garris, Ahlers, & Driskell, 2002; Gee, 2009; Prensky, 2001; Sweetser & Wyeth, 2005; Wood, Griffiths, Chappell, & Davies, 2004), and others that overlap with the ones discussed here (Bedwell, Pavlas, Heyne, Lazzara, & Salas, 2012; Floryan, 2009).

I now review concepts related to game-based assessment, to separate how GBAs differ from similar tools.

Simulations

Although not directly related to game use, high-fidelity simulations are highly related to gamification. Simulations have been defined in their simplest terms as a "form of integrated measurement… in which the candidate is expected to perform some task that is closely related to the work of the job" (Tippins, 2015, p. 557). While this is a hiring-specific definition, simulations are often used in other domains and have also been described as "a serious attempt to accurately represent a real phenomenon" (Crawford, 1984, p. 8). Simulations can be both low and high fidelity, meaning low and high in realism. The type that is most relevant here is high-fidelity assessments, or those that provide a realistic environment and materials to show a task as it would be presented on the job (Motowidlo & Dunnette, 1990). An example would be a virtual assessment center where candidates are shown around a virtual office, and asked to write emails to stakeholders or have a phone call with a supervisor.

Both high-fidelity simulations and gamified tools are rich in media, measure multiple constructs, and assess the candidate with lower transparency than other tests may (such as an aptitude or physical ability test). However, they are also substantively different. One factor separating simulations from their game counterparts is that the games have the added element of fun. Simulations are meant for measurement while serious games are meant for both measurement and some amount of enjoyment (Mislevy et al., 2014). Additionally, simulations are often highly job-related and provide tasks that are similar to the job. Gamified tools may do this, or they may assess job-relevant constructs like personality and cognitive ability in ways that are distinct from job-related tasks.

Gamification

One buzzword that is used frequently in both business and education is *gamification*. This term was coined in 2008 by the digital media industry, and has more recently being defined by Deterding and colleagues as "the use of game elements in non-game contexts" (2011, p. 2). A non-game context in this case is one in which a game would not normally be expected, or a context that is not confined just to entertainment. To put it in other words, gamification is when game elements are layered onto an existing program or context with the intention of increasing the enjoyment and experience of the user (DuVernet & Popp, 2014; Kapp, 2012). The existing program or context does not become a game at this point; it is only "augmented with characteristics borrowed from games" (Landers & Landers, 2015). An example of this would be adding a point system and leaderboard to a traditional training program so that users can easily track their progress and compare it to that of their peers. The training itself remains unchanged, but game elements are layered onto the tool to add fun and competition. As I will argue later,

GBAs are not just gamified tools; rather the game elements are more deeply incorporated in their structure and purpose.

Serious games

The term serious game is used in the training and learning literature (also called learning games or training games); serious games take gamification one step further in that they are "full-fledged games for non-entertainment purposes" (Deterding et al., 2011, p. 3; Landers, 2014). In these games education, rather than entertainment, is the purpose (Michael & Chen, 2005). Instead of simply layering game elements onto an existing program or tool, serious games have these elements woven into the foundation of the tool or program. As we will see shortly, GBAs incorporate game elements in a similar way but are not used solely for educational purposes the way serious games are.

Game-based assessment

All of the concepts discussed above have several things in common. They include game elements that encourage fun and engagement. They assess the user over the course of the interaction whether it be to provide feedback to the user or to a third party (such as an instructor). However, many of these assessment tools have only been discussed and studied in the training or academic literatures (Chin, Dukes, & Gamson, 2009; Horn & Cleaves, 1980; Landers, 2014; Mislevy et al., 2014; Wilson et al., 2009). Although not formally researched, assessments that incorporate games are also experiencing a burgeoning use in hiring. These tools evaluate candidates on a variety of well-known predictors: cognitive ability, personality, and job knowledge. Therefore, I put forth a new definition of game-based assessment that can be used within the selection arena.

A game-based assessment, for the purposes of this study, is a tool or method that incorporates multiple game elements to evoke and/or measure one's ability or skill. This is different than gamification and similar to serious gaming in that the game elements are more fully integrated into the measurement of the relevant construct. GBAs also require that more than one game element be present, otherwise the tool will be gamified rather than game-based. The constructs assessed in these tools are used by organizational members to assess ability and fit for a job. This introduces a different dynamic than the training and education literature as there is an immediate decision that will be made based (at least partially) on the outcome, and there is highly-motivating goal in place for the candidate. This requires finely tuned assessments and a careful eye toward participant reactions, as these have large implications for the organization.

This is not to say that GBAs cannot be used for educational purposes- in fact, there are assessments such as SimCityEDU: Pollution Challenge! (Mislevy et al., 2014) that incorporate multiple game elements to evoke and measure the relevant constructs. However, the GBAs I focus on are designed to measure relevant work characteristics and generally meet guidelines for use in selection.

An example of a GBA in a selection context may be one in which a candidate is being measured on his or her ability to innovate. He or she could have a game scenario in which there is no correct answer. The assessment may then incorporate game elements like autonomy, multiple pathways to success, competition, and feedback in the evoking and measuring of innovation. These tools can be used to mirror the constraints of an environment in which innovation is required, allow for the demonstration of the ability, and aid in measurement of a construct that is difficult to capture.

I suggested earlier that because GBAs are so new to the selection context, there is no existing concept that truly captures the applicant's experience. I now move on to discuss the impact GBAs can have on an applicant, most notably in inducing a flow state. Below I discuss the original conceptualization of flow, flow in the human-computer interaction literature, and the conceptualization of flow I take for this study.

The role of flow

As discussed earlier, GBAs are unique in that they are one of the first assessment tools designed primarily to absorb the candidate. We have seen a progression from cognitive ability tests and paper/pencil measures to more complex and higher fidelity tools such as situational judgment tests and assessment centers. However, it has not been the goal of any of these types of assessments to engage or enthrall the candidate. Those are useful supplemental benefits (Pulakos, 2005), but not the primary purpose. GBAs are notable in that a large part of the appeal in using them is the candidate experience. Consequently, there is a lapse in theory to explain what this engagement process looks like when an applicant interacts with the assessment.

In order to fill this gap, I put forth the concept of *flow*. Flow has been described by its founder as "a narrowing of the focus of awareness, so that irrelevant perceptions and thoughts are filtered out; by a loss of self-consciousness; by a responsiveness to clear goals and unambiguous feedback; and by a sense of control over the environment" (Csikszentmihalyi, 1975a, p. 72). Flow describes a state of deep immersion, where an individual is subsumed in his or her task. Flow was originally conceptualized as having eight dimensions: challenge-skill balance, clear goals and feedback, sense of control, the merging of action and awareness, concentration on the task at hand, loss of self-consciousness, and the transformation of time (Csikszentmihalyi, 1990).

I rely on flow as a central construct for a few reasons. Flow is preferable to a less immersive state such as enjoyment because games and game-based tools offer a more intense and multi-dimensional experience than can be captured through enjoyment alone. Flow can be ascribed to a specific tool, unlike some constructs like engagement which are "not focused on any particular object, event, individual, or behavior" (Schaufeli, Salanova, González-Romá, & Bakker, 2002, p. 74). Flow has validated measures, and is an original and novel construct unlike others that are more of a repackaged version of flow (i.e. cognitive absorption). Lastly, there is ample support for flow within game-based systems as I will explore in later sections. I now transition into a more comprehensive explanation of each dimension of the original flow concept.

Csikszentmihalyi's original conceptualization

The first dimension of flow is a challenge-skill balance. This occurs when the situation calls for "a bundle of opportunities for action" and the individual has the skills to match those opportunities (Csikszentmihalyi, 1990, p. 49). When an individual faces a challenge, he or she will be stimulated. Having the ability to match the challenge keeps the individuals in the realm between boredom (from not having enough challenge) and anxiety (being overwhelmed by the challenge), which is where flow occurs.

The second element requires clear goals and feedback. Goals and feedback enable high levels of immersion in the activity because they keep the individual focused and on-task. Feedback provides the individual the chance to track progress toward the goal, and the goal almost always provides a sense of success when challenge and skill are matched.

Third, the individual will feel a sense of control over the situation when he or she is in flow. At this time, there is a sense of peace and power that the individual is able to effect change in any way he or she sees fit. Of course, this sense is just a feeling, and the individual can very

well lose control (think of sports or games where one can lose). But at the time, one experiences a deep sense of control and with it the ease and pleasure that comes from mastery over a domain.

Fourth, the merging of action and awareness refers to all attentional resources being directed to the task at hand. The individual loses a sense of him or herself as separate from the activity. Even though there is a large amount of effort and focus directed at the activity, the individual has no sense of the strain. This is tied closely with the fifth dimension, concentration, and the sixth dimension, loss of self-consciousness. In the case of concentration, the individual forgets about every day worries. He or she has a focused band of awareness and it is restricted to the activity. Similarly in the case of loss of self-consciousness the individual is not only one with the activity but one with the world. He or she fails to recognize the details of the environment and simply feels united with it. There is no energy to scrutinize one's appearance or to have self-doubts; all is consumed during flow.

Lastly is the transformation of time. This is one of the most intuitive appeals of the concept, as almost everyone has had the experience of looking up at the clock and realizing hours have gone by in what feels like a much shorter span of time. The transformation of time can occur reversely as well, where a movement that takes a few seconds feels like it is stretched to a minute or more, such as with a complex move in dance or gymnastics. The crux of this dimension is that to the individual in flow, time does not pass at its usual rhythm. One is freed from that steady march for however long they are able to maintain the state.

These eight concepts were originally hypothesized as elements of flow by Csikszentmihalyi (1990). However, there have been many contradictions in the literature since this early postulation. The first is that the elements that constitute flow are not agreed upon. For example, in his book Csikszentmihalyi discusses the autoletic personality, which is an

individual's tendency to be able to experience a state of flow, as a stable tendency rather than an element of flow. In a later publication he and Jeanne Nakamura openly state that the challenge-skill balance, and clear goals and feedback are precursors to flow rather than dimensions of the actual state (2002). Despite this one of the most-used measures of flow includes the autoletic personality, challenge-skills balance, clear goals, and feedback as dimensions of flow (Jackson & Eklund, 2002; Jackson & Marsh, 1996). A second related issue is that the number of elements in flow is not agreed upon. Some work treats flow as a one-factor construct (Hoffman & Novak, 1996), and some takes a composite of certain elements and label them as flow (Bakker, 2008; Ghani, 1995; Ullén et al., 2012). Lastly, as a result of the differing dimensions there is no agreed upon definition of flow. Many authors use the challenge-skills balance to encapsulate flow (Chiang, Lin, Cheng, & Liu, 2011; Csikszentmihalyi & LeFevre, 1989) and others simply list all the elements in their definition of flow and allow that to suffice (Hoffman & Novak, 1996; Novak, Hoffman, & Duhachek, 2003).

Flow as a concept lacks consistency in its definition and elements (Choi, Kim, & Kim, 2007). Due in part to the muddy origins of the construct and its elements, flow researchers treat the elements in whatever way fits their line of work. I do the same by taking a human-computer interaction (HCI) approach to flow, described in more detail below.

Flow in human-computer interaction

Flow can be used to understand performance in fields as varied as sports (Jackson, 1995; Jackson & Csikszentmihalyi, 1999) to internet shopping (Guo & Poole, 2009; Smith & Sivakumar, 2004). There has been a large contingent of researchers focused on flow and technology systems. For example, some work has examined how those who experience flow when working with computers are more likely to continue to use those systems (Agarwal &

Karahanna, 2000; Ghani & Deshpande, 1994; Webster, Trevino, & Ryan, 1993). Others have looked at flow in relation to online or video games (Chiang et al., 2011; Chou & Ting, 2003; Cowley, Charles, Black, & Hickey, 2008; Fu, Su, & Yu, 2008; Hsu & Lu, 2004; Sweetser & Wyeth, 2005).

It is evident in these works that flow has been adapted to fit the relevant context. Researchers in the human-computer interaction arena conceptualize flow in a way that is pertinent to users' interactions. Across a review of the flow-HCI literature, a few of Csikszentmihalyi's elements are reoccurring (please see Table 2 for a summary of how elements of flow are conceptualized in different HCI studies). The first two prominent ones are concentration and loss of self-consciousness. In the HCI literature, these are discussed via the mesmerizing quality of computers (Webster et al., 1993). Because computers can offer rich, vivid media environments, they are better able to engage the full attentional resources of the user (Hoffman & Novak, 1996). Users can also perform many tasks on a computer that are engaging in their own right (e.g., shopping, gaming) which only enhances the absorptive quality.

The third common element is control; control is easily induced with technology, given the user possesses the right skills and self-efficacy. Computers in particular offer flexibility (one example is undoing mistakes such as a mistyped word), as well as many alternative activities (web surfing, document creation, etc.) (Ghani, 1995; Webster et al., 1993). Games in a technology-mediated setting are especially good at giving the user a sense of control (Chen, 2007; Cowley et al., 2008).

The last flow element that is featured in many HCI works is enjoyment or intrinsic motivation (Ghani & Despande, 1994; Hoffman & Novak, 1996; Trevino & Webster, 1992). When users are interacting with a technology system, they often have goals and feedback. They

may be looking for information, completing a task for work, or engaging in a recreational activity like shopping or gaming. These goals and feedback create engagement, which when coupled with a sense of autonomy and concentration can be very fulfilling (Ryan & Deci, 2000). A user will often find him or herself enjoying the activity and motivated to continue for one's own sake rather than a constraint placed by an external force (Csikszentmihalyi, 1975a). While this is not listed as an element in Csikszentmihalyi's original works, he makes extensive reference to it, going so far as to include the term "intrinsic reward" into the title of seminal flow works (Csikszentmihalyi, 1975b; Csikszentmihalyi & Larson, 1978; Csikszentmihalyi, 1979). The state of flow is inherently motivating because of its pleasurable aspects, so it is no surprise that intrinsic motivation is heavily present in the HCI literature as an element of flow (Ghani, 1995; Hoffman & Novak, 1996; Hsu & Lu, 2004; Trevino & Webster, 1992).

Table 2. *Conceptualizations of the elements of flow*

Author and Year	**Journal**	**Dimensions**	**Summary**
Chen (2007)	Communications of the ACM	Csikszentmihalyi's 8	Link flow with games
Ghani & Deshpande (1994)	Journal of Psychology	Antecedents: Challenge, Sense of control During: Total concentration, Enjoyment	Examine job characteristics and flow theory to understand the experience of people who use computers at work
Ghani (1995)	Carey (1995)	Challenge-skill balance, Control, Intrinsic motivation	Examine how a mix of flow elements and other aspects like creativity and learning relate to flow
Ghani, Supnick, & Rooney (1991)	International Conference on Information Systems	Total concentration, Enjoyment	Examine the role of control and challenge in inducing flow in FTF vs. computer-mediated work groups

Table 2 *(cont'd)*

Author and Year	Journal	Dimensions	Summary
Hamari et al. (2016)	Computers in Human Behavior	Challenge-skill balance	Investigate the impact of flow (operationalized as heightened challenge and skill), engagement, and immersion on learning in game-based learning environments
Hoffman & Novak (1996)	American Marketing Association	Seamless sequence of responses, Intrinsically enjoyable, Accompanied by loss of self-consciousness, Self-reinforcing	Understand consumer navigation in computer-mediated environments using the concept of flow
Hoffman & Novak (2009)	Journal of Interactive Marketing	Congruence of skill and challenge, Interactivity, Vividness, Motivation, Involvement, Attention, Telepresence, Flow, Increased learning, Perceived behavioral control, Exploratory mindset, Positive subjective experience	Examine the conceptualization and measurement of flow in online environments, and marketing outcomes
Hoffman, Novak, & Yung (1999)	Marketing Science	Seamless sequence of responses, Intrinsically enjoyable, Accompanied by loss of self-consciousness, Self-reinforcing	Create model of what makes a compelling online experience
Hsu & Lu (2004)	Information & Management	Total involvement, Enjoyment, Control, Concentration, Intrinsic interest	Examine how and whether flow and the technology acceptance model predict online game use
Koufaris (2002)	Information Systems Research	Intrinsic enjoyment, Perceived control, Concentration/attention focus	Examine the role of the technology acceptance model, flow, and consumer behavior on online consumer behaviors
Novak, Hoffman, & Duhachek (2003)	Journal of Consumer Psychology	Seamless sequence of responses, Intrinsically enjoyable, Accompanied by loss of self-consciousness, Self-reinforcing	Examine flow, characteristics of flow, and how flow differs depending on state in regards to web buying experiences

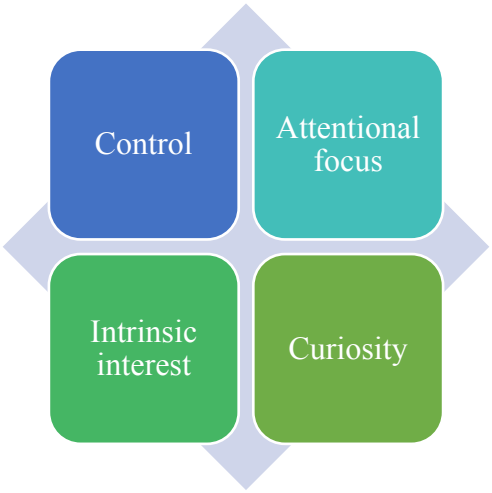
Table 2 *(cont'd)*

Author and Year	Journal	Dimensions	Summary
Pilke (2004)	International Journal of Human-Computer Studies	Csikszentmihalyi's 8/9	Examine which computer tasks induce flow
Trevino & Webster (1992)	Communications Research	Control, Attentional focus, Curiosity, Intrinsic interest	Examine how technology and characteristics of technology influence flow and how this in turn affects attitudes, effectiveness, quantity, and barrier reduction
Webster, Trevino, & Ryan (1993)	Computers in Human Behavior	Control, Attentional focus, Curiosity, Intrinsic interest	Examine factor structure/ correlates of flow in human computer interactions

Having reviewed the literature of flow in an HCI context, it is clear that there are common elements that are especially pertinent. I use an HCI-based conceptualization of flow for this paper because GBAs are always administered through a technologically-mediated form and I believe it is important to take the effect of this technology on flow into account. I rely on Webster, Trevino, and Ryan's four-dimensional definition of flow because it captures the elements that are common and likely the most theoretically relevant for HCI (1993). In their paper the authors review and define flow in the context of HCI, and validate the four dimensions of control, attentional focus (which I see as very similar to concentration and loss of self-consciousness), intrinsic interest, and curiosity (please see Figure 2 for a visual of these elements). While curiosity is not as widely studied in the flow literature, it does play a part in other related literatures like that of intrinsic motivation (Malone, 1981) and cognitive absorption (Agarwal & Karahanna, 2000). Curiosity is important in flow with technology because it explains why piquing one's sensory and cognitive facilities leads to continued engagement in the

activity. The user wonders what content or surprises he or she will find ahead, and continues on course to satiate this interest. I see this as especially relevant in GBA because it is highly novel. I expect curiosity to be an important element of flow, along with the aforementioned three, in regards to GBAs. Moving forward, I treat these four elements as a composite for an overall flow variable.

Figure 2. *Elements of flow used in the present study*



Inducing flow from game-based assessments

How is it then that users go from engaging with a GBA to experiencing a flow state? A number of studies have looked at the relationship between gaming and flow. One set of authors found that both violent and non-violent video games have the ability to induce flow (Chiang et al., 2011) and another found that playing against either a human or computer-controlled opponent in a game was related to feelings of flow (Weibel, Wissmath, Habegger, Steiner, & Groner, 2008). Other research found that flow experiences can act as a precursor to addictive gaming behaviors (Chou & Ting, 2003). Bowman postulated that the appeal of Pac-Man was due to the flow characteristics of clear goals and immediate feedback (1982). In the learning arena, some research has found that flow has a role in increasing engagement and subsequent

perceptions of learning in game-based learning environments (Hamari et al., 2016). In summary, there is strong empirical support for this link.

The same is true conceptually. One can make connections between GBAs overall and flow, but these arguments would inevitably be centered on the game elements that make up the GBA, as they are the building blocks. For example, I could say that some GBAs assign the applicant an avatar, and that this personification draws the applicant into the assessment. The connection then is the avatar, which is at the element level. This means that the best illustration is to make linkages between key game elements that can be present in GBAs, and the dimensions of flow used in this study. To choose game elements that are generally present in game-based tools, I return to the literature reviewed earlier (Table 1). My goal is to identify game elements that reoccurred across the taxonomies, and that could be easily adapted to a selection context.

The first frequent element is focused goals and rules. The rules portion refers to guiding principles that direct the player (Wilson et al., 2009). Goals elucidate the outcomes that result in wins, and are integral to the motivational structure of the game (Garris et al., 2002). Focused rules and goals feature in nine out of the ten game element taxonomies covered in Table 1. This means that they are a fundamental aspect of games. I include them here for a few reasons in addition to their centrality. Focused goals and rules are highly motivating (Locke & Latham, 2002), so they are effective when designing an assessment to be engaging. Goals and rules can motivate without adding confounding noise the way other motivating game elements do (for example, *interaction* in which construct measurement is contaminated by helping behaviors from other participants). They also do not require high investment the way other game elements do (i.e. sensory stimulation, 3D environments) and thus are useful for a greater range of situations.

The second element of interest is feedback. Feedback provides information on the accuracy of the player's behavior, and the discrepancy between one's current state and the goals of the game (Shute & Ke, 2012). Feedback is one of the most important elements of a game because it keeps the player engaged and working towards goals. It can also function as a form of reward. Feedback is included as a general element because it hard to keep out of a game scenario- it is fundamental to many other game elements such as points systems, leaderboards, and badges. While it is explicitly mentioned in only about half of the taxonomies, it is often present in some form through other game mechanisms.

The last element used for illustration is challenging tasks. Challenging tasks act as barriers between the player's current state and the end goals, so they contribute to the fun and competition of the game (Wilson et al., 2009). There has to be some ambiguity for a challenging task to be present, so the challenges of the game cannot be too obvious. Challenges are also essential to games because without them, players would cease to play. This is represented in the fact that seven out of the ten taxonomies include challenging tasks. It is important to note that challenging tasks can be included in traditional selection measures, especially in cognitive tasks, work samples, and simulations. However, because they are almost always present in effective games they are still included here in order to make linkages to flow.

I have identified three focal game elements that are common across taxonomies and in some cases also easily adapted to a selection scenario. I now explain how these three elements relate to the four elements of flow, to show the natural progression from engaging with a GBA to experiencing a flow state.

The first link that can be made is between the game element of focused goals/rules, and the attentional focus found in flow. Both rules and goals in a game setting require engagement on

behalf of the player, lest he or she misstep and be penalized. A player has to monitor him or herself to follow rules, and depending on individual characteristics, he or she may engage in meta-cognition about the discrepancy between the current state and the goal state. Both of these activities demand cognitive and affective resources (Baumeister, Muraven, & Tice, 2000; Muraven & Baumeister, 2000) including attentional focus.

The second link is between the game element of feedback and the control one feels in flow. Games allow some measure of control (otherwise it would be called a movie), and the feedback in game-based tools allows players to incorporate this feedback into their gaming strategy and manipulate the environment accordingly. This then means that he or she gains information about available resources and performance discrepancies (feedback), and can affect change as needed (using control). The tight coupling between feedback and subsequent control embeds the player into a state of flow.

Lastly, the challenging task element is connected to both intrinsic interest and curiosity. The Job Characteristics Model sheds light on this by linking skill variety, task identity, and task significance to intrinsic motivation at work (Hackman & Oldham, 1975). While these three job dimensions are not perfect correlates, they certainly illuminate the idea that tasks that require adaptation, a diversity of skill, and are meaningful will increase intrinsic motivation. I expect the logic to apply to games and flow; when a person is being stimulated he or she will be driven to continue engaging with the GBA. The logic behind the challenging task and curiosity link is similar, and rests on one key characteristic of challenging tasks in games: they have to be ambiguous. The player should not be able to fully anticipate the next step in the game. The mystery of the game captivates the user and invokes curiosity. Challenging tasks, and their

ability to interest people and make them eagerly anticipate what is coming next, are a natural precursor to flow.

The connections listed here are by no means an exhaustive list of the links that can be made between game elements and flow. There are many more game elements that can be used in a GBA, and other potential dimensions of flow that are not discussed here. For example, one can see the natural connection between the element of novelty and the flow dimension of intrinsic interest, or time pressure and flow's attentional focus. This section gives a sampling of the ways GBAs and flow can be tightly interwoven. I propose that the relationship between GBAs and flow is substantially different than the relationship between traditional computerized measures and flow. With that I suggest:

Hypothesis 1: There will be a higher level of flow when participants take a game-based assessment than when they take a traditional assessment.

With the link between GBAs and flow being made, I move on to describe potential moderators of this relationship.

Moderators to the GBA-flow relationship

There are three moderators that have particular relevance in a game-based context: technology self-efficacy, openness to experience, and one that I introduce for the purposes of this study and call *preference for games*. Technology self-efficacy and openness to experience have been explored in computerized settings (please see Table 3 for a summary of past work of the two moderators in computer or mobile test settings). Preference for games is intended to capture a specific bias or liking of game-based tools and systems. I discuss them each in turn to better explain characteristics that can enhance or inhibit the experience of flow.

Table 3. *Previous work on technology self-efficacy and openness to experience in studies about computerized/mobile testing*

Author and Year	**Construct name**	**Scale**	**Assessment Method**	**Findings**
		Technology self-efficacy		
King, Ryan, Kantrowitz, Grelle, & Dainis (2015)	Computer/mobile self-efficacy	4-item measure from Barbeite & Weiss (2004)	Computerized and mobile customer service orientation, cognitive ability, and supervisor SJT	Not significantly related to test reactions or performance
Saadé & Kira, 2009	Computer self-efficacy	Developed for study	Online learning management system (LMS) with tests related to management information systems and information technology	Computer self-efficacy mediated the relationship between anxiety and perceived ease of use of the LMS
Wiechmann & Ryan (2003)	Computer self-efficacy	8-item measure from Levine & Donitsa-Schmidt (1997)	Computerized in-basket to assess management experience	A combined construct assessing reverse-scored computer self-efficacy and anxiety was negatively related to liking, process fairness, face validity, test ease, and self-assessed performance
		Openness to experience		
Hertel, Naumann, Konradt, & Batinic, 2002	Openness to experience	NEO FFI derived from Costa & McCrae (1985) and translated from Borkenau & Ostendorf (1993)	Computerized personality measure based on the Big Five	No differences in openness between sample who opted for computerized measure and those who opted for paper-and-pencil measure using convenience sampling
Salgado & Moscoso (2003)	Openness to experience	IP 5F from Salgado (1998)	Computerized personality measure based on the Big Five	Not significantly related to test reactions
Wiechmann & Ryan (2003)	Openness to experience	12-item measure from Costa & McCrae's 1992 NEO-PI	Computerized in-basket to assess management experience	Openness to experience positively related to face validity

Technology self-efficacy

Although GBAs are highly engaging and can lead to a flow state, not everyone will experience this. One of the primary reasons for this can be comfort with the medium being used. GBAs are most often administered through a computer, tablet, or smartphone. This means that an individual's belief about his or her capacity to use technology is an important factor in predicting flow during GBA use. I use technology self-efficacy to capture an individual's feelings of competence in using computers and smartphones and describe how low self-efficacy will impede certain elements of flow.

The first relevant element of flow, control, posits that a person should feel he or she can modify or exert command over the situation. If the technology is unfamiliar, the user is prevented from feeling empowered because he or she cannot fully manipulate the device. This impedes the experience of flow as one is flustered at his or her lack of confidence in navigating technology. Bandura's work on self-efficacy would support this, suggesting that belief about one's ability affects whether the person thinks he or she can successfully execute the behavior needed to reach an outcome (1977). In other words, self-efficacy affects if a person thinks he or she can control the event. Judge and colleagues have similarly suggested that self-efficacy and locus of control share a common core (Judge, Erez, Bono, & Thoresen, 2002). Therefore I expect that when the user feels he or she lacks the skills to navigate the assessment, he or she will also feel less control over the GBA.

Having low self-efficacy beliefs with regards to technology will also take away attentional focus from the assessment itself. Attentional focus allows for immersion, and for the user to ignore the passing of time (Csikszentmihalyi, 1990). When that attention is directed both toward the task and towards meta-cognitively deciding how to use the tool to complete the task,

flow is broken. One of the mountain climbers interviewed about his flow experience described this utter focus by saying "When you're [climbing] you're not aware of other problematic life situations. It becomes a world unto its own, significant only to itself. It's a concentration thing. Once you're into the situation, it's incredibly real, and you're very much in charge of it. It becomes your total world" (Csikszentmihalyi, 1990, p. 58). If using the technology competes with experiencing the assessment, either because he or she cannot navigate the space or is feeling anxious about his or her capabilities, the user will find it near impossible to enter flow.

> *Hypothesis 2a: Technology self-efficacy will moderate the relationship between a game-based assessment and flow in that those who are more self-efficacious with technology will experience higher rates of flow. No moderating effect is expected for a traditional assessment.*

Openness to experience

In addition to accepting the medium through which the assessment is delivered, candidates must also be receptive to the *idea* of a gamified tool. Gamification and game-based tools are by no means universally accepted: a recent study by the Pew Research Center asked a number of tech analysts and stakeholders to share their thoughts on gamification. While many responses were positive, those surveyed voiced concerns about gamification being a passing fad, or about it being too prevalent. Some skeptics criticized games' inability to capture the full range of "human motivations, perceptions, cognitions, and practices" and saw gamified tools as a poor facsimile (Anderson & Rainie, 2012, p. 5). It is hard to say, in a hiring situation, whether these skeptics would prefer lower fidelity tools like computer measures, or if they reject anything other than face-to-face assessment but the point is that not all candidates will be supportive of games that play in the gray area between the two. In Pew's analysis, participants also said that

gamifying everything was pandering to our "already over met desire to be entertained" (Anderson & Rainie, 2012, p. 6). Some pointed out that companies want to manipulate our base instincts by incorporating games in order to get higher engagement (be it by sales or at work). One stakeholder shed light on these critiques, pointing out that humans do not yet understand the games "generate alternate realities" and that because we only see them as a means of entertainment, we cannot grasp their full utility (Anderson & Rainie, 2012, p. 6). Games are pushing the edge of innovation and while they are being increasingly incorporated in our daily life (Chou, 2017), they are still novel. This means that in order for candidates to react positively to them in a selection context, the candidate must be open to this novelty.

I posit the use of the personality trait openness to experience to identify candidates who will be able to experience flow while participating in GBAs. Specifically I rely on Costa and McCrae's facets because they have been extensively validated (McCrae, 2002) and widely used in the psychological literature (Brown & Ryan, 2003; Fredrickson, Tugade, Waugh, & Larkin, 2003; Judge & Bono, 2000; Whiteside & Lynam, 2001). According to their model, openness has six facets: "(a) ideas (e.g., having intellectual curiosity), (b) actions (e.g., valuing experimentation and learning), (c) fantasy (e.g., having an active imagination), (d) aesthetics (e.g., being intrigued by art, poetry, and music), (e) feelings (e.g., often experiencing strong emotions), and (f) values (e.g., believing that moral issues and social policies should change rather than be based on religion or principle)" (Colquitt, Hollenbeck, Ilgen, LePine, & Sheppard, 2002, p. 403; Costa & McCrae, 1992). Specifically, I hypothesize that those who have a high standing on the facets of *ideas*, *actions*, and *fantasy* will be the most likely to experience greater flow experience because there are so many parallels between these facets and elements commonly found in games.

The *ideas* facet is relevant because games are often challenging, employing different levels and some ambiguity to keep the user motivated (Bedwell et al., 2012). Gamifying is popular in educational arenas for the very reason that it meets intellectual curiosity needs; the chance to learn in a problem-based scenario, and the balance between one's skills and the requirements of the job captivates many participants (Dickey, 2005; Kiili, 2005; Landers, 2014). Those who are intellectually curious and willing to wrangle with the abstruse and surprising nature of games will be more likely to feel absorbed by them.

Candidates high in *actions*, or those who want to learn and experiment, will also find merit in games. Games often allow for some degree of control and protection from early mistakes. In many gamified tools, including GBAs, participants choose their path (Shute & Ke, 2012; Wilson et al., 2009). These choice options range in scale, from which avatar represents you to which planet you visit next on your expedition (Ratan & Dawson, 2015). This control is satisfying for those who crave experimentation, because they can shape their own progress. Similarly, many games offer adaptive challenges where the situation gets harder as one's performance improves (Bedwell et al., 2012; Shute & Ke, 2012). This allows players to experiment with new strategies. Candidates can also learn as they go in adaptive situations because the landscape is constantly changing. Those with a high standing on the *actions* facet will respond well to the protective nature of many games. Oftentimes users are given a buffer from early mistakes as they learn the rules of a new game or level; they can fail without much consequence (Dickey, 2005). The control, coupled with adaptive challenge and protections, makes for an ideal place to try new tactics while learning.

Lastly, the *fantasy* facet is a natural fit for game-based tools. Games by their very nature exist in an alternate reality. They require some buy-in from the user from the beginning. Games,

and GBAs, often take this a step further by introducing imaginative narratives or characters (e.g. Mislevy et al., 2014; Artic Shores, n.d.). Those high in openness are more likely to accept or even enjoy these inventive factors, further encouraging a transition into flow.

I speculate that games align well with and can satiate those who are high in openness. There will then be an interactive effect where:

> *Hypothesis 2b: Openness to experience, specifically the facets of ideas, actions, and fantasy, will moderate the relationship between a game-based assessment and flow in that those who are higher in openness will experience higher rates of flow. No moderating effect is expected for a traditional assessment.*

I do not expect there to be differential effects at the facet level, and I conceptually treat the openness to experience moderator as a composite of the three facets.

Preference for games

In addition to openness to experience, some work has started exploring other game-based preferences and their effect on participant reactions. The reasoning behind this is that positive attitudes towards or a preference for games can offer fresh insight that cannot be gained through a broader openness measure, because preference for games speaks specifically to a proclivity for game elements, as opposed to general new experiences. For example, Landers and Armstrong tailored their Technology-Enhanced Training Effectiveness Model to incorporate attitudes towards game-based learning, and found that participants with more positive attitudes anticipated greater benefits from the learning (Landers & Armstrong, 2015).

Building on this idea that game-related attitudes offer prediction of important reactions, I introduce a measure of preference for games to capture the extent to which a candidate has a bias toward game-based tools and processes. Much of this measure relates to the discussion in the

openness to experience section about how many people are opposed to incorporating gamification into everyday life; there are those who are receptive to these innovations (those high on openness to experience) and those who will like the changes because they gravitate toward games.

I posit that those who have a preference for games (those who score highly on the measure) would be more likely to experience flow when interacting with a GBA because they have an intrinsic interest in the experience. I also suggest that those with a preference would devote more attentional resources to the GBA; both the heightened intrinsic interest and attentional focus would allow a candidate with a preference for games to experience a greater rate of flow.

Hypothesis 2c: Preference for games will moderate the relationship between a game-based assessment and flow in that those who have a preference for games will experience higher rates of flow. No moderating effect is expected for a traditional assessment.

Next I move on to discuss how GBAs and flow affect reactions to the assessment process and the organization overall. I begin with an overview of the main frameworks in the applicants' reactions literature before explaining how my hypotheses fit into these frameworks. In this section, I explain concepts from the "upper" half of the model, specifically the relationship of flow, perceptions of job-relatedness, and justice perceptions.

Applicant reactions

Applicant reactions are concerned with how applicants view the process by which they apply for jobs. Researchers in this area look to see how perceptions during the hiring process affect perceptions of the organization and subsequent outcomes, such as accepting offers, recommendation intentions, etc. (Ryan & Ployhart, 2000). This area has grown in prominence in

the last three decades for a few reasons. Candidates who have a negative experience during the selection process are less likely to accept offers, and more likely to engage in behaviors the company would find negative such as dissuading friends and colleagues from applying, or filing legal complaints against the company (Hausknecht et al., 2004). With the surge of research, two primary streams have emerged.

The first is a justice and fairness approach to applicant reactions (Gilliland, 1993; 1994). This has been looked at from both a procedural justice lens (how fair was the procedure used to decide the outcomes) and a distributive justice lens (how fair are the outcomes). In the case of procedural justice, applicants look to see the extent to which the process used to decide their outcomes is fair. Theorizing in this area says in order for the process to be seen as fair, it should be: applied consistently, free of bias, make use of accurate information, have a self-correcting mechanism, conform to an ethical code, and ensure that the opinions of affected groups are taken into account (Colquitt, Conlon, Wesson, Porter, & Ng, 2001; Leventhal, 1980). A branch of procedural justice, referred to as interactional justice, suggests that what is said to applicants and how it is said also affects justice perceptions (Bies & Moag, 1986). This has great importance in selection in regards to face-to-face interactions and communication, but has been studied less in relation to assessment.

With distributive justice, individuals evaluate their outcomes to see if they are consistent with a distribution rule (Gilliland, 1993). There are three such rules that are especially relevant for selection. The first relates to *equity*; applicants will evaluate their returns against what they input to the system, and how this compares to a referent other. When the inputs are perceived to be greater than the outputs, applicants feel a lack of equity and can become dissatisfied with the process, the organization, and themselves (Hausknecht et al., 2004). Distributive justice also

hypothesizes that in a selection context, *equal* outcomes should be given regardless of job-irrelevant characteristics like sex or race (Gilliland, 1993). The final potential rule of distributive justice is that rewards should be distributed based on *need* so that those who have special needs should be given preferential treatment. The ideal balance of these three can depend on the individual, context, or both (Bond, Leung, & Wan, 1982; Cook & Hegtvedt, 1983; Gomez-Mejia & Welbourne, 1994).

The second major stream examines test-taker attitudes during the hiring process (Arvey, Strickland, Drauden, & Martin, 1990; Chan, Schmitt, Jennings, Clause, & Delbridge, 1998). This has been referred to as the *test perceptions model* (Ployhart & Harold, 2004) and researchers in this area look at the role of motivation, anxiety, and attitudes on test performance and how these can relate to subgroup differences. A host of work by Chan and colleagues shows evidence for the idea that these perceptions matter; face validity affects cognitive test performance through the mediator of motivation (Chan, Schmitt, DeShon, Clause, & Delbridge, 1997), belief in the test is positively related to both cognitive and personality test performance (Chan, Schmitt, Sacco, & DeShon, 1998b), predictive validity perceptions positively correlate with cognitive ability test performance (Chan, 1997) and so on.

I use concepts from both of these frameworks in my study. The relevant concepts (perceptions of job-relatedness and justice perceptions) are further discussed below.

Flow and perceptions of job-relatedness

I posit that the experience of flow will lead to a number of outcomes relevant to applicant reactions. The first pathway that I focus on is the relationship between flow and perceptions of job-relatedness. Perceptions of job-relatedness have been defined as "the extent to which a test either appears to measure content relevant to the job situation or appears to be valid" (Gilliland,

1993, p. 703). Although the author in this original paper tried to distinguish perceptions of job-relatedness from face validity, later conceptualizations adopted face validity as one factor of job-relatedness, and perceived predictive validity as the other (Bauer et al., 2001; Chan & Schmitt, 2004; Chan et al., 1998a; Ryan & Ployhart, 2000). This is due in part to the work of Smither and colleagues who treated perceptions of job-relatedness as a composite of predictive and face validity (Smither, Reilly, Millsap, Pearlman, & Stoffey, 1993). For this paper, I treat perceptions of job-relatedness the same way Smither and colleagues do. I define the factors along similar lines as previous researchers and say face validity is the extent to which content of the assessment appears to be related to content of the job, and predictive validity as the extent to which an assessment will predict future job performance.

Because flow is foreign to the selection literature, there is virtually no work connecting it with perceptions of job-relatedness. In addition to the dearth of research, it is not theoretically clear what the relationship should be. Below I examine different perspectives that could explain the connection. These perspectives make competing predictions, so I divide these into positive and negative forecasted directions (please see Table 4 for a summary of these possible results):

Positive relationship between flow and perceptions of job-relatedness. One scenario is that those who experience flow, a positive and reaffirming state, will then have a generally positive reaction to the assessment. When users are in flow, they report being happier, more potent (active, alert, and excited), more satisfied, and possessing higher positive affect than when there are not in flow (Chiang et al., 2011; Csikszentmihalyi & LeFevre, 1989; Konradt, Filip, & Hoffman, 2003; Rogatko, 2009). The feelings-as-information perspective submits that when someone is in a positive mood, he or she will make more positive judgments (Schwarz & Clore, 1983; Wyer & Carlston, 1979). The halo effect can also be used to explain this. Studies have

shown that a global attribution about an object can alter independent assessments of its characteristics, even if there is sufficient information to form more nuanced opinions (Nisbett & Wilson, 1977). Therefore we may expect that the buoyancy candidates have coming off a flow-like state will result in an overall positive evaluation of the assessment, including its perceived job-relatedness.

Another scenario draws on the importance of test method. Researchers have not fully been able to parse the importance of test method (how the assessment is administered) and test content (what the assessment measures), but some work indicates that interactivity and rich media positively affect perceptions of job-relatedness (Bruk-Lee et al., 2016; Chan & Schmitt, 1997; Richman-Hirsch, Olson-Buchanan, & Drasgow, 2000; Toldi, 2011; Tsacoumis, 2015). It is possible that the engagement, control, and inbuilt media the GBA offers will lead users to feel it is job-related. However, there is a body of research that indicates media richness does not matter or that more complex media can hurt validity perceptions (Kanning, Grewe, Hollenberg, & Hadouch, 2006; Lievens & Sackett, 2006; Potosky & Bobko, 2004; Truxillo & Hunthausen, 1999; Wiechmann & Ryan, 2003), so takeaways from this research are mixed.

Negative relationship between flow and perceptions of job-relatedness. The most obvious reason that GBAs may relate negatively to perceptions of job-relatedness is they are not necessarily designed to mirror work tasks and environments. Job simulations and work samples enjoy high perceptions of job-relatedness ratings because they are similar to the target job and job tasks (Hausknecht et al., 2004; Rynes & Connerly, 1993; Smither et al., 1993; Tsacoumis, 2015). These tools make use of behavioral consistency and point-to-point correspondence with the criterion to predict how the individual will behave on the job (Funder & Colvin, 1991). Assessments like these can mirror a work setting, such as an online job simulation that virtually

leads one through a traditional office space. They can also elicit highly job-similar behaviors, such as asking the participant to write an email or have a phone call with a role player. While GBAs measure abilities and traits that are important to job performance such as cognitive ability and personality, traditional measures of these constructs do not fare as well in ratings of perceptions of job-relatedness as the aforementioned work-related assessments (Macan et al., 1994; Pulakos, 2005; Truxillo, Steiner, & Gilliland, 2004). Therefore, due to the lack of work contextualization and the low job-relatedness ratings of the constructs GBAs normally assess, these tools may not induce a positive relationship between flow and perceived job-relatedness.

Exploratory question 1: What is the relationship between flow and perceived job-relatedness?

Table 4. *Framework to understand the relationship between flow and perceptions of job-relatedness*

Characteristic	Flow is positively related to perceptions of job-relatedness	Flow is negatively related to perceptions of job-relatedness
Description	(1) When GBAs induce flow, applicants are in a high positive affect, satisfied state and assign positive evaluations to the assessment overall (2) Interactivity and high media-richness induce higher perceptions of job-relatedness because they increase engagement and control	A lack of job-related stimuli and failing to evaluate job-relevant behaviors (such as talking to a supervisor or writing an email) will lead applicants to feel the assessment is inappropriately contextualized and have low perceived job-relatedness
Boundary conditions	This possibility is more likely when: a) The GBA makes use of more complex graphics, sounds, and narratives	This possibility is more likely when: a) The GBA does not attempt to replicate work problems or environments b) The GBA measures traits and stable abilities (personality, cognitive ability) rather than work-related knowledge
Relevant theories	(1) Feelings-as-information (Schwarz & Clore, 1983; Wyer & Carlston, 1979) Halo effect (Nisbett & Wilson, 1977) (2) Media richness theory (Daft & Lengel, 1986)	Justice perceptions (Gilliland 1993; 1994) Behavioral consistency (Funder & Colvin, 1991)

While the relationship between flow and perceived job-relatedness is unclear, we can make somewhat more substantial hypotheses about the direct relationship between assessment method and perceived job-relatedness.

Assessment method and perceived job-relatedness

GBAs are attractive because they have the potential to increase engagement and enjoyment through the incorporation of game elements. These same game elements can also

make gamified or game-based systems seem inappropriate for work or unnecessary. Game thinking is still burgeoning in organizations, and can be viewed with a skeptical approach. This is exacerbated by the fact that most game-based assessments currently available measure cognitive or personality aspects, which tend to have lower job-relatedness ratings from candidates (Macan et al., 1994; Pulakos, 2005). The novel and sometimes fantastical nature of GBAs means that, when flow is controlled for, we may expect lower ratings of perceived job relevance than more traditional tools to which candidates have been exposed and accustomed.

Hypothesis 3: A traditional assessment will have a stronger direct positive effect on perceived job-relatedness than will a game-based assessment, controlling for the effect of flow.

Having discussed the connections between the assessments, flow, and perceived job-relatedness, I move on to discussing the next, well-established link in the model.

Perceived job-relatedness and justice perceptions

A large body of work connects perceived job-relatedness to justice perceptions (Chan & Schmitt, 2004; Elkins & Phillips, 2000; Gilliland, 1993; Kluger & Rothstein, 1993; Schmitt, Oswald, Kim, Gillespie, & Ramsay, 2004; Smither et al., 1993; Truxillo, Bauer, & Sanchez, 2001). Perceived job-relatedness is important for perceptions of procedural justice because it means that appropriate information is being collected during the process (Leventhal, 1980). This use of accurate information is one of the basic tenants of procedural justice. People want to feel they are being evaluated by information that matters in high stakes situations such as job opportunities. The most recent meta-analysis of applicant reactions found a strong relationship between perceptions of job-relatedness and procedural justice (r = .51), as well as its two factors (r = .50 for face validity, r = .54 for predictive validity) (Hausknecht et al., 2004).

Perceptions of job-relatedness also connect to distributive justice meta-analytically (r = .29 for face validity, r = .34 for predictive validity), though to a lesser degree than procedural justice. Gilliland's seminal paper does not provide a rationale for why perceptions of job-relatedness are important to distributive justice, and consequently many studies look at how perceptions of job-relatedness contribute to overall perceptions of fairness instead of to the two facets independently (Bauer et al., 1998; Chan et al., 1998a; Macan et al., 1994; Ryan, Greguras, & Ployhart, 1996; Schmitt et al., 2004; Zibarras & Patterson, 2015). One could speculate that because some applicants want job decisions to be distributed equally and/or equitably, and without regard to job-irrelevant characteristics, collecting job-relevant information makes it more likely that pertinent information will be used to make the distribution decision.

I provide a few examples of the work that connects perceived job-relatedness to procedural and/or distributive justice. Gilliland found perceived job-relatedness influenced distributive justice perceptions more for applicants who were rejected (1994). Ryan and Chan found perceived job-relatedness to be correlated with both process fairness and outcome satisfaction for a sample of licensed psychologists (1999). Another study found that job relevance was a significant predictor of process fairness using regressions in a police sample (Truxillo, Bauer, Campion, & Paronto, 2002). Some of these same authors found that both content and predictive validity related positively to procedural fairness for a video-based test (though the same was not true for a multiple choice based test) (Truxillo et al., 2001). Overall, there is strong evidence to support the following hypothesis:

Hypothesis 4: Perceptions of job-relatedness will be positively related to perceptions of procedural and distributive justice.

Because procedural and distributive justice are highly correlated (Cropanzano & Ambrose, 2001; Colquitt et al., 2001), I make joint hypotheses for them throughout this paper.

While I discuss the relationships in the applicant reactions literature it is important to keep the role of the selection decision in mind. The rationale for this can be found in the next section.

Moderating role of selection decision in the perceptions of job-relatedness to justice perceptions relationship

One of the biggest influences on how applicants feel about the selection process is the outcome itself; that is, whether or not they were issued an offer (Ryan & Ployhart, 2000). Therefore examining outcomes such as justice perceptions, which are heavily influenced by the distributive outcome, without discussing the role of the decision is remiss. Quite a bit of work has already examined the role of decisions in fairness perceptions. Kluger and Rothstein found that students who did not meet the hiring standard saw the test as less fair than those who did (1993). In Lounsbury's and colleagues' study, those who failed the employment test rated the test as less fair (Lounsbury, Bobrow, & Jensen, 1989). A later study replicated these effects (Bauer, Maertz, Dolen, & Campion, 1998), and other studies have found similar decreases in justice perceptions for those who are told they did not meet the hiring requirements (Bernerth, Feild, Giles, & Cole, 2006; Macan et al., 1994; Smither et al., 1993; Weichmann & Ryan, 2003).

The finding is generally that being rejected from an organization is related to lower justice perceptions, both procedurally and distributionally. Taken with the literature I reviewed previously on the positive relationship between perceived job-relatedness and justice, one may expect that if someone does not perceive the GBA as job-related, he or she will have very negative fairness perceptions if he or she is rejected from the job. To further break down this

doubly negative effect, I first look at distributive justice. People have a tendency to think they are better or superior to others. This has been referred to as superiority bias, illusory, superiority, and unrealistic optimism (Alicke, 1985; Codol, 1975; Hoorens, 1993; Weinstein, 1980). When people are not given the outcome they expected, especially when it is related to their abilities and skills in as focal a context as work, distributive justice perceptions are affected. Placing moderators such as hiring ratios and perceived competitiveness aside, applicants expect that they have the skills and abilities to get hired and feel violated when they are not. Now if this is taken in combination with an assessment that applicants do not feel appropriately simulated the requirements of work (low perceived job-relatedness which violates the requirements of procedural justice), there will be a particularly negative impact on justice perceptions. This is an effort to lessen the ego-threat associated with being rejected (Chan et al., 1997; Weiner, 1986). Individuals engage in a self-serving bias where they attribute their failure to aspects of the test or process (in this case, its unfairness). However, when the applicant is given the offer, he or she experiences distributive justice and is less inclined to rely on perceived job-relatedness as a cue for justice perceptions. Perceptions of job-relatedness are a more dominant cue of justice perceptions only when people are rejected from the job. Therefore, I suggest that:

> *Hypothesis 5: The relationship between perceived job-relatedness and justice perceptions (both procedural and distributional) is moderated by the selection decision such that the relationship is stronger when the individual is not offered a job and weaker when the individual is offered a job.*

The preceding section covered an important link in the applicant reactions literature: I posited arguments for how flow may or may not relate to perceived job-relatedness, and how perceived job-relatedness (if it is present) in turn should be positively linked to justice

perceptions. This explicates one half of the model. For the rest of this section, I discuss the "lower" half of the model. I explore the role that GBAs and flow have on perceived person-organization (PO) fit. I then connect perceived PO fit to organizational attractiveness to show the second mechanism by which an applicant may feel positively about an organization's process after engaging with a GBA. The concepts in the lower half of the model are described in the following sections.

Game-based assessment, flow, and perceived PO fit

Perceived PO fit. Fit research originates from person-environment fit theories, which say that when an individual fits or matches the environment, good outcomes occur (Holland, 1985; Tom, 1971. Fit theories have evolved over time, with some of the earliest work looking at the match between individuals and different vocations (Parsons, 1909). Later theories, such as the theory of work adjustment, used fit to explain satisfaction (Dawis & Lofquist, 1984). Around the same time, Holland discussed the benefits of congruence between a person and his or her work environment (Holland, 1959; 1997). More recent iterations have focused on the fit of a person with the job and the organization (Kristof-Brown, Zimmerman, & Johnson, 2005). Since many organizations use standardized selection processes across jobs, person-organization fit has been a heavy topic of study in the recruitment and hiring areas (Cable & Judge, 1996; Carless, 2005; Chapman, Uggerslev, Carroll, Piasentin, & Jones, 2005; Dineen, Ash & Noe, 2002). I take that focus here as gamified tools are often used across the organization, and reflect on the organization as a whole (e.g., Deloitte's gamified leadership academy, PWC's recruitment game Multipoly, Loreal's innovation game Brandstorm).

Defining PO fit in these literatures has been elusive, in part because what is meant by "fit" can vary depending on the perspective. Kristof arrived at a comprehensive definition by

stating that PO fit is "the compatibility between people and organizations that occurs when: (a) at least one entity provides what the other needs, or (b) they share similar fundamental characteristics, or (c) both (Kristof, 1996, pp. 4-5). In order to highlight the perspectives that feed into this construct, I review them in the next section (please see Table 5 for a summary of fit perspectives and their corresponding theories).

Table 5. *Fit perspectives and corresponding theories*

Tradition	Theory	Author	Premise
Complementary		Muchinsky & Monahan (1987)	When a person's characteristics make whole the environment or add to it what is missing
Complementary	Needs-supplies	Caplan (1987), Edward (1991)	Organization satisfies individual's needs, desires, or preferences
Complementary	Abilities-demands	Caplan (1987), Edward (1991)	Individual has the abilities required to meet the organization's demands
Supplementary		Muchinsky & Monahan (1987)	Supplements, embellishes, or possesses characteristics which are similar to other individuals
Supplementary	Similarity-attraction	Byrne (1971)	Individuals are attracted to others based on similarity, and are likely to interact and make connections with similar others
Supplementary	Self-categorization theory	Turner, Hogg, Oakes, Reicher, & Wetherell (1987)	A greater amount of similarity in social categories (background, attitudes, and lifestyles) means people will form a common identity with referent others
Supplementary	Social identity theory	Tajfel (1974)	Individuals will identify social groups to which they belong, and will view these groups as prestigious in order to enhance self-esteem
Supplementary	ASA	Schneider (1987)	Individuals are attracted to organizations whose values, goals, and culture are similar to the individuals' personalities, values, and goals; people leave when mismatch occurs
Supplementary	Social comparison	Festinger (1954)	Individuals will compare themselves with others, including others who are unknown
Supplementary	Balanced state theory	Heider (1958)	Individuals will maintain balance, including within their relationships, to maintain cognitive consistency

There are two (blurred) sets of dichotomies that demarcate the PO fit literature. The first comes from a needs lens. When the individual is looking to see if the organization can fulfill his or her desires, preferences, and needs, he or she would be coming from a *needs-supplies* perspective (Caplan, 1987; Edwards, 1991). This means the individual is looking for certain elements to be present, and would want to work in an organization that can meet those preferences. On the other hand if the point is to see if the individual can provide the abilities that the organization demands, the lens is then aptly *abilities-demands.* Here, the individual's abilities would need to be high enough that he or she can meet work goals and exhibit adequate job performance (but likely not so high that the person is bored or overqualified for the job). Demands- abilities fit is rarely studied from a PO perspective (for the exception see Hutcheson, 1999). It is possible that if a GBA falls closer to a work simulation that an applicant could evaluate abilities-demands fit from it. However in this study, I test the model using a tool that does not heavily simulate a work environment so I do not make hypotheses about this kind of fit.

For this study I take a needs-supplies perspective because I am interested in the candidates' reactions to the organization's image. I would expect that in the case of GBAs, applicants use the selection process to evaluate organizational values rather than the difficulty of the job because most current GBAs are domain general. Therefore, this will be a time for applicants to process signals about the organization's culture and values, rather than the specific job tasks. This fits with previous rationalizations that employees are looking to accept jobs that will gain them access to needed resources, so a needs-supplies perspective is the most sound (Cable & DeRue, 2002; Simon, 1951).

The second axis in the PO fit literature is supplementary versus complementary fit. In complementary fit, the "weakness or needs of the environment are offset by the strength of the

individual, or vice versa" (Muchinsky & Monahan, 1987, p. 271). Both needs-supplies and abilities-demands perspectives generally fall into this line of research (though there are reasonable arguments to be made to the contrary, see Ostroff, 2012; Westerman & Cyr, 2004). Complementary fit from the individual's side can be characterized as a desire for psychological need fulfillment, where attitudes of fit are influenced by an organization's ability to supply what is needed to meet an individual's desires (Edwards, 1991). In supplementary fit, one actor is said to embellish or possess characteristics that are similar to the other (Muchinsky & Monahan, 1987). In theories that take a supplementary perspective, such as the attraction-selection-attrition model or the similarity-attraction paradigm, individuals look to find shared features between themselves and a referent other. This similarity, or congruence, can have positive outcomes like satisfaction, psychological comfort, and enhanced communication (Cable & Edwards, 2004; Dawis & Lofquist, 1984; Erdogan & Bauer, 2005). Applicants looking for supplemental fit would want to see similarities between themselves and the organization, rather than a "filling the gap" approach offered by complementary fit.

Upon closer inspection, complementary fit is fraught with ambiguous measurement strategies. Complementary fit can entail filling a void, but it can also be denoted as fulfilling a preference. It is not always clear if this preference is aligned with the individual's existing values, or whether it is a gap he or she is looking to fill. This can come down to a simple choice in the measure of "I do" or "I can" to "I should" or "I prefer". When the measurement focuses on idealized situations, it can be interpreted as complementary. The individual is looking for something from the organization that he or she may not yet possess. Yet when the measurement is around enacted values (the dos or cans), this becomes supplementary fit because one is now

looking at similarity between what the individual already does or believes in, and the organization's values.

To further illustrate the misalignment, I provide two examples (please see Table 6 for a more thorough review of definitions and measurement of fit). In one study PO fit is said to be measured as a correlation between the value profiles of the organization and the individual (Chatman, 1989). However, the individual values are sorted according to how important the organizational characteristic is. This is likely to be interpreted as "what characteristics do you prefer in an organization", and an individual can easily answer with characteristics he or she is looking for but does not possess already. Despite the ambiguous measurement, the Chatman article conceptualizes fit as supplementary. Another example is the Bretz and Judge article where they define fit using the language of "match", yet measure fit as whether or not an applicant accepts a job from a hypothetical organization (1994b). This measurement is too vague to capture whether the applicant was thinking about fulfilling his or her needs, or about finding a place that is congruent. Perhaps part of the reason for this miscalibration is there are very few articles that intentionally measure complementary PO fit (for exceptions see Van Vianen, 2000 and Westerman & Cyr, 2004). The majority of articles rely on a supplementary perspective, though their measurement does not always align with that.

Table 6. *Definitions, measurement, predictors, and outcomes of person-organization fit*

Author	Type of Fit	Definition	Measure of Fit	Antecedent(s)	Outcome(s)
Ambrose, Arnaud, & Schminke (2007)	PO fit, supplementary	Congruence (or fit) between individual and organizational attributes and values; Kristof (1996) definition	Match between individual's level of cognitive moral development and organization's ethical climate		Commitment, job satisfaction, intention to quit
Bretz & Judge (1994a)	PO fit, supplementary	State of congruence between individual and environmental characteristics	Sum of difference between measures about organizational environment and personal needs		Tenure, satisfaction, career success
Bretz & Judge (1994b)	Unclear, fit measured through job acceptance intention	Match between individual characteristics and human resource system characteristics	Policy capturing approach where participants given profile of organization with manipulated characteristics and asked if they would accept an offer at the company	Individual values (individual work, locus of control, fairness, work/family conflict)	
Cable & Judge (1994)	PO fit, supplementary	Match between dispositions and pay systems	Relevant personality characteristic regressed on pay preferences	Individual values (materialism, locus of control, individualistic, self-efficacy, risk aversion)	

Table 6 *(cont'd)*

Author	Type of Fit	Definition	Measure of Fit	Antecedent(s)	Outcome(s)
Cable & Judge (1996)	PO fit, supplementary	Personal characteristics are aligned with organizational attributes	Measure of how much individuals feel their values match or fit values of current employees in the organization	Demographic similarity between job seeker and recruiter, perceived value congruence	Job choice, organizational commitment, job satisfaction, intentions to quit, recommendation intentions
Cable & Parsons (2001)	PO fit, supplementary	Compatibility between people and the organizations in which they work	Cable & Judge (1996) measure about individual value match with organization	Collective, formal, sequential, fixed, serial, and investiture socialization tactics; pre-entry value congruence	Turnover
Carless (2005)	PO fit, supplementary	Estimate the match between person and his/her work environment by estimating match between personality, attitudes, and values	Judge & Cable (1997) measure about match between individual's values, goals, and personality and those of the current employees at the referent organization		Organizational attractiveness, intentions to accept offer, job offer decision

Table 6 *(cont'd)*

Author	Type of Fit	Definition	Measure of Fit	Antecedent(s)	Outcome(s)
Chatman (1991)	PO fit, complementary	Congruence between patterns of organizational values and patterns of individual values	Correlating individual's Q-sort profile of desirable organizational values with organization's profile	Time with firm members, perceptions of similarity to successful firm members, firm acceptance ratio, number of offers, social interaction with firm members, time with firm mentor, formal training, perceptions of socialization process	Job satisfaction, intent to leave, length of membership
Dineen, Ash, & Noe (2002)	PO fit, supplementary	Objective PO fit as individual's actual, or measured, value congruence with an organization and subjective PO fit as level of perceived value congruence	Objective PO fit as correlation between individual profile and company profile and subjective PO fit as how well the values of the organization reflect the individual's values	Objective PO fit, level of PO fit feedback	Subjective PO fit, organizational attractiveness
Goodman & Svyantek (1999)	PO fit, complementary	Match of an individual's values with the value system in a specific organizational context and the potential effects of this match (or lack thereof)	Comparison of organization's perceived culture to individual's idealized culture		Contextual performance, task performance

Table 6 *(cont'd)*

Author	Type of Fit	Definition	Measure of Fit	Antecedent(s)	Outcome(s)
Kim, Cable, & Kim (2005)	PO fit, supplementary	Employees' subjective beliefs about how well their personal values match the organizational culture	Cable & DeRue (2002) measure about personal values matching organizational values	Institutionalized socialization tactics	
Lauver & Kristof-Brown (2001)	PO fit, supplementary	Compatibility between people and organizations that occurs when at least one entity provides what the other needs, they share similar fundamental characteristics, or both	Cable & Judge (1996) measure about individual value match with organization		Job satisfaction, intent to quit, contextual performance
Posner (1992)	PO fit, supplementary	Values congruency via alignment with the organization's six core principles	Clarity, consensus, and intensity of individual's alignment with organization's core stated principles	Demographic characteristics	Motivation, commitment, and teamwork

Table 6 *(cont'd)*

Author	Type of Fit	Definition	Measure of Fit	Antecedent(s)	Outcome(s)
Posner, Kouzes, & Schmidt (1985)	PO fit, supplementary	Extent to which personal values are compatible with values of the organization	Measure of value congruence where individuals rated how compatible their values are with the organization's		Feelings of personal success, organizational commitment, self confidence in understanding personal and organization goals, ethical behavior, job and personal stress, concern for stakeholders
Resick, Baltes, & Shantz (2007)	PO fit, supplementary	Measure of a person's perception of his or her compatibility with an organization's culture and members	Combination of Cable & Judge (1996) and Saks & Ashforth (1997) measures		Satisfaction with internship, intention to accept an offer, offer acceptance
Valentine, Godkin & Lucero (2002)	PO fit, supplementary	Kristof (1996) definition	Netemeyer et al. (1997) measure about similarity of organization's values to individual's values	Corporate ethical values, organizational commitment	
Van Vianen (2000)	PO fit, complementary	Kristof (1996) definition	Comparison of newcomer's culture preferences with organizational culture supplies		Commitment, turnover intentions

Table 6 *(cont'd)*

Author	Type of Fit	Definition	Measure of Fit	Antecedent(s)	Outcome(s)
Vancouver & Schmitt (1991)	PO fit, supplementary	Degree to which organizational members agree on the priorities of organizational goals	Comparing goal importance rating of one individual with all other members of the organization		Job satisfaction, organizational commitment, intent to quit
Westerman & Cyr (2004)	PO fit, supplementary and complementary	Correspondence between the individual's values and organization's existing value systems	Comparison of organization's value system and individual's ideal value system; comparison of averaged organizational personality profile and individual personality profile; comparison of organization's work characteristics and individual's ideal work characteristics		Employee satisfaction, organizational commitment, intention to remain

Despite this, there are a few authors that have examined the respective roles of the two forms of PO fit. Cable and Edwards tested three models, two that predicted value congruence or psychological need fulfillment would be the best predictors of attitudinal outcomes respectively, and one that posited simultaneous effects (2004). The simultaneous effects model predicted job satisfaction, intent to stay, and organizational identification better than did either individual model. In another paper, confirmatory factor analysis was used to see how employees distinguish among different kinds of fit. The authors found that employees are able to separate PO fit, needs-supplies fit, and abilities-demands fit (Cable & DeRue, 2002). Here, the authors viewed PO fit as supplementary and needs-supplies/abilities-demands as complementary, so the results indicated that employees are able to differentiate supplementary and complementary fit, and take a more nuanced view of complementary fit. Ultimately, the most recent meta-analysis in this area shows that the two types of fit do have differential relationships. Supplementary fit is more strongly associated with organizational commitment and intent to quit, and complementary (defined as needs-supplies) was more strongly associated with job satisfaction (Kristof-Brown et al., 2005). These findings provide support that both types of fit play a role an employee's experience with the organization.

I retain this dual view for my study. I believe that GBAs and flow experiences can elicit both complementary and/or supplementary fit. Specifically, I draw on signaling theory to explain the mechanisms by which these connections are made (Connelly, Certo, Ireland, & Reutzel, 2011; Spence, 1973). Signaling theory explains how a party can communicate by sending information (a signal) to reduce information asymmetry with another party. In signaling theory, the signaler relays a message to the receiver. The receiver interprets that signal, and makes a choice. That choice is seen as feedback to the signaler about the effectiveness of the signal.

Signaling is studied frequently in the recruitment process (Suazo, Martínez, & Sandoval, 2009), and I use it in the next section to explain how GBAs and the state of flow can act as a signal to the applicant about what the organization values.

Game-based assessment, flow, and the connection to perceived PO fit. Games are seen as engaging and innovative (Altman, 2015; Drew, Lamer, Bruk-Lee, LeVine, & Wrenn, 2012; Meister, 2013). Because GBAs are new and sparse in the selection arena, organizations that use them can benefit by appearing to be on the cutting-edge. GBAs can serve as a signal that the organization values innovation and creativity. Values such as innovation are frequently shown to be a desirable part of an organization's culture that leads to many positive outcomes (Chatman & Jehn, 1994; O'Reilly, Chatman, & Caldwell, 1991; Hurley & Hult, 1998; West & Farr, 1990).

Relatedly, the state of flow when using technology gives the applicant a sense of control, attentional focus, intrinsic interest, and curiosity (Webster et al., 1993). The flow state can also serve as a signal to the organization's values. When an applicant is feeling in control while engaging with the GBA, he or she may feel that the organization allows its employees to have control over their work, and that autonomy and independence is valued. Intrinsic interest may speak to the organization's emphasis on intellectually stimulating work. Curiosity can indicate that the organization is creative and imaginative in its approach. These signals- autonomy, independence, intellectual simulation, creativity, and imagination- appear again and again in work on values. For example, England includes autonomy and creativity in his list of Western-based, organizational values (1967). Rokeach discussed the values of independence, imagination, and intellectualism in his taxonomy, though this was not restricted to a work setting (1973). A few decades later, another iteration stated openness, creativity, and autonomy as part of its list of

values that an organization or individual can enact (McDonald & Gandz, 1991). Schwartz developed a set of ten types of values that validate in dozens of countries around the world (1992; 1994). Self-direction and stimulation rated among these. Although not all of these taxonomies are organizationally based, there are recurrent themes across work and non-work settings. One could also hypothesize that general human needs extend across environments and indeed, definitions of values include words like "transsituational", "universal requirements", (Schwartz & Bardi, 2001, p. 269) and "omnipresent" (McDonald & Gandz, 1992, p. 219). Researchers have stated that humans establish stable values over the course of our lives, and that these are unlikely to change upon entrance to an organization (Ravlin & Meglino, 1989). Therefore I take non-work taxonomies as additional support for the importance of focal values.

There are many parallels between GBAs, flow, and values that individuals and organizations can embody. The experiences with GBAs and in flow can signal to an applicant that similar experiences may await him or her upon joining the organization. Applicants may feel the organization chose specific assessments because it values the characteristics the assessment embodies, and the experiences that it elicits. An applicant may think that instead of staying with a traditional, self-reported, computerized measure, the organization opted for a tool that was more engaging because it cares about the applicant's experience and how the tool reflects on the culture of the organization. Once an applicant interprets the GBA as a signal of the organization's values, fit can occur two ways. An applicant may feel that he or she already possess the values (innovation, autonomy, intellectual curiosity, imagination, etc.) the organization is espousing and feel congruent with the organization. On the other hand, the applicant may recognize the signals and feel the organization can fulfill a need he or she has. The

values derived from the assessment should be the same, but applicants can take either a value congruence or need fulfillment perspective.

Hypothesis 6a: A game-based assessment will have a stronger direct positive effect on perceived PO fit, both complementary and supplementary, than will a traditional assessment.

Hypothesis 6b: Flow will partially mediate the relationship between a game-based assessment and perceived PO fit, both complementary and supplementary. No partial mediation is expected for a traditional assessment.

I have discussed perceived PO fit, and explained how GBAs and flow can signal to an applicant that he or she would fit at the organization. I now connect perceived PO fit to organizational attractiveness to clarify why an applicant may want to work at a place with which he or she feels fit.

Perceived PO fit and organizational attractiveness

Much of the fit literature shows that perceptions of fit influence job seeking attitudes and intentions (Cable & Judge, 1996; Cable & DeRue, 2002; Carless, 2005; Chatman, 1991; Lauver & Kristof-Brown, 2001; O'Reilly et al., 1991; Uggerslev, Fassina, & Kraichy, 2012). One particular attitude of interest is organizational attractiveness, or "perceptions about the appeal or image that a company or organization maintains" (Hausknecht et al., 2004, p. 643). Two theories can be used to explain the link between PO fit and organizational attractiveness: Schneider's attraction-selection-attrition (ASA) model and Byrne's similarity-attraction paradigm (Byrne, 1971; Schneider, 1987). The ASA model submits that individuals will be attracted to and opt to work in organizations that share characteristics with the individual. Along the same lines, the similarity-attraction paradigm says individuals are attracted to and try to make connections with

entities that are similar to them. A host of research supports the link between fit and attractiveness (Bretz, Ash, & Dreher, 1989; Cable & Judge, 1997; Judge & Bretz, 1992). In order to accommodate both supplementary and complementary fit, one may say that individuals are attracted to organizations that embody the individual's current values or idealized values. I then hypothesize that:

Hypothesis 7: Perceived PO fit (both complementary and supplementary) will be positively related to organizational attractiveness.

Moderating role of selection decision in the perceived PO fit to organizational attractiveness relationship

Earlier I reviewed work on the effect of a selection decision on fairness; the synopsis is that rejected individuals have lower fairness perceptions of the process and outcome. I expect the same to be true for organizational attractiveness; applicants who are not offered the job will be less attracted. However, the moderating effect will show that perceived PO fit is a more dominant cue of organizational attractiveness when applicants are given an offer. The reason for this is fit becomes even more relevant at this point. Let us examine the opposite scenario where an applicant feels some amount of fit but then is rejected from the job. Attractiveness is going to be lowered, regardless of fit, because joining the organization is not an option at this time. Therefore, the time that fit perceptions matter (in the case that a selection decision has been provided) is when the offer is extended.

Hypothesis 8: The relationship between perceived PO fit (both complementary and supplementary) and organizational attractiveness is moderated by the selection decision such that the relationship is stronger when the individual is offered a job and weaker when the individual is not offered a job.

I now move to the final set of hypotheses which discuss how perceived justice and organizational attractiveness perceptions can influence a candidate to accept an offer from an organization.

Acceptance intentions

Acceptance intentions are frequently used as an outcome in the study of applicant reactions to understand whether the applicant wants to continue the relationship with the employer that started during the recruitment process (Herriot, 1989). Acceptance intentions portray how favorably the applicant views the selection process, job, and organization (Macan et al., 1994; Rynes, 1993; Rynes & Barber, 1990). Intentions measures also have practical implications (Ryan & Ployhart, 2000), as they are one of the key indicators of whether an applicant would eventually accept the offer (which is the outcome in which organizations are most invested). Because this is a laboratory-based study, I use them here as the closest indicator to acceptance behavior under a simulated scenario.

Justice perceptions and acceptance intentions

Justice is often cited as a predictor of relevant work intentions and of later behaviors (Gilliland, 1993; Hausknecht et al., 2004; Smither et al., 1993). To reiterate a previous concept, the premise of justice theory within applicant reactions is that when applicants feel the selection process (procedural) and the outcome of the selection process (distributive) are fair, applicants are more likely to engage with and even accept offers from an organization. Let us examine this by each type of justice individually. The origins of procedural justice are in dispute resolution and the original authors referred to present-day procedural justice or voice as *process control* (Thibaut & Walker, 1975). They found that disputants were willing to give up distributive control (at the time called decision control) in order to have greater process control. Gilliland

reinforces this idea by saying procedural justice will be more strongly related to acceptance decisions than distributive justice (1993).

Other research has framed this same idea by saying procedural justice has a greater effect on evaluations of others, and distributive justice is more important in evaluating personal outcomes (McFarlin & Sweeney, 1992). Models about intentions reinforce the postulations of justice theory. For example, the theory of planned behavior advances the idea that intentions are a function of beliefs, attitudes, and subjective norms (Ajzen, 1991). Perceived behavioral control is a predictor of intentions in this theory: when a person feels he or she is given an opportunity to perform, there are then stronger intentions. This conjecturing dovetails neatly with the tenants of procedural justice (Ryan & Ployhart, 2000). Both original conceptualizations of justice and the theory of planned behavior suggest that people prefer to be in control of the process, and that control can lead to positive outcomes like acceptance intentions.

While procedural justice seems to be of greater importance, fair outcomes are important too. Other research in the justice arena shows that distributive justice matters greatly for pay outcomes (Folger & Konovsky, 1989; Sweeney, 1990). Distributive justice elicits equity assessments, and perceived equity results in satisfaction and positive emotions (Colquitt et al., 2001; Greenberg, 1990). When these positive equity assessments are attributed to the focal organization, one may expect that:

Hypothesis 9: Procedural and distributive justice will be positively related with acceptance intentions.

Organizational attractiveness and acceptance intentions

The connection between organizational attractiveness and acceptance intentions is fairly intuitive. Organizational attractiveness, or the appeal of the company to the applicant, relates to

how favorably the applicant views the organization. When applicants like or respect a company, they are more likely to want to enter that system. This link has been shown empirically (Hausknecht et al., 2004; Macan et al., 1994; Rynes, 1993; Rynes & Barber, 1990). One can also rely on real world examples: many of the USA's most highly rated places to work (such as American Express, Acuity, and SAS Institute) are also the hardest places to get hired (Fortune, 2016; Shen, 2016). When attractive organizations issue offers, they are accepted.

Hypothesis 10: Organizational attractiveness will be positively related to acceptance intentions.

METHODS

Rationale for design

My study manipulated assessment method (rather than assessment construct) (Arthur & Villado, 2008). Therefore, my focus was on controlling or measuring any confounding effects in order to get as pure an assessment of the two methods as possible. I took several steps to reach this goal: I held the constructs as constant as I could across the two methods. I piloted my assessments to understand the extent to which participants notice the core characteristics (game elements) in the GBA versus the non-GBA. Lastly, in my main study I assessed a number of other variables so I could understand the role of other non-method related differences.

By virtue of looking at assessment method, I was interested in external validity. I simulated the use of a GBA and a non-GBA in a realistic setting as much as possible. My GBA was quite ecologically valid in that it has been frequently tested and is used by companies already. It has the core characteristics one would expect from an operational GBA: game elements, measurement of work-related constructs, and reasonable psychometric support. I imbued the non-GBA measure with similarly realistic characteristics. To do so I used item types that are often seen in company settings (i.e. Likert and cognitive ability items), I administered the assessment through a computer, and I used measures with content that is appropriate in a work setting.

Pilot study

Prior to conducting the experiment, I conducted a pilot study to establish similarity between the GBA and non-GBA. My primary interests were convergent validity between the constructs measured, amount of time to complete assessment, and reaction to method of delivery.

Sample

To determine the sample size, I used G*Power to conduct a power analysis for correlations (Faul, Erdfelder, Lang, & Buchner, 2007). I used an exact, a priori, two-tailed scenario with a desired correlation of .30, alpha of .05, power of .80, and null hypothesis of 0. My recommended sample size was 84.

I collected a Mechanical Turk sample of 101 participants which was a 62% response rate from the 163 who were given a qualification to proceed to the pilot from the pre-screen. This sample was primarily female (62.4%) and White (73.3%), with some African-American (10.9%), and Asian (7.9%) participants; other participants marked themselves as 'other' or chose not to answer. The average age was 36.25 years of age (*SD* = 9.06).

This sample had on average 16.55 years of work experience (*SD* = 8.73) in a range of industries (please see Table 7). Of those who responded to the question, some had no experience with the hiring process (17.8%), some had it gone through it up to five times (47.5%), and others had been through it six or more times (34.6%). About half the participants had no exposure to GBAs during these hiring experiences (45.5%), some had been exposed one to three times (39.6%), and others had interacted with GBAs several times (14.8%).

Table 7. *Industries of Mechanical Turk pilot participants*

Industry	***N***	**Percentage**
Arts, Design, Entertainment, Sports, and Media	5	5.0%
Business and Financial Operations	10	9.9%
Community and Social Service	1	1.0%
Computer and Mathematical	9	8.9%
Construction and Extraction	2	2.0%
Education, Training, and Library	8	7.9%
Food Preparation and Serving Related	6	5.9%
Healthcare Practitioners and Technical	2	2.0%
Healthcare Support	11	10.9%
Legal	1	1.0%
Life, Physical, and Social Science	2	2.0%

Table 7 *(cont'd)*

Industry	***N***	**Percentage**
Management	8	7.9%
Office and Administrative Support	14	13.9%
Personal Care and Service	1	1.0%
Production	2	2.0%
Protective Service	1	1.0%
Sales and Related	10	9.9%
Retired	2	2.0%
Other (please specify)	5	5.0%
Total	*101*	*99.2% (due to rounding)*

Measures

Game-based assessment. For my study I used a GBA by the company Arctic Shores, a UK-based company that specializes in GBAs that are used for both selection and development. The GBA is called Cosmic Cadet and is a five-module assessment designed to measure executive functioning, attention control, processing capacity, processing speed, performance under pressure, learning agility, resilience, persistence, social confidence, affiliative, risk appetite, managing uncertainty, and innovation potential. The assessment typically takes between 25 and 30 minutes.

There is some psychometric support for the assessment based on a diverse sample (paid testers, students, air traffic controllers, other full-time workers). Based on the information in Artic Shores' technical report, the alpha for each construct ranges from .50 to .91, with an average of .65 (Artic Shores, personal communication, November 27, 2016). The test-retest of the assessment is high (.85). Construct validity between Cosmic Cadet and several other measures has been examined. The table below provides a summary of these relationships.

Table 8. *Convergent validity coefficients for Cosmic Cadet constructs*

GBA Construct	**Corresponding Measure and r**
Executive functioning	Cut-e's Scales FX .22* Cut-e LCT .22* Cognitive Monitoring Scale .30* Saville Assessment .41 (average)
Processing capacity	Cut-e's Scales FX .32* Cut-e LCT.42* META's Disruptive Talent -.32 (average) Cognitive Monitoring Scale .31* Saville Assessment .39***
Processing speed	Cut-e's Scales FX .20* Cut-e LCT .30*
Performance under pressure	Watson-Glaser .30**
Learning agility	Saville Assessment .37***
Resilience	The Brief Resilience Scale .30** Work Acceptance & Action Questionnaire .30** META's Disruptive Talent Resilience .32*
Persistence	International Personality Item Pool: Persistence .41**
Social confidence	Social Confidence (Cut-e's Shapes) .52**
Affiliative	Sociability (Cut-e's Shapes) .38**
Risk appetite	Bomb Risk Elicitation Task .50** Cut-e's Scales FX .22* Cut-e LCT .21* Cognitive Monitoring Scale .38 (average)
Managing uncertainty	Multiple Stimulus Types Ambiguity Tolerance Scale-II .30* Cognitive Monitoring Scale .34 (average)
Innovation potential	Creative Personality Scale .20**

** $p < .01$, * $p < .05$

There is also evidence of criterion-related validity for the cognitive portions of the assessment. In an evaluation with 181 air traffic controllers, processing capacity, learning agility, and executive functioning had correlation coefficients ranging from .39 to .47 (uncorrected) with a composite of performance.

Non-game-based assessment. In this section I outline the measures I used in my non-GBA assessment for the pilot (and my main study). Items and examples can be found in Appendix A.

Cognitive measures. In order to assess executive function, processing speed, processing capacity, learning agility, attention control, and performance under pressure I used items from the International Cognitive Ability Resource (ICAR, 2014) and the digit span test. The ICAR is a public domain and open source bank of cognitive ability items that is available to research groups. I used eight 3D rotation items, eight matrix reasoning items, and eight letter and number series. These items have been administered in very large samples (~97,000 participants) and data from subsets of these samples have shown reliabilities ranging from .68 to .93 for item types, discriminant validity among item type, and convergent validity with standardized tests and cognitive ability measures such as the Shipley-2 (Condon & Revelle, 2014; Shipley, Gruber, Martin, & Klein, 2009). In Table 9, I match the item types to the construct they are measuring and identify my scoring method.

Table 9. *Cognitive ability item types/tasks, definition, scoring method, and rationale*

GBA Construct	**Definition**	**Matched Item Type**	**Component**	**Rationale**
Executive functioning	Control processes responsible for planning, assembling, coordinating, sequencing, and monitoring other cognitive operations (Salthouse, Atkinson, & Berish, 2003, p. 566)	Letter and number series 3D rotation	Score	These items require sequencing, assembling, and coordinating. In previous construct validations, spatial reasoning tasks like the 3D rotation have had a moderate correlation with executive functioning (Artic Shores, personal communication, November 27, 2016).

Table 9 *(cont'd)*

GBA Construct	Definition	Matched Item Type	Component	Rationale
Processing speed	Ability to automatically and fluently perform relatively easy or over-learned elementary cognitive tasks, especially when high mental efficiency (i.e., attention and focused concentration) is required (McGrew, 2009, p. 6)	Letter and number series Matrix reasoning	Score Time	The narrow abilities of processing speed include number facility and speed of reasoning, so this scoring system takes into account both the time and accuracy of completing number and reasoning tasks.
Processing capacity	Ability to apprehend and maintain awareness of elements of information in the immediate situation (Shrank, 2010)	Digit span	Score	This test is a measure of working memory which relates to one's ability to maintain awareness of a number of elements in the immediate situation (McGrew, 2009; Wechsler, 1939).
Learning agility	Desire and ability to learn from experience and apply this learning to novel situations (Artic Shores, personal communication, November 27, 2016)	Matrix reasoning	Score	These items require pattern recognition, which is similar to the ability to identify previous learnings and apply them in a novel situation. In previous construct validations of Cosmic Cadet, pattern recognition items have had a moderate correlation with learning agility (Artic Shores, personal communication, November 27, 2016).

Table 9 *(cont'd)*

GBA Construct	Definition	Matched Item Type	Component	Rationale
Attention control	The ability to adapt one's attention from moment to moment (Artic Shores, personal communication, November 27, 2016)	Digit span	Score	Attention control is a component of working memory (Engle & Kane, 2004), so a working memory task like the digit span will capture important variance in attention control.
Performanc e under pressure	Ability to perform under difficult or constrained circumstances	3D rotation Digit span Letter and number series Matrix reasoning	Score	Because these items will be timed, the overall score under time constraints will represent ability to perform under pressure.

The ICAR items are normally administered without time restraints, but I imposed time constraints as some constructs cannot be appropriately assessed without it. Certain cognitive abilities, like spatial abilities, have evolved in interaction with environmental demands such as time pressure. It is considered more ecologically valid to administer tests of these abilities under some form of time pressure (Peters, 2005). Setting time limits is in keeping with other cognitive measures like the military's ASVAB and college admissions' SAT or GRE (Mead & Drasgow, 1993). The original creators of the ICAR do not preclude the use of time restraints but rather did not use them in validations of the items so as to get a more stringent evaluation of the items' utility (Condon & Revelle, 2014). It was therefore appropriate to include some time constraints for my study.

These restraints were set generously for the three sets of ICAR items, as my intention was not to rush the participants but rather to add some time urgency. These time limits were determined both through the pilot testing and through norms used by other cognitive tests. For

the matrix reasoning items the pilot group had a mean of 34.36 seconds (*SD* = 18.67) per item. For Raven's Advanced Progressive Matrices (Raven, Raven, & Court, 1998) the authors allot between 25 and 67 seconds per item depending on difficulty. I used similar norms and allotted eight minutes for eight items of moderate difficulty (60 seconds an item).

For the 3D rotation tasks, the pilot group had a mean of 30.90 seconds (*SD* =19.96). This is shorter than the Vandenberg and Kuse (1978) mental rotation which often allows 15 to 30 seconds per item (Peters, 2005; Voyer, 2011). Because of the high number of options and the added cognitive load of adding options for "None of these", I allotted 60 seconds per item here as well.

For the letter and number series, the pilot group had a mean of 31.59 seconds per item (*SD* = 17.76). In similar tasks such as the Primary Mental Abilities number series (Thurstone & Thurstone, 1962) and the Adult Development and Enrichment Project letter series (Blieszner, Willis, & Baltes, 1981) participants are allowed 13.5 seconds per item. I allotted 30 seconds an item for the eight items as the items in the ICAR include a "None of these" option which is more cognitively taxing than choosing from bounded multiple choice answers.

To assess processing capacity I used a digit span test modelled after the Wechsler Digit Span Test (1939). In the original task the participant is auditorily presented with one number each second, in increasing lengths until he or she fails to repeat the string correctly twice. I administered visual strings at the same interval, starting with two digits and advancing up to nine digits, which is the common maximum capacity cited for working memory (Engle, 2002).

Resilience. I used the Connor-Davidson Resilience Scale which is a longer measure that discusses factors that contribute to one's ability to be resilient (e.g., "know where to turn for help") (Connor & Davidson, 2003). The Cronbach's alpha for this measure in the original paper

was .89. It was .96 for my pilot sample and .92 for my main study sample. The scale had high convergent validity with hardiness and stress scales. It had divergent validity with a sexual experiences scale. One item was excluded from this scale because it is inappropriate for work-related measures ("Sometimes fate or God can help").

Persistence. I used an 8-item measure from the IPIP on industry/perseverance/persistence. This is a portion of the Values in Action scale (Goldberg et al., 2006; Peterson & Seligman, 2004). This scale has an alpha of .81 in previous studies, .89 for both my pilot sample and my main study sample.

Social confidence. I used the 10 social confidence items derived from the Jackson Personality Inventory available in the IPIP (Jackson, 2004; Goldberg et al., 2006). These items have an alpha of .87 in previous studies, .96 in my pilot sample, and .90 in my main study sample.

Risk appetite. To assess risk appetite, I used the Bomb Risk Elicitation Task, or BRET (Crosetto & Filippin, 2012). In this task, participants are given a 10 X 10 figure with 100 boxes in it, and told that each opened box nets a point. However, one of the boxes has a bomb in it. The box with the bomb is randomly determined after they have chosen how many boxes to open. Starting from the upper left corner, participants have to choose how many of the 100 boxes to open; the closer a participant gets to opening all 100 boxes, the higher risk appetite he or she is said to have. This game is a more true measure of risk appetite because it does not capture loss aversion, which has been shown to have gender differences (Charness & Gneezy, 2010; 2011). BRET explained significance variance in other measures of risk. The authors also conducted the task in a variety of ways (increased reward, larger matrix of boxes) but sample means were not significantly different. The GBA risk appetite module is based off the BRET, and the two have

been strongly correlated in previous studies. In this study, there was a mean of 37.08 (*SD* = 22.19) indicating that the sample was on the more risk-adverse side.

Managing uncertainty. For managing uncertainty I used an unpublished measure- the General Risk scale (Nye, personal communication). This scale assesses non-domain specific preference for risk, and is 14 items long. This scale had an alpha of .96 in my pilot study and .95 with my main study sample.

Innovation potential. For this construct I used the 12-item Propensity to Innovate scale from Burningham and West (1995) which measures an individual's disposition toward change at work. This scale has reliabilities ranging from .72 to .85 in other studies that have used it (Bunce & West, 1995; 1996), .89 in my pilot study, and .84 in my main study sample.

The measures listed in this section were retained for the main study because they had significant correlations with the GBA and thus demonstrated reasonable convergent validity. These results are further explicated in the pilot analysis section.

Non-game-based assessment measures that were excluded from the main study. Some measures were included in the pilot to see if they had convergent validity with the GBA constructs, but then were excluded in the main portion of the study due to low and nonsignificant correlations with the GBA constructs. These measures can be found in Appendix B, with descriptions below.

Cognitive ability items. There were seven mini-tests included in the pilot from a vendor. These tests assessed data ordering, focus, ability to read and interpret graphs, inductive reasoning, pattern finding, following instructions, and logical sequencing. However, the vendor did not correctly collect IDs so while the participants did complete these assessments, their data

were not able to be matched to the other survey measures. These assessments were not included in any other parts of the study, and the results are not reported.

Brief Resilience Scale. In order to assess resilience I used Smith and colleague's Brief Resilience Scale which measures the ability to bounce back after stressful experiences (Smith et al., 2008). This measure was designed to be a more "pure" measure of resilience in that it only measures resilient behavior as opposed to also measuring the availability of resources that make resilience possible. This measure has good construct validity with high convergent validity with other resilience and coping measures, and high reliability (.80-.91 across four samples, and .94 for my pilot sample). This scale had some convergent validity with the GBA resilience construct in previous samples, but correlated nonsignificantly .01 with GBA resilience in my pilot study.

IPIP Sociability. In my pilot I used the 10 sociability items derived from the HEXACO Personality Inventory available in the IPIP (Ashton, Lee, & Goldberg, 2007; Goldberg et al., 2006). These items had an alpha of .96 in my study, and a nonsignificant correlation of .04 with the GBA construct of social confidence.

Affiliative Tendency Scale. I used the 25 items from Mehrabian (2000) in the pilot to capture one's preference for people. This scale had an alpha of .90 in the pilot, but a nonsignificant -.02 correlation with the GBA affiliation construct.

Multiple Stimulus Types Ambiguity Tolerance Scale-II. For managing uncertainty, I used the same scale that was utilized in previous construct validations of Cosmic Cadet- the Multiple Stimulus Types Ambiguity Tolerance Scale-II (McLain, 2009). This scale is a more recent iteration of other measure of ambiguity, and improves on these measures with higher reliability and stronger alignment to the underlying theories. The average reliability across three samples in the original manuscript is .81. In previous studies this scale has had reasonable convergent

validity with another measure of ambiguity tolerance, and discriminant validity with a social desirability scale. These items had an alpha of .96 in my study, but a nonsignificant correlation of .09 with the GBA construct of managing uncertainty.

Creative Personality Scale. For innovation potential I included the Creative Personality Scale (Gough, 1979). This scale derives from the Adjective Check List (Gough & Heilbrun, 1965) and was developed with the help of many kinds of creativity outcomes (faculty evaluations, personality assessment staff observers, life history interviewers, and subject matter experts). This scale has high convergent validity with related scales, and a reliability of .76 across two samples. In it participants are given 30 adjectives and asked to check those that apply to them. They are then given 1 point for each of the positive items checked, and lose one point for each of the negative items checked. This scale was used in the original construct validation of Cosmic Cadet, but had a low correlation. In my pilot it had a nonsignificant correlation of -.02 with innovation potential.

Reactions and manipulation check measures. In order to understand how participants perceived and reacted to the two assessments, I used the following measures in the pilot.

Attitudes toward mobile testing. I used the seven-item scale from King and colleagues (2015) to understand the participants' perceptions toward the use of mobile devices to complete assessments. The scale is composed of one multiple-choice item that asks about preferences for mobile over computerized assessments, and six items that use a Strongly Disagree to Strongly Agree scale to assess fairness, organizational attractiveness, and application and acceptance intentions for the two devices. The alpha was .79 in the original study and .73 in my pilot.

Awareness of game elements. While my study manipulates and is principally interested in assessment method, I still recognize that game elements are a primary characteristic that

distinguishes GBAs from non-GBAs. Therefore I collected some information to see whether participants noticed the presence of game elements. I collected information about the three common game elements I referred to in the introduction (focused goals/rules, feedback, and challenging tasks) as well as the element of sensory stimuli to see if the participants notice the game-like feel of the assessment. I included an 8-item measure that I created to evaluate participants' awareness of engaging with a GBA ($\alpha = .88$ for the GBA, $\alpha = .71$ for the non-GBA).

Demographics**.** I collected age, race, gender, years of work experience, industry, number of times participating in the hiring process, and experience with GBAs from my pilot sample.

The pilot study measures can be found in Appendix B.

Procedure

Prior to beginning my pilot and main study, I conducted a pre-screen to identify people who worked full-time, had a smartphone on which the GBA app could be downloaded, and would be comfortable downloading an app on their phone to complete a HIT (a Human Intelligence Task, or what MTurk calls its self-contained tasks). I also asked about gender and commute time to work as distractors, so that participants would not be made aware of the demands of the study. These participants were paid $.01 for about a 30-second prescreen survey. A total of 1,206 participants responded of whom 637 were qualified for the pilot or main study. Of these, 163 were given access to the pilot (of which 101 successfully completed). Most of the other qualified participants (458) were reserved and given access to the main portion of the study, which is discussed in more detail in the main study section.

Mechanical Turk participants completed the pilot virtually. They were first asked to complete a consent form. They were then asked to respond to the attitudes toward mobile testing

scale before completing either of the assessments so that their responses would not be affected by the test-taking experience. The purpose of this measure was to understand whether preferences in assessment device relate to test performance.

Participants took both the GBA and the non-GBA, and the order of presentation was randomized across participants. The GBA was taken on a smartphone and the non-GBA was taken on a computer. After each assessment, participants were asked about their awareness that game elements were present ($\alpha = .88$ for GBA manipulation check measure and $\alpha = .71$ for non-GBA manipulation check measure). At the end participants were asked to provide feedback about significant differences between the two methods of assessments, as well as any other general comments. Participants were also asked if they were genuinely paying attention during the study, to which 100% answered yes.

Analysis

The order in which the participants took the GBA and non-GBA was correlated with the assessment scores and the awareness of game elements composites to see if there were order effects. There were no significant correlations.

Attitude toward mobile testing results. I first looked at the attitudes toward mobile testing measure to understand how my pilot sample felt about assessment devices. For the first item which asked participants how they would compare taking a test on a mobile device and a computer, more than half (63.4%) agreed that completing an assessment on a mobile device was worse than completing it on a computer. About a quarter (26.7%) did not have a preference, and 8.9% preferred a mobile test. As shown below in Table 10, the other six items in the scale indicated that many participants felt it was equally fair to administer an assessment on a mobile device as opposed to a computer, and felt that offering mobile assessments was good for the

company's brand and reputation. These results indicated that that there was a proclivity toward taking computer-based assessments but that mobile assessments overall also have a positive perception , so in interpreting my results I needed to consider that there was a slight preference for computerized assessments.

Table 10. *Attitude toward mobile testing measure means and standard deviations*

Item	***M***	***SD***
I would prefer to complete tests on a mobile device versus completing them on a computer.	2.26	1.14
I would be more likely to apply for a job at a company that allowed me to complete a pre-employment test on my mobile device such as a smart phone, tablet, or iPad versus a company that allowed taking the same test only on a computer.	2.83	1.23
It is equally fair to use a test given on a mobile device as it is to use the same one given on a computer to make a hiring or promotion decision for a job.	3.56	1.09
I believe a company that allows me to take its test on my mobile device would be a better place to work compared to a company that only allows its test to be taken on a computer.	2.88	1.04
I would be more likely to accept a job offer from a company that allows me to take its test on my mobile device versus a company that only allows its test to be taken on a computer.	2.76	1.09
Having the option to complete this assessment on a mobile device positively represents the company's brand image.	3.47	1.06

Note: Items are on a 1 (Strongly disagree) to 5 (Strongly agree) scale

I then correlated the attitude toward mobile testing scale with performance on the assessment constructs to understand whether having a preference for mobile assessments was related to my measures. I did find positive correlations with resilience, social confidence, general risk, innovation, and creativity measures on the non-GBA, and with affiliation and social confidence on the GBA. These significant correlations were moderate in nature, ranging from .21 to .34 as seen in Table 11.

Table 11. *Correlations between items 2-7 on the attitudes toward mobile testing scale and assessment scale/construct*

	Attitude toward mobile testing measure composite (Items 2-7)
Non-game based assessment	
3D Rotation	.00
Digit Span	.00
Matrix Reasoning	.04
Letter and Number Series	.14
Brief Resilience	.27**
Connor Davis Resilience	.34**
IPIP Persistence	.23*
IPIP Social Confidence	.31**
IPIP Sociability	.21*
Affiliative Tendency Scale	.14
Multiple Stimulus Types Ambiguity Tolerance II	.23*
General Risk	.33**
Bomb Risk Elicitation Task	-.02
Propensity to Innovate	.28**
Creative Personality Scale	.23*
Game-based assessment	
Affiliation composite	.25*
Attention Control composite	.03
Executive Function composite	-.01
Innovation Potential composite	.18
Learning Agility composite	-.04
Managing Uncertainty composite	.20†
Performance Under Pressure composite	.05
Persistence composite	.12
Processing Capacity composite	-.05
Processing Speed composite	-.01
Resilience composite	.03
Risk Propensity composite	.13
Social Confidence composite	.28**

** $p < .01$, * $p < .05$, † $p < .07$

Manipulation checks. Next, I was interested to see whether there was significant differences between the perceived game elements in the GBA and the non-GBA. To evaluate this, I created a composite item for each of the four game elements (focused goals/rules, feedback, challenging tasks, and sensory stimuli). There were significant differences for three out of the four elements. Participants felt there were more focused goals/rules in the GBA condition ($M = 4.57$, $SD = .82$) than the non-GBA condition ($M = 4.40$, $SD = .80$); $t(100) = 2.12$, $p < .036$. Participants felt there was more feedback in the GBA condition ($M = 4.08$, $SD = 1.07$) than the non-GBA condition ($M = 1.96$, $SD = 1.14$); $t(100) = 13.76$, $p < .001$. The same was true for sensory stimulation or feeling like the assessment was game-like ($M = 4.61$, $SD = .87$ for the GBA and $M = 2.70$, $SD = 1.34$ for the non-GBA); $t(100) = 13.06$, $p < .000$. There was not a significant difference between the GBA ($M = 4.26$, $SD = .90$) and non-GBA ($M = 4.24$, $SD = .87$) conditions for challenging tasks.

This indicated that there were significant perceived differences between the two assessments due to the presence of game elements as expected.

Timing. For the next analysis I looked to see if the times between the two assessments were similar. The GBA took an average of 17.90 minutes to complete ($SD = 9.02$). The non-GBA took a total of 29.28 minutes ($SD = 12.03$) with an average of 17.92 minutes on the cognitive portion of the assessment and 11.37 on the non-cognitive measures. The timings for the non-GBA take into account several added measures that were included for piloting and removed for the main study. With the removal of five measures, and the addition of only one (Extraversion) in the main study, these times were considered sufficiently equivalent.

Convergent validity. Lastly, I looked to see whether my GBA and non-GBA were measuring equivalent constructs by correlating them, as can be seen in Tables 12-14.

The first table examines the relationship between the non-GBA cognitive measures and the GBA cognitive composites. We see a moderate relationship with all of the non-GBA measures except the digit span and the GBA composite of processing capacity ($r = .28$, $p < .01$ for 3D rotation, $r = .26$, $p < .05$ for letter and number series, $r = .29$, $p < .01$ for matrix reasoning). There was an inverse relationship between the letter and number series, and GBA attention control meaning as a person's score on the letter and number series went up, his or her attention control score went down ($r = -.24$, $p < .05$). This was counterintuitive. There was also a marginally significant relationship between the letter and number series and GBA performance under pressure ($r = .19$, $p < .07$). Overall there was weak convergent validity here.

Table 12. *Cognitive measure score composites and construct correlations*

GBA composites	**3D Rotation**	**Digit Span**	**Letter and Number Series**	**Matrix Reasoning**
Attention Control	.01	-.13	-.24*	-.12
Executive Function	.19	-.08	-.01	.10
Learning Agility	.14	.11	.08	.18
Performance Under Pressure	.10	.08	.19†	.14
Processing Capacity	.28**	.04	.26*	.29**
Processing Speed	.14	-.15	.02	.06

** $p < .01$, * $p < .05$, † $p < .07$

The second table reflects the non-GBA cognitive measure timings and the GBA composite scores. Participants were told to complete the non-GBA tasks as quickly and accurately as possible while also having moderate time pressure, and almost all of the GBA components were timed. I had predicted that some constructs could not be adequately measured without a reference to time, and therefore amount of time should relate to scores on the task. There was one relationship that turned out as expected, and that was that those who took longer

on the non-GBA matrix reasoning section had on average lower GBA processing speed scores (r = -.21, p < .05). Thus there was partial support for the idea that time on a task can relate to GBA scores.

Table 13. *Cognitive measure times and construct correlations*

GBA composites	3D Rotation Time	Letter and Number Series Time	Matrix Reasoning Time
Attention Control	.01	.19	.14
Executive Function	.08	.07	.04
Learning Agility	-.02	-.16	-.18
Performance Under Pressure	.00	.09	-.11
Processing Capacity	.11	-.11	-.10
Processing Speed	-.17	-.11	-.21*

* p < .05

Lastly, I looked at the relationship between the non-GBA personality measures and the GBA personality composites. Several constructs were validated as expected: managing uncertainty, persistence, and risk propensity. A few other constructs were validated through scales that do share theoretical underpinnings. GBA innovation potential was moderately related to the BRET (r = .42, p < .01), and one could argue that those who are not afraid to take risks would also be more innovative, as that requires putting one's novel ideas out for display and critique. GBA resilience was positively related to persistence (r = .23, p < .05), logical in that those who get back up after they have been knocked down are also persistent. GBA social confidence shared positive correlations with the BRET (r = .25, p < .05) and General Risk Scale (r = .28, p < .01), with the reasoning that those who are less risk adverse will also be bold in their social decisions.

Noticeably, the BRET and the General Risk scale correlated with the GBA dimensions more than I hypothesized. The BRET (correlated with four additional GBA dimensions) is unlike

Likert-style personality measures and is arguably more “game-like”. The General Risk scale was correlated with three additional measures unexpectedly. The GBA, and games in general, assess risk throughout; the player makes decisions of speed vs. accuracy. Therefore there may be a method effect or general factor measured outside of just the constructs.

Five scales did not yield any significant correlation: Affiliative Tendency, Brief Resilience, Creative Personality, IPIP Sociability, IPIP Social Confidence, and Multiple Stimulus Types Ambiguity Tolerance II. All of these other than IPIP Social Confidence were removed for the main study. The social confidence scale was retained because there were no other validated scales to replace the affiliative tendencies, and I wanted to collect a measure of this to be comprehensive in the constructs assessed by the non-GBA. While some relationships did not pan out as expected, overall there was moderate support for convergent validity between the GBA and non-GBA personality measures and constructs.

Table 14. *Personality measure composites and construct correlations*

GBA Composite	Affiliative Tendency Scale	Bomb Risk Elicitation Task	Brief Resilience	Connor Davis Resilience	Creative Personality Scale	General Risk	IPIP Persistence	IPIP Sociability	IPIP Social Confidence	Multiple Stimulus Types Ambiguity Tolerance II	Propensity to Innovate
Affiliation	-.02	-.02	.04	.20†	-.13	.11	.10	.06	.08	-.03	.17
Innovation Potential	.09	.42**	.07	.12	-.02	.17	.12	-.03	.03	.01	.10
Managing Uncertainty	.12	.40**	.14	.16	.05	.31**	.08	.05	.11	.09	.16
Persistence	.05	-.18	.11	.21*	.05	-.29**	.31**	.04	.06	-.08	.06
Resilience	.08	-.38**	.01	.10	.01	-.10	.23*	.06	.07	-.11	.01
Risk Propensity	.02	.44**	-.02	.07	-.14	.26*	.06	-.04	.00	-.07	.11
Social Confidence	.03	.25*	.09	.23*	-.09	.28**	.14	.04	.10	-.01	.23*

** $p < .01$, * $p < .05$, † $p < .07$

Main study

Sample

I conducted the main experiment using two samples: an undergraduate sample from the university subject pool and an adult, working sample from Amazon's Mechanical Turk. In order to determine the sample size needed, I used 3 different methods of power analysis. The first was to ensure I had sufficient power for my ANOVA analysis. This was calculated in G*Power (Faul, Erdfelder, Lang, & Buchner, 2007) using a desired medium effect size of .25, an alpha error probability of .05, a power of .95, a numerator degrees of freedom of 1, and 4 desired groups (because I have a 2 by 2 design). The recommended sample size was 210. My second method was an a-priori structural equations modeling sample size calculator (http://www.danielsoper.com/statcalc3/calc.aspx?id=89). I used the same effect size, probability level, and power, and input the two latent variables and eight observed variables of my model. The sample size recommended here was 225. Lastly, a few loose rules of thumb have suggested that ten participants for each variable or sample sizes over 200 are generally advisable (Tanaka, 1987). I restricted participants in both samples to people who have smartphones, as this was needed for those in the GBA condition.

The first sample was drawn from the participant pool at Michigan State University. I provided credit for students who participate in an hour long study. While some researchers have bemoaned the overuse of undergraduates in research studies (Anderson, 2003; Sears, 1986), undergraduate students are actually the target audience of many GBA endeavors. Organizations wish to appear novel and cutting-edge so that they can attract young, entry-level talent (Meister, 2013). Therefore, I was very interested in the perceptions of this soon-to-be entry-level workforce. Additionally, because part of my study design requires participation on multiple days,

I used an undergraduate sample because I was more effectively able to reach them on their MSU emails the day after they take the assessment.

A total of 202 students signed up to take the study, of whom 161 earned credit for completing it (79.7%). Eleven more people were removed for failing attention and/or timing checks. Of the remaining 150, 22 were unusable because of an error in the game-based assessment that resulted in no data being collected. My final student sample was 128, and it was primarily female (69.5%), with a mean age of 20.53 (*SD* = 2.81). This sample was mostly White (62.5%), as well as African American (8.6%), Asian (24.2%), and American Indian or Alaskan Native (0.8%). In addition, 5.5% of the sample identified as being Hispanic or Latino.

About half of the participants had no previous exposure to GBAs during the hiring process (43.8%), and another half had interacted with GBAs up to 5 times (46.1%) with the remaining 10.1% interacting 6 or more times.

The second sample derived from Mechanical Turk, Amazon's online marketplace for work. I restricted my pool here to workers who have high approval rates (above 98% and 5,000 approved assignments), were US-based, and who work full-time in order to get a broader sample of adults with work experience. The reason for collecting a second sample in MTurk was to collect data from a sample with work experience, get more variability in my moderator variables, and have participants from a larger range of ages and socioeconomic backgrounds (Mason & Suri, 2012). I paid these workers $8 for an hour's work; the first portion was about 45 minutes and paid $5 and the second portion was about 15 minutes and paid $3. The payment was disproportionally loaded on the second portion to encourage completion. This payment structure is higher than the recommended rate of the federal minimum wage (DeSilver, 2017) and was recognized as both incentivizing and ethical by the participants.

My final MTurk sample was 254 which is 55.4% of the 458 participants who were given the qualification to proceed to the main study from the prescreen. Of these, six people were removed due to the GBA data collection failure mentioned earlier in the student section. This resulted in a final MTurk sample of 248 people, of whom 51.6% were female, and who had an average age of 36.32 (*SD* = 9.72). This sample was mostly White (81.9%), as well as African American (6.5%), Asian (6.0%), or American Indian or Alaskan Native (0.4%). In addition, 8.5% of the sample identified as being Hispanic or Latino.

This group had an average of 17.42 years of work experience (*SD* = 9.70), and worked in a variety of industries as seen in the table below. A small portion had not participated in the hiring process before (13.7%) but about half had at least 3 times (48.0%), with another 20.2% having experienced it at least ten times, and the last 18.1% having experienced it ten or more times. About a third of participants in the MTurk sample had no exposure to GBAs during the hiring process (36.3%) with another 44.4% having experienced them at least three times. The remaining 19.3% reported having four or more experiences.

Table 15. *Industries of Mechanical Turk main study participants*

Industry	***N***	**Percentage**
Arts, Design, Entertainment, Sports, and Media	14	5.6%
Business and Financial Operations	26	10.5%
Community and Social Service	1	0.4%
Computer and Mathematical	20	8.1%
Construction and Extraction	9	3.6%
Education, Training, and Library	32	12.9%
Food Preparation and Serving Related	6	2.4%
Healthcare Practitioners and Technical	13	5.2%
Healthcare Support	16	6.5%
Legal	3	1.2%
Life, Physical, and Social Science	5	2.0%
Management	13	5.2%
Office and Administrative Support	29	11.7%
Personal Care and Service	4	1.6%
Production	5	2.0%

Table 15 *(cont'd)*

Industry	***N***	**Percentage**
Protective Service	2	0.8%
Sales and Related	22	8.9%
Other (please specify)	12	4.8%
Total	*248*	*100.0%*

Measures

The GBA from the pilot was used in the main study. Most of the non-GBA measures from the pilot were used in the main study. The exceptions were the seven mini cognitive tests, and the five personality scales mentioned in the pilot section. Additionally, an extraversion scale was added to the non-GBA assessment during the main study to achieve similar construct coverage as the GBA. To do this I used the eight items from the warmth and seven items from the gregariousness facets of extraversion in the NEO-PI R self-report version (Costa & McCrae, 1992). These were included after the sociability, social confidence, and affiliation measures drew nonsignificant correlations with the socialness measures of the GBA. These scales have had alphas between .76 and .79 in other samples (McCrae & Costa, 2005). The alpha was .92 for my main study sample.

The reactions measures for the main study can be found below. Unless otherwise noted, responses were scored on a 1 (Strongly disagree) to 5 (Strongly agree) scale because most of the original measures utilized this range (exceptions are measures from Arvey et al., 1990; Cassidy & Eachus, 2002; and Webster et al., 1993). Attention checks (e.g. "There are seven continents.") were included throughout the scales. All of the following measures can be found in Appendix C with the exception of the openness to experience and extraversion items, for copyright reasons.

Belief in testing. There are a number of beliefs that can affect reactions to a test, but it is not always clear which of these beliefs were preexisting, and which occur as a result of the test

(Chan et al., 1998b; McCarthy et al., in press). In this study I wanted to understand the effect of method-specific characteristics on outcomes rather than the effect of an individual's pre-existing proclivity toward testing. In order to understand and parse out the effect of general attitudes about the usefulness of testing for hiring, I measured belief in testing with the goal of examining whether my two experimental groups differed in their belief, and controlling for this belief if there was a significant difference. I used a 3-item measure from Arvey and colleagues to evaluate whether individuals believe in the use of employment assessments (1990). This measure had an alpha of .89 for my study.

Flow. I relied on Webster, Trevino, and Ryan's measure of flow in an HCI setting (1993). In the original paper this measure was adapted from another HCI flow study which was based on some of Csikszentmihalyi's original work (1975a). The scale in the 1993 paper had an alpha of .82. I adapted these items slightly to refer to a general assessment context. This measure had an alpha of .90 for my study.

Openness to experience. The three facets of ideas, actions, and fantasy were measured using the NEO-PI R self-report version (Costa & McCrae, 1992). Each facet has an 8-item measure with alphas from .58 to .80 for the facets used in this study (McCrae & Costa, 2010), and an overall alpha of .87 for my study.

Technology self-efficacy. To assess technology self-efficacy, I used a measure of computer self-efficacy and modified it to fit both computers and smartphones. Computer self-efficacy derives from Bandura's self-efficacy theory and an individual's efficacy judgement as it relates to multiple computer application domains (Smith, 2001). I used Cassidy and Eachus's scale which was developed to better understand how discomfort with computers affects people's use of them. The 47-item scale was initially developed using faculty and staff at a university.

This scale was construct validated and factor analyzed to reduce the number of items. This shortened 30-item scale was shown to have both construct and criterion-related validity. The reliability of the 30-items scale was .97 in the original study (Cassidy & Eachu, 2002). For this study I used an even shorter 8-item version, used by another researcher (Shank, 2011). I used the Computer User Self-Efficacy (CUSE) scale to evaluate how capable participants are at navigating computers, but I adapted the items to refer to working generally with computers instead of with computer software as the original items state because my participants are not working with software. I also adapted the items to apply to smartphone apps to fit a GBA context. I expected there to be similarities between how efficacious participants feel with smartphones versus computers, but I collected measures on both to look at reactions to the computerized tests in an exploratory manner. This measure had an alpha of .94 for my study.

Preference for games. I created a six-item measure designed to capture the extent to which a candidate has a bias toward game-based tools and processes. These items were edited by two subject matter experts, faculty members with expertise in psychometrics and gaming, as a form of validation. The measure had an alpha of .87 in this study, and the items can be found in Appendix C.

Perceived job-relatedness. To assess job-relatedness, I used Smither and colleagues' measure (1993), though I changed the word "examination" in the items to "assessment" to better align with the wording in my other measures. My definition of job-relatedness stemmed from Smither and colleagues in that it encompasses both face validity and predictive validity. In the original study the 10-item scale had an overall alpha of .82, and in this study the alpha was .91. Here the face validity sub-scale had an alpha of .89, and the predictive sub-scale had an alpha of .91.

Perceived PO fit. To measure perceived PO fit, I relied on Cable and Judge's 3-item measure (1996). This measure has high reliability ($\alpha > .80$) in previous studies (Cable & Parsons, 2001; Judge & Cable, 1997) and an alpha of .94 in this study. This measure is designed to assess supplementary fit. I altered the items slightly to ask about idealized and preferred values in order to assess complementary fit. I also edited the items slightly for clarity, and add instructions so that the participants could differentiate between current/practiced values and preferred values.

Justice perceptions. Procedural and distributive justice perceptions were assessed with five items from Smither and colleagues (1993). These items had reasonable reliability ($\alpha = .68$ for procedural, $\alpha = .86$ for distributive) and were adapted slightly to fit the selection context. In this study there was a high alpha ($\alpha = .88$ at Time 1 and $\alpha = .87$ at Time 2). One item for procedural justice was altered to read "Overall, I believe the assessment [process] was fair". This measure was assessed twice, once right after taking the assessment and again after the selection decision. The purpose for this is to see the changes, particularly in distributive justice, as a result of the selection decision as have been seen in other studies (Gilliand, 1994; Ployhart & Ryan, 1998). I chose not to use the Bauer and colleagues' Selection Procedural Justice Scale because it is based on Gilliland's ten procedural justice rules, many of which do not apply here (2001).

Organizational attractiveness. For attractiveness, I used a 5-item measure from Highhouse, Lievens, and Sinar (2003). In the original paper this measure was created using previous measures as a template, and was designed not to assess intentions but rather preliminary attitudes. This measure has been used with high reliability in other studies ($\alpha > .80$) (Jones, Willness, & Madey, 2014; Lievens, Hoye, & Schreurs, 2005). In this study there was a high alpha ($\alpha = .92$ at Time 1 and $\alpha = .93$ at Time 2). This measure was also assessed twice (right after

taking the assessment and again after the selection decision) to better understand the role the selection decision has in affecting attractiveness perceptions.

Acceptance intentions. For acceptance intentions I used the same one item as other studies that have offered a manipulated selection decision and then asked about acceptance intentions (Ployhart & Ryan, 1998; Weichmann & Ryan, 2003). Other researchers have also used general measures about intentions towards the company (intentions to reapply, intentions to stay in the selection process, intentions to recommend to a friend) (Bauer et al., 1998; Macan et al., 1994; Ployhart & Ryan, 1997). For this study I provided limited information about the company, so there was little for the participant to use to make judgments about anything other than simply accepting an offer. I controlled for other potential influences by asking the participant to disregard other factors that are normally relevant when accepting a job offer.

Awareness of game elements. The same eight-item measure used to assess focused goals/rules, feedback, challenging tasks, and sensory stimuli in the pilot was administered during the main study ($\alpha = .63$). This was driven in part by those who took the non-GBA ($\alpha = .45$), though the GBA alpha was below accepted levels too ($\alpha = .69$).

Alternative explanation checks. I included several measures to see if there are alternative explanations for differences in reactions. The measures related to perceived time to take the assessment, ease of faking, opportunity to perform, and perceived performance. These factors can sometimes affect candidates' reactions to assessments (Bauer et al., 2001; Gilliland, 1993), so I measured them here to better understand whether they vary across assessment and use them as controls if needed. The perceived time to take assessment items were created for this study and the scale had $\alpha = .55$. The ease of faking scale had $\alpha = .65$ in the original study (Gilliland & Honig, 1994) and $\alpha = .82$ in my study. Opportunity to perform has $\alpha > .80$ in other

studies (Bauer, Truxillo, Paronto, Weekley, & Campion, 2004; Maertz, Mosley, Bauer, Posthuma, & Campion, 2004) and α = .93 in my study. The perceived performance measure had α = .67 in the original study (Smither et al., 1993) and α = .90 in my study

Demographics. I collected age, race, gender, years of work experience, and experience with GBAs from both samples. From the student sample I also collected intended major and GPA. From the MTurk sample, I collected the industry in which the participant works and number of times participating in the hiring process.

Procedure

This study utilized a 2 (GBA vs. non-GBA) by 2 (selected for the job vs. not selected for the job) experimental design. Both samples participated virtually. Before taking the assessment, both samples read a brief vignette to put them in the mind frame of a job seeker at an organization who is being assessed for hiring. The organization was described as a multinational technology company that specializes in Internet-related services and products (Google, 2017). Participants were told they are applying for a sales position, as this is one of the most popular jobs in the US currently (Bureau of Labor Statistics, 2016). The reason for this particular company description and position is to denote some connection to technology, but keep the job general and accessible enough that most participants could imagine what the responsibilities would be. Participants were also told that this assessment is a screen for the company more broadly, and that later stages of the interview process will involve more job-specific hiring methods.

For undergraduate sample. Participants viewed the consent form and provided their school email addresses. They then completed the measures for belief in testing, openness to experience, preference for games, and technology self-efficacy.

Participants in the GBA condition were then asked to download the appropriate app on their smartphones, and given a passcode before playing in order to link their data to the computerized measures. After taking the GBA, they returned to the computer survey to complete the measure of flow, awareness of game elements, alternate explanation checks (perceived time to take assessment, ease of faking, opportunity to perform, and perceived performance), perceived job-relatedness, perceived PO fit, justice perceptions, and organizational attractiveness. Lastly, they completed demographics questions before being told that the organization will make a selection decision the next day based on their performance.

Participants in the non-GBA condition proceeded to take the assessment on the computer. After completing the assessment they took the same flow, awareness of game elements, alternate explanation checks (perceived time to take assessment, ease of faking, opportunity to perform, and perceived performance), perceived job-relatedness, perceived PO fit, justice perceptions, organizational attractiveness, and demographic measures and told a selection decision will be made the next day.

The next day both groups received a selection decision via their school emails. There was little or no explanation to accompany it in order to control for effects of interactional justice (Bies & Moag, 1986; Cohen-Charash & Spector, 2001). Participants were then asked again to complete the perceived justice and organizational attractiveness measures. They then were asked the appropriate acceptance intentions question based on whether they were selected or not. When making this decision, they were told to rely on the interactions they had with the organization rather than external factors such as job location or competitiveness of the job market. Finally, they were asked if they had any general comments about the study, read the debriefing form, and were thanked for their participation.

For MTurk sample. Participants who responded to the HIT were informed immediately that this was a two-part study, and they should only participate in the first part if they were willing to take a short follow-up the next day. Participants were asked to provide their MTurk ID so they could be matched to part two of the study. Participants started the survey by reading the informed consent, and completing the measures for belief in testing, openness to experience, preference for games, and technology self-efficacy. Those assigned to the GBA condition took the assessment on their phones, and those in the non-GBA continued in the survey. Both groups were then asked to complete the measures of flow, awareness of game elements, alternate explanation checks (perceived time to take assessment, ease of faking, opportunity to perform, and perceived performance), perceived job-relatedness, perceived PO fit, justice perceptions, organizational attractiveness, and demographic measures. They were then told the selection decision would come the next day.

I contacted the same workers using the messaging feature of the MTurk R program by Leeper (2017). I stated at the beginning of the survey that only workers who completed part one would be paid. Participants were then asked to enter their MTurk IDs so their data could be matched. They were given the selection decision and asked to complete the perceived justice, organizational attractiveness, and acceptance intentions measures before being asked if they had any general comments about the study and reading the debriefing form.

RESULTS

Confirmatory factor analysis for measures

Individual measures. First I looked at the dimensionality of each of my individual measures. I tested this using a confirmatory factor analysis (CFA) using the lavaan package in R. The full-information maximum likelihood method was used, and the factor loading of the first indicator was automatically fixed to 1. I examined each measure with the dimensions loading onto individual factors, and then loaded onto a general factor. For flow, when the dimensions of control, attention, curiosity, and interest were separated, the fit was better ($\chi 2 = 186.82$, $p < .000$, SRMR = .04, RMSEA = .09, CFI = .96) than when the dimensions were loaded onto a single factor ($\chi 2 = 1003.20$, $p < .000$, SRMR = .12, RMSEA = .21, CFI = .74). This indicated the measure for flow is multidimensional.

I then looked at perceived job-relatedness and found that the fit was better when the dimensions of face validity and predictive validity were loaded onto separate dimensions ($\chi^2 = 310.21$, $p < .000$, SRMR = .09, RMSEA = .14, CFI = .90) than when they were loaded onto one general factor ($\chi^2 = 814.50$, $p < .000$, SRMR = .12, RMSEA = .24, CFI = .73). This indicated the measure for perceived job-relatedness is multidimensional.

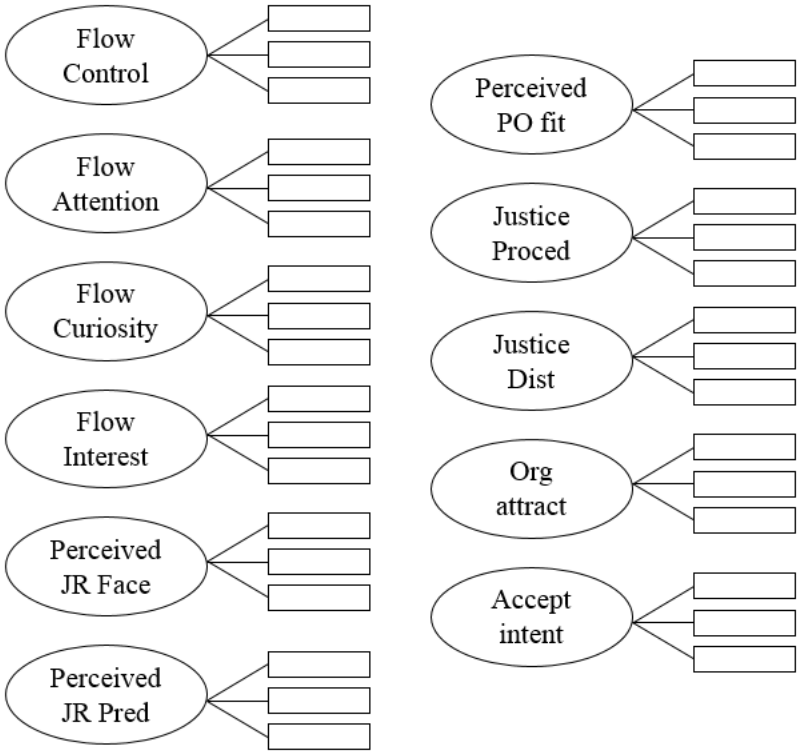
For the next measure, I examined perceived PO fit and found that when supplementary and complementary fit are separated, the overall model fit was essentially identical ($\chi^2 = 86.25$, $p < .000$, SRMR = .03, RMSEA = .16, CFI = .97) to when the dimensions were loaded onto one general factor ($\chi^2 = 86.25$, $p < .000$, SRMR = .03, RMSEA = .16, CFI = .96). This indicated perceived PO fit could be treated as a unidimensional measure.

Last I looked to see whether the justice dimensions emerged in a CFA. I found that when procedural and distributive justice were loaded onto separate dimensions the fit was slightly

better (χ^2 = 279.96, p < .000, SRMR = .18, RMSEA = .23, CFI = .80) to when they were both loaded onto to the same dimension (χ^2 = 408.37, p < .000, SRMR = .11, RMSEA = .26, CFI = .71). This offered some support for the idea that the justice measure was multidimensional.

Distinctness of measures. After examining the multidimensionality of individual measures I looked to examine whether my reactions measures were distinct. Based on the results of in the individual measure CFAs I tested the hypothesized model in which my reactions items (flow, perceived job-relatedness, perceived person-organization fit, justice, organizational attractiveness, and acceptance intentions) were all included in a CFA, accounting for multidimensionality (meaning flow, perceived job-relatedness, and justice all had multiple latent factors). This model, shown in Figure 3, did not reach recommended levels of fit (Hu & Bentler, 1999): χ^2 = 3464.59, p < .000, SRMR = .07, RMSEA = .08, CFI = .83. This may have been in part because many of the item response distributions were negatively skewed.

Figure 3. *Original hypothesized model*



Another model, in which all indicators were loaded onto one latent factor, was also tested. This model had even worse fit than the previous models: $\chi^2 = 9520.99$, $p < .000$, SRMR = .13, RMSEA = .14, CFI = .44. This offered some, albeit weak, evidence for the ideas that the items were measuring distinct constructs.

I then moved to testing a few alternative models. In Alternative Model 1 (seen in Figure 4 for all alternative models) I combined organizational attractiveness and acceptance intentions onto one latent factor because the covariance between them was high in the original model. This model had very similar fit to the originally hypothesized model: $\chi^2 = 3488.57$, $p < .000$, SRMR = .08, RMSEA = .08, CFI = .84. I then tested Alternative Model 2 in which I also loaded perceived job-relatedness and justice perceptions onto one latent factor. The fit was slightly worse than

Model 1: $\chi^2 = 4819.45$, $p < .000$, SRMR = .09, RMSEA = .10, CFI = .75. Lastly, I tested Alternative Model 3 in which the items for perceived PO fit and organizational attractiveness were loaded onto one factor due to their high collinearity, and acceptance intentions was loaded onto its own factor again (Model 3). This model was not much different than the other two alternative models: $\chi^2 = 4530.76$, $p < .000$, SRMR = .09, RMSEA = .09, CFI = .77. Overall, the best fit came from the original hypothesized model. This meant that for any analyses with flow, perceived job-relatedness, and justice, I looked at the dimensional relationships as well as the overall outcome.

Figure 4. *Alternative models*

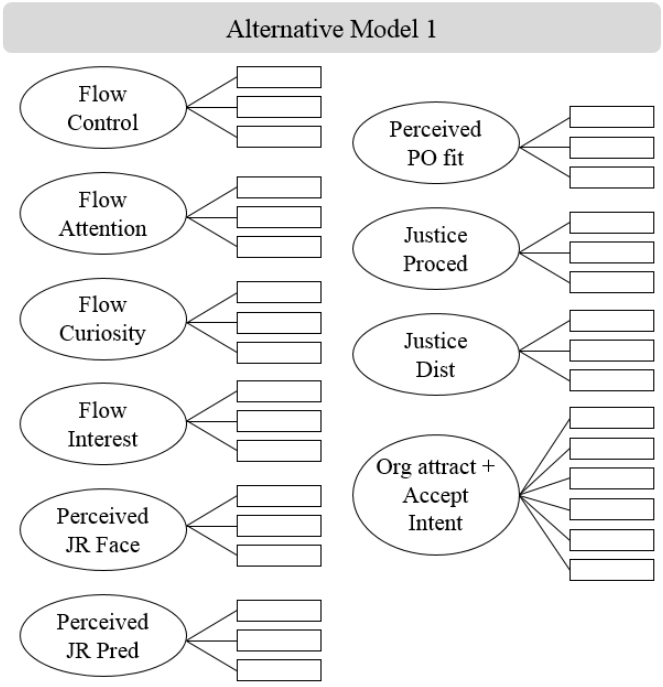
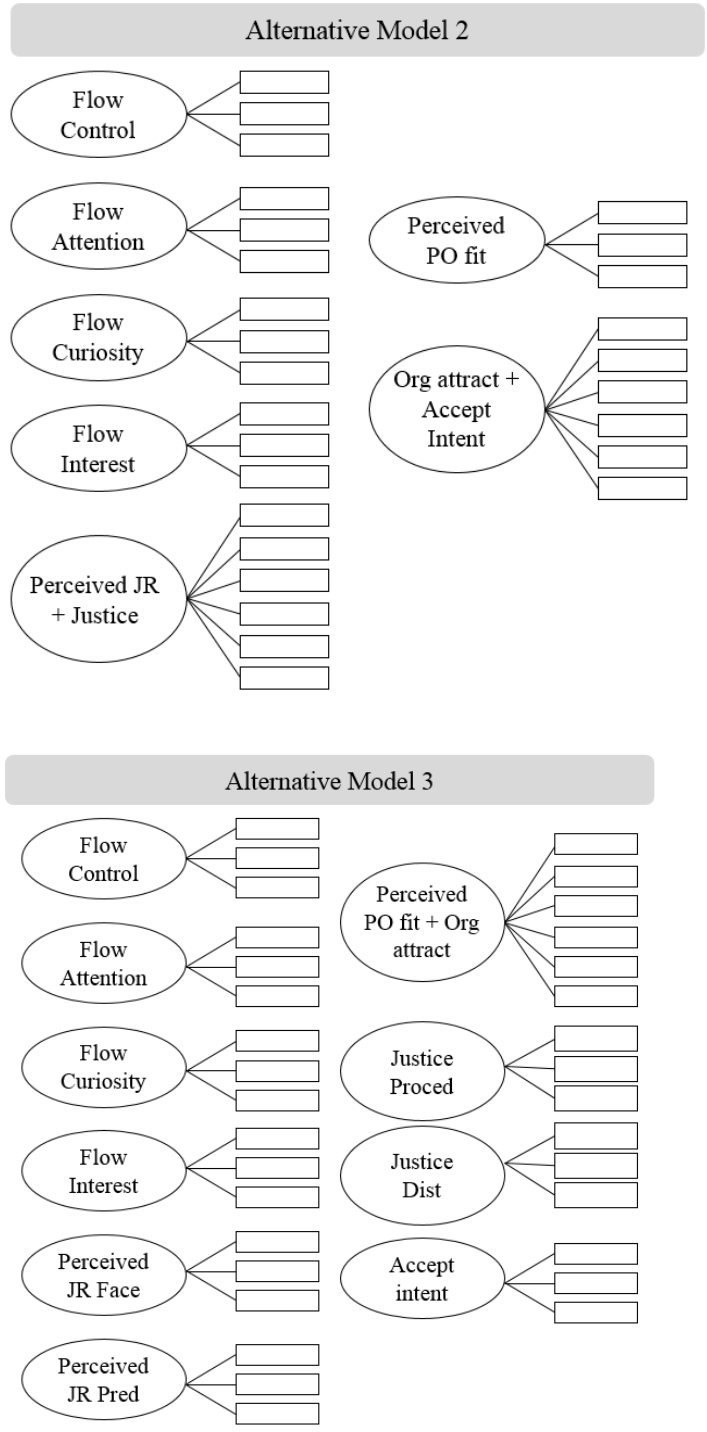


Figure 4 *(cont'd)*



Composites

Next I computed a composite for each reaction measure. In the case of justice perceptions and organizational attractiveness, I looked to see whether there was a significant difference from Day 1 to Day 2 using a paired samples t-test. For procedural justice perceptions, people felt the assessment was more fair on the first day ($M = 3.28$, $SD = 1.08$) than the second day ($M = 3.02$, $SD = 1.07$), $t(286) = 3.88$, $p < .00$. The same was true for organizational attractiveness on Day 1 ($M = 3.36$, $SD = .99$) than Day 2 ($M = 3.24$, $SD = 1.06$), $t(371) = 2.62$, $p < .01$. This was as expected, and I used the Day 2 scores for both variables in my analyses.

I also looked at the two versions of the Computer User Self-Efficacy measure to see if there were differences between smartphone and computer comfort. The sample was slightly more comfortable with smartphones ($M = 4.42$, $SD = .65$) than computers ($M = 4.31$, $SD = .75$) but was highly comfortable with both. I used a composite of the two in the analyses.

Below is a table of the relationships among the primary reactions measure composites and their dimensions (where relevant based on the CFA results).

Table 16. *Correlations among reactions measures*

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	Flow composite	-															
2	Flow- Control	**.61**	-														
3	Flow- Attention	**.70**	**.23**	-													
4	Flow- Curiosity	**.87**	**.41**	**.39**	-												
5	Flow- Interest	**.90**	**.37**	**.52**	**.84**	-											
6	Technology self-efficacy	**.35**	**.19**	**.37**	**.23**	**.30**	-										
7	Openness to experience	**.24**	.09	**.16**	**.23**	**.25**	**.31**	-									
8	Preference for games	**.27**	.09	*.12*	**.32**	**.28**	**.20**	**.45**									
9	Perceived job-relatedness composite	**.35**	**.35**	.08	**.37**	**.31**	-*.11*	.02	**.17**	-							
10	Perceived job-relatedness - Face	**.27**	**.30**	.08	**.26**	**.22**	-.08	.05	*.12*	**.89**	-						
11	Perceived job-relatedness - Predictive	**.36**	**.32**	.06	**.40**	**.33**	-*.12*	-.02	**.18**	**.88**	**.58**	-					
12	Person-organization fit	**.40**	**.21**	**.21**	**.42**	**.38**	**.16**	**.16**	**.24**	**.35**	**.28**	**.35**	-				
13	Justice perceptions composite	**.27**	**.23**	**.14**	**.26**	**.22**	-.03	.01	.05	**.49**	**.42**	**.45**	**.26**	-			
14	Justice perceptions- Procedural	**.29**	**.23**	**.16**	**.29**	**.24**	.04	.04	.04	**.41**	**.37**	**.36**	**.25**	**.91**	-		
15	Justice perceptions- Distributive	**.20**	**.18**	*.10*	**.20**	**.18**	-.06	-.01	.06	**.44**	**.37**	**.41**	**.22**	**.93**	**.65**	-	
16	Organizational attractiveness	**.47**	**.19**	**.35**	**.44**	**.45**	**.13**	.08	**.15**	**.40**	**.33**	**.36**	**.45**	**.55**	**.50**	**.50**	-
17	Acceptance intentions	**.37**	**.13**	**.28**	**.35**	**.35**	.09	.05	**.14**	**.30**	**.26**	**.28**	**.35**	**.45**	**.36**	**.44**	**.72**

Note: **Bold** indicates $p < .01$, *Italics* indicates $p < .05$

Controls

I then looked to see if there were significant differences between my two samples on the reactions measures (flow, perceived job-relatedness, perceived PO fit, justice perceptions, and organizational attractiveness). There were for many of these variables (as can been seen in Table 17), so I included a dummy variable for sample source in my future analyses as a control. The student sample had less positive reactions overall, which was somewhat unexpected but could have been a result of their inexperience with the hiring process or lower levels of engagement with the study.

Table 17. *Means and standard deviations of primary reactions variables by sample*

	Student		**MTurk**		
	M	***SD***	***M***	***SD***	**Sig**
Flow	2.87	.82	3.74	.78	.00
Perceived job-relatedness	2.31	.79	2.46	.93	.12
Perceived person-organization fit	3.20	.91	3.35	.84	.10
Justice perceptions	2.85	.80	3.19	.92	.00
Organizational attractiveness	2.91	.82	3.53	.92	.00
Acceptance intentions	3.02	1.10	3.63	1.13	.00

*Note: M*s and *SD*s are across all dimensions and time points. All 5 point scales

I next looked to see if there were differences between the two assessment method conditions for my expected controls and alternative explanation checks (belief in testing, perceived time to take the assessment, ease of faking, opportunity to perform, and perceived performance). There were for perceived time to take the assessment (the GBA was seen as taking a more reasonable amount of time, ease of faking (the non-GBA was seen as more fakeable), and perceived performance (people felt they did better on the GBA), as seen in Table 18. Because ease of faking and perceived performance relate to the assessment method (i.e., non-GBAs are often more transparent about the constructs of interest thus may be seen as more easily faked, and do not provide feedback so could be seen as more difficult) these were not used as controls

because they were thought to account for characteristics inherent to the assessment itself. Perceived time to take the assessment was added as a control in later analyses.

Table 18. *Means and standard deviations of controls by condition*

	GBA		Non-GBA		
	M	***SD***	***M***	***SD***	**Sig**
Belief in testing	3.35	1.00	3.42	.91	.44
Perceived time to take the assessment *(higher means more reasonable time)*	3.90	.91	3.52	1.01	.00
Ease of faking	2.33	.86	3.68	.94	.00
Opportunity to perform	2.61	1.17	2.67	1.16	.62
Perceived performance	3.23	1.23	2.16	.99	.00

Note: All 5 point scales

Individual hypothesis testing

Next I tested the individual links in my model. All analyses were conducted using SPSS unless otherwise noted. Sample source and perceived time to take the assessment were included as controls or covariates in all analyses unless noted.

I first looked at Hypothesis 1 to see whether there were different rates of flow between those who took the GBA as opposed to those who took the non-GBA. This was tested through an ANOVA. For all four facets and the overall composite of flow, those in the GBA condition experienced higher rates of flow than those in the non-GBA condition (see Table 19 below). Hypothesis 1 was supported.

Table 19. *ANOVA results for condition and flow*

	GBA		Non-GBA		*F*	*df1*	*df2*	*p*
	M	***SD***	***M***	***SD***				
Control	3.52	0.90	3.15	0.98	8.26	1	371	.00
Attention	3.92	1.08	3.53	1.21	8.42	1	371	.00
Curiosity	3.73	1.06	3.03	1.21	22.99	1	372	.00
Interest	3.96	1.03	3.01	1.22	52.56	1	372	.00
Composite	3.78	0.79	3.18	0.88	39.73	1	372	.00

Next I used moderated regressions to test whether the relationship between assessment method and the dimensions of flow was moderated by individual characteristics and preferences. The predictor, assessment method, was dummy coded with the non-GBA condition as the referent. The composites for technology self-efficacy, openness to experience, and preference for games were centered. Fifteen hierarchical regressions were then conducted: an overall flow composite and four flow dimensions for each of the three moderators. Technology self-efficacy was the only one of the three that significantly moderated the method-flow connection. The flow composite was significant (b = .25, $t(375) = 2.27$, $p < .05$) as was the curiosity dimension (b = .39, $t(375) = 2.31$, $p < .05$) as can be seen in Tables 20 and 21. Neither the flow composite nor any dimension moderated regressions were significant for openness to experience or preference for games; the composite analysis is reported for these in Tables 22 and 23 respectively. Hypothesis 2a was supported, but 2b and 2c were not.

Table 20. *Technology self-efficacy as a moderator of the method-flow composite relationship*

	R	***R²***	***R² Change***	***b***	***t***
Step 1	.61	.37			
Intercept				2.36**	6.07
Sample source				-.52**	-6.49
Perceived time to take assessment				.32**	8.13
Step 2	.67	.45	.08**		
Technology self-efficacy				.10	1.32
Assessment method dummy code				.44**	6.30
Step 3	.68	.46	.01*		
Technology self-efficacy x Assessment method dummy code				.25*	2.27

** $p < .01$, * $p < .05$

Table 21. *Technology self-efficacy as a moderator of the method-flow curiosity dimension relationship*

	R	***R²***	***R² Change***	***b***	***t***
Step 1	.49	.24			
Intercept				2.22**	3.72
Sample source				-.40**	-3.22
Perceived time to take assessment				.42**	7.02
Step 2	.54	.29	.05**		
Technology self-efficacy				-.03	-.23
Assessment method dummy code				.51**	4.76
Step 3	.55	.29	.01*		
Technology self-efficacy x Assessment method dummy code				.39*	2.31

** $p < .01$, * $p < .05$

Table 22. *Openness to experience as a moderator of the method-flow composite relationship*

	R	***R²***	***R² Change***	***b***	***t***
Step 1	.61	.37			
Intercept				1.97**	5.96
Sample source				-.57**	-7.34
Perceived time to take assessment				.34**	8.77
Step 2	.67	.45	.08**		
Openness to experience				.23**	2.92
Assessment method dummy code				.44**	6.26
Step 3	.67	.45	.00		
Openness to experience x Assessment method dummy code				.00	.00

** $p < .01$, * $p < .05$

Table 23. *Preference for games as a moderator of the method-flow composite relationship*

	R	***R²***	***R² Change***	***b***	***t***
Step 1	.61	.37			
Intercept				2.259**	8.11
Sample source				-0.55**	-7.11
Perceived time to take assessment				0.33**	8.82
Step 2	.68	.47	.09**		
Preference for games				0.14*	2.44
Assessment method dummy code				0.47**	6.72
Step 3	.69	.47	.00		
Preference for games x Assessment method dummy code				.12	1.43

** $p < .01$, * $p < .05$

I then followed up on the significant moderation for technology self-efficacy with a simple slopes analysis. The results showed that there was a significant positive linear relationship between technology self-efficacy and the flow composite for those in the GBA condition (b = .35, SE = .09, $p < .00$) but not those in the non-GBA condition (b = .10, SE = .07, p = .19). This means that people in the GBA condition were significantly more likely to experience a flow state

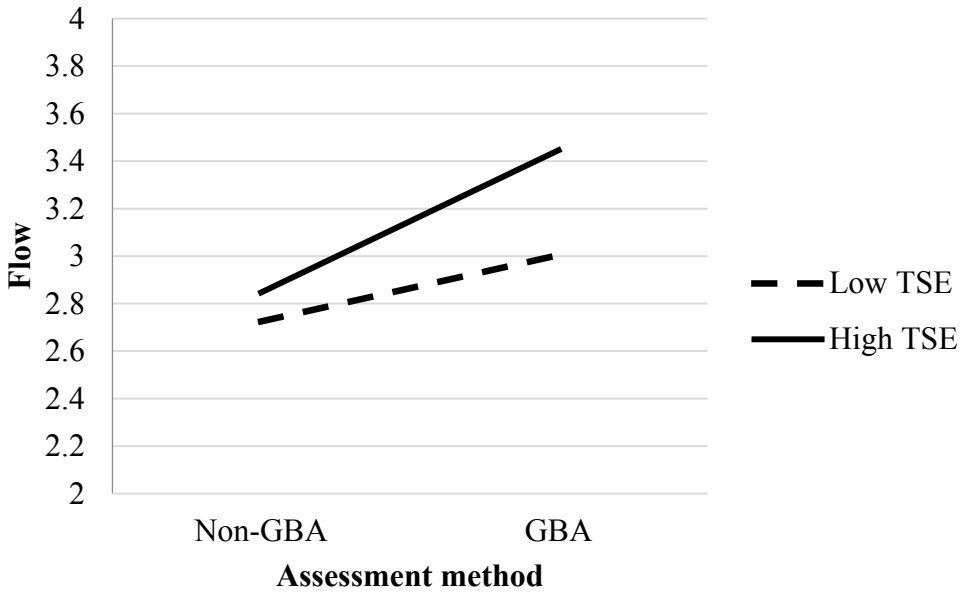
when they had high technology self-efficacy as opposed to low, but that there was no significant difference in flow between those high and low in technology self-efficacy for those in the non-GBA condition. The regression results can be found in Table 24, and Figure 5 demonstrates the simple slopes.

Table 24. *Regression results for moderation of technology self-efficacy and flow composite*

	GBA	**Non-GBA**
Intercept	1.70	2.36
Slope for technology self-efficacy	.35**	.10

** $p < .01$

Figure 5. *Moderating effect of technology self-efficacy on assessment method-flow composite relationship*



I also conducted a simple slopes analysis for the flow curiosity dimension. The results again showed that there was a significant positive linear relationship between technology self-efficacy and the flow curiosity dimension for those in the GBA condition ($b = .37$, SE = .14, $p < .00$) but not those in the non-GBA condition ($b = -.03$, SE = .11, $p = .82$). This was interpreted to

mean that those in the GBA condition felt higher levels of curiosity-related flow when they had high technology self-efficacy as opposed to low, but that again there was no difference in curiosity-related flow between those high and low in their comfort with technology for those in the non-GBA condition. These results mirrored the pattern for the general flow dimension. Table 25 represents this below.

Table 25. *Regression results for moderation of technology self-efficacy and flow curiosity dimension*

	GBA	**Non-GBA**
Intercept	1.01	2.22
Slope for technology self-efficacy	.37**	-.03

** $p < .01$

Next I was interested in Exploratory Question 1, the relationship between flow and perceived job-relatedness. This was tested through correlations. The flow composite and perceived job-relatedness were positively related ($r = .38$, $p < .001$). The dimensions of each were almost all also positively correlated, as shown in Table 26. The exceptions to this were the correlations between the attention dimension of flow and perceived job-relatedness; neither of these relationships was significant indicating that absorptive elements of the assessment did not relate to perceptions of job-relatedness.

Table 26. *Correlations between flow and perceived job-relatedness*

		1	2	3	4	5	6	7
1	Flow Composite	-						
2	Flow- Control	.60**	-					
3	Flow- Attention	.60**	.17**	-				
4	Flow- Curiosity	.83**	.34**	.22**	-			
5	Flow- Interest	.85**	.27**	.36**	.79**	-		
6	Perceived job-relatedness composite	.38**	.35**	.05	.37**	.32**	-	
7	Perceived job-relatedness- Face	.27**	.29**	.05	.24**	.20**	.89**	-
8	Perceived job-relatedness- Predictive	.40**	.33**	.03	.42**	.37**	.88**	.57**

** $p < .01$

For the next link, I speculated that the non-GBA would have a stronger positive effect on perceived job-relatedness, controlling for the effect of flow. I used an ANOVA in which the flow composite was included as an additional covariate (because I wanted to account for all aspects of flow and was not particularly interested in how much of the variance was accounted for by each dimension). Those who were in the non-GBA condition ($M = 2.42$, $SD = .92$) did experience significantly higher overall perceived job-relatedness than those in the GBA condition ($M = 2.38$, $SD = .85$), $F(1,371) = 13.34$, $p < .001$. This finding was replicated for the face validity dimension of perceived job-relatedness, but not the predictive validity dimension (as can be seen in Table 27 below). Hypothesis 3 was partially supported.

Table 27. *ANOVA results for condition and perceived job-relatedness*

	GBA		Non-GBA		*F*	*df1*	*df2*	*p*
	M	*SD*	*M*	*SD*				
Perceived job-relatedness composite	2.38	0.85	2.42	0.92	13.34	1	371	.00
Perceived job-relatedness- Face	2.41	0.98	2.65	1.06	22.91	1	371	.00
Perceived job-relatedness- Predictive	2.36	0.91	2.19	1.01	2.06	1	371	.15

For Hypothesis 4, I looked to see whether perceptions of job-relatedness were positively related to justice perceptions. I examined the composite and dimension correlations for both, as shown in Table 28. All dimensions significantly correlated, offering support for this link. In an exploratory manner, I tested whether this was true for both the GBA and non-GBA condition, and it was supported in both.

Table 28. *Correlations between perceived job-relatedness and justice perceptions*

		1	2	3	4	5
1	Perceived job-relatedness composite	-				
2	Perceived job-relatedness- Face	.88**	-			
3	Perceived job-relatedness- Predictive	.87**	.54**	-		
4	Justice perceptions composite	.40**	.35**	.35**	-	
5	Justice perceptions- Procedural	.39**	.35**	.34**	.91**	-
6	Justice perceptions- Distributive	.33**	.27**	.30**	.90**	.63**

** $p < .01$

Next I looked at whether selection decision moderated the relationship between perceived job-relatedness and justice perceptions. I created a dummy variable for whether a participant received an offer or not, with not being given an offer as the referent. I centered perceived job-relatedness (the composite, face validity, and predictive validity). I then conducted moderated regressions for the composite and two facets of perceived job-relatedness regressed onto the composite for justice perceptions, as well as Day 2 procedural and distributive justice dimensions.

Several of these relationships were significant. When the composite for perceived job-relatedness was used, there was significant moderation as seen in Table 29. The same was true when face and predictive validity dimensions were used as the predictor, as can be seen in Tables 30 and 31.

Table 29. *Selection decision as a moderator of the relationship between perceived job-relatedness composite and justice perceptions composite*

	R	***R²***	***R² Change***	***b***	***t***
Step 1	.15	.02			
Intercept				1.52**	5.20
Sample source				-.07	-.75
Perceived time to take assessment				.07	1.50
Step 2	.61	.38	.35**		
Perceived job-relatedness composite				.38**	5.17
Selection decision dummy code				.74**	8.97
Step 3	.62	.39	.01*		
Perceived job-relatedness composite X Selection decision dummy code				.26*	2.78

** $p < .01$, * $p < .05$

Table 30. *Selection decision as a moderator of the relationship between perceived face validity and justice perceptions composite*

	R	***R²***	***R² Change***	***b***	***t***
Step 1	.15	.02			
Intercept				1.86**	6.27
Sample source				-.08	-.81
Perceived time to take assessment				.09	1.88
Step 2	.55	.30	.28**		
Perceived face validity				.20**	3.13
Selection decision dummy code				.69**	7.97
Step 3	.57	.32	.02**		
Perceived face validity X Selection decision dummy code				.29**	3.44

** $p < .01$, * $p < .05$

Table 31. *Selection decision as a moderator of the relationship between perceived predictive validity and justice perceptions composite*

	R	***R²***	***R² Change***	***b***	***t***
Step 1	.15	.02			
Intercept				1.57**	5.48
Sample source				-.07	-.78
Perceived time to take assessment				.07	1.49
Step 2	.61	.37	.34**		
Perceived predictive validity				.37**	5.64
Selection decision dummy code				.80**	9.65
Step 3	.61	.37	.01*		
Perceived predictive validity X Selection decision dummy code				.19*	2.16

** $p < .01$, * $p < .05$

Upon closer study, these relationships seem to be driven by distributive justice as all three of the moderated regressions were significant when this dimension was used as the dependent variable ($b = .36$, $t(370) = 3.50$, $p = .001$ for perceived job-relatedness composite; $b = .36$, $t(370) = 3.85$, $p = .000$ for face validity; $b = .28$, $t(370) = 2.92$, $p = .004$ for predictive validity), but none were significant when procedural justice was used as the dependent variable. This made sense, as I was examining whether selection decision was a moderator.

To further explore the moderated regression, I did a simple slopes analysis for perceived job-relatedness and its dimensions with distributive justice as the dependent variable. The results showed there was a significant positive relationship between perceived job-relatedness and distributive justice perceptions for both those who were not selected ($b = .28$, SE = .08, $p < .00$) and those who were selected ($b = .64$, SE = .07, $p < .00$) but with a steeper slope for those who were selected. This meant that there was a stronger relationship between perceived job-relatedness and distributive justice perceptions for those who were offered the position than for

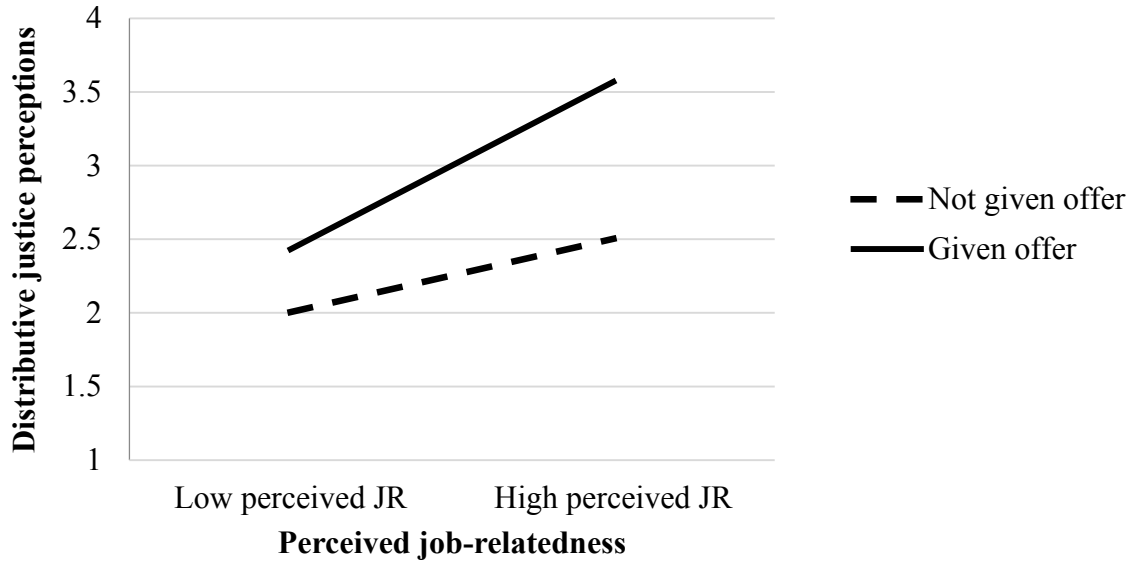
those who were not but that in both groups, perceived job-relatedness and distributive justice were related. Results can be found in Table 32, and Figure 6 demonstrates the simple slopes.

Table 32. *Regression results for moderation of perceived job-relatedness and distributive justice*

	Not selected	**Selected**
Intercept	1.58	1.46
Slope for perceived job-relatedness	.28**	.64**

** $p < .01$

Figure 6. *Moderating effect of selection decision on perceived job relatedness-distributive justice perception relationship*



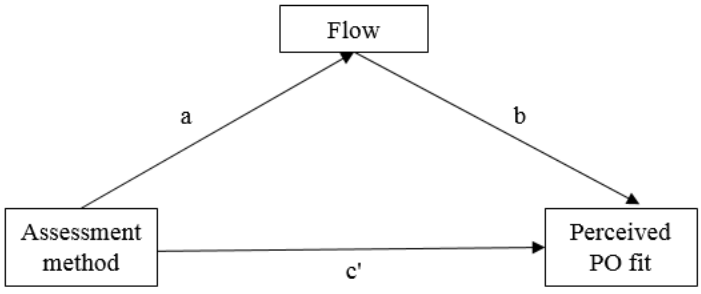
The relationships for the two dimensions of perceived job-relatedness followed similar patterns. Hypothesis 5 was partially supported in that selection decision was a significant moderator, but in a different direction than expected and only for distributive justice. Because these results were different than expected, I looked again at this relationship but as an exploratory three-way interaction, this time including method condition. The result was not

significant, F(17,223) = .95, p = .52 indicating no difference in the moderation for those who took the GBA and non-GBA.

For Hypothesis 6a, I tested whether there was a difference between assessment method and relationship to perceived PO fit. I first regressed the two control variables onto perceived PO fit, and then I regressed a dummy variable for assessment method. There was no significant change in the R^2 when the dummy variable was added, b = -.14, $t(375)$ = -1.56, p = .12. This hypothesis was not supported.

For Hypothesis 6b, I tested whether flow mediated the relationship between assessment method and perceived PO fit. The model is depicted in Figure 7. The path coefficient indicated by the letter (a) stands for the relationship between the predictor (a dummy variable for assessment method in which 0 was the non-GBA condition and 1 was the GBA condition) and the mediator (flow). The path marked (b) represents the relationship between the mediator (flow) and the outcome variable (perceived PO fit). The coefficient (c') is the direct relationship between the predictor and outcome. In order to get the indirect effect, or the relationship of the antecedent to the outcome via the mediator, one can multiply (a) by (b).

Figure 7. *Mediating effect of flow*



This was tested using the PROCESS program (Hayes, 2013). I used a bootstrap approach to calculate the confidence intervals, as this method does not make assumptions about the distribution of the sample. The results for the mediation model can be found in Table 33. All reported values are unstandardized. I found that assessment method indirectly influenced perceived PO fit through the effect of flow. Those in the GBA condition experienced higher rates of flow than those in the non-GBA condition (a = .57), and those who experienced a higher rate of flow were more likely to have higher rates of perceived PO fit (b = .54). A bias-corrected bootstrap confidence interval for the indirect effect (ab = .31) based on 10,000 bootstrap samples did not include zero in its confidence interval (.22 to .42). However, assessment method did have a significant influence on perceived PO fit when flow was accounted for, therefore this was considered a partial mediation (c'= -.37, $p < .01$).

For two of the dimensions of flow – curiosity and interest – this same relationship held and there was partial mediation. For the other two dimensions – control and attention – the direct effect of assessment method on perceived PO fit was not significant when the dimension of flow was accounted for, indicating full mediation as can be seen in Table 33 below. Hypothesis 6b was partially supported.

Table 33. *Mediating effect of flow on assessment method and PO fit*

Flow dimension	Total effect of IV	Direct effect of IV	Mediation by Flow				
		c'	a	b	ab	LL CI for ab	UL CI for ab
Composite	-.07	-.37**	.57**	.54**	.31	.22	.42
Control	-.07	-.14	.36**	.21**	.08	.03	.14
Attention	-.07	-.13	.34**	.19**	.06	.02	.13
Curiosity	-.07	-.31**	.67**	.36**	.24	.16	.34
Interest	-.07	-.40**	.91**	.37**	.34	.24	.45

** $p < .01$, * $p < .05$

For the next relationship, Hypothesis 7, I looked at a well-established relationship: whether perceived PO fit would be positively related to organizational attractiveness (on Day 2). It was, $r = .43$, $p < .001$ so this hypothesis received support.

Next I looked to see whether selection decision would moderate the relationship between perceived PO fit and organizational attractiveness (on Day 2). The same offer dummy variable from Hypothesis 5 was used, and perceived PO fit was centered. This moderation was not significant, as can be seen in Table 34. Therefore Hypothesis 8 was not supported.

Table 34. *Selection decision as a moderator of the perceived PO fit- organizational attractiveness relationship*

	R	***R²***	***R² Change***	***b***	***t***
Step 1	.32	.10			
Intercept				1.50**	4.37
Sample source				-0.39**	-3.80
Perceived time to take assessment				0.13*	2.57
Step 2	.58	.34	.24**		
Perceived person-organization fit				0.46**	6.34
Selection decision dummy code				.58**	6.40
Step 3	.58	.34	.00		
Perceived person-organization fit X Selection decision dummy code				.12	1.16

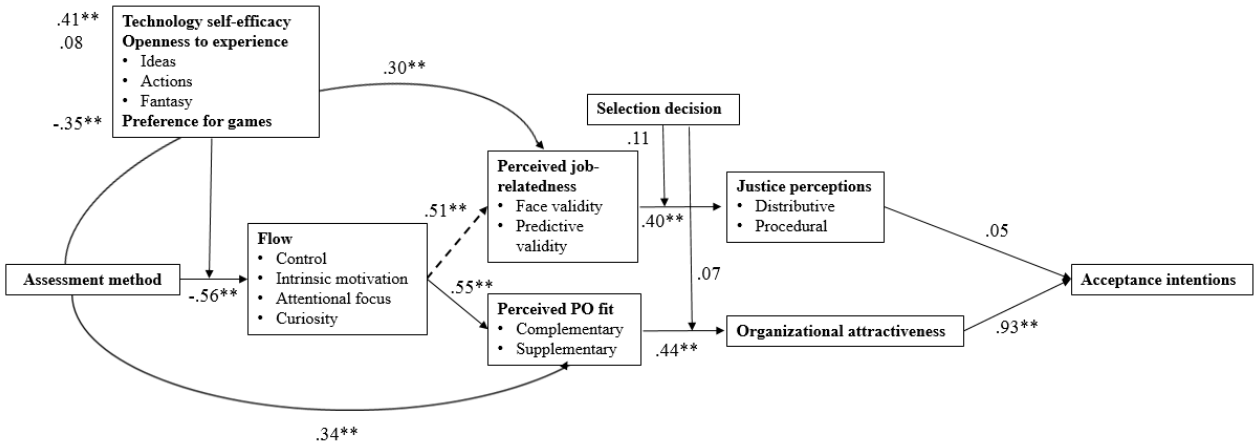
** $p < .01$, * $p < .05$

Finally, I looked to see if justice perceptions and organizational attractiveness were positively related to acceptance intentions. Both the composite ($r = .36$, $p < .001$), procedural ($r = .30$, $p < .001$), and distributive ($r = .36$, $p < .001$) of justice were related to acceptance intentions. Likewise organizational attractiveness (on Day 2) was strongly related to acceptance intentions ($r = .71$, $p < .001$) offering support for Hypotheses 9 and 10.

Overall model testing

Lastly, I looked to see whether there was fit for the overall model. I used the structural equations modeling function in MPlus (Muthén & Muthén, 1998-2017). I used maximum likelihood estimation with 1,000 bootstrap samples. For assessment method, the GBA condition was coded as 1 and the non-GBA as 2. The unstandardized path coefficients can be seen in Figure 8 below.

Figure 8. *Path coefficients for full model*



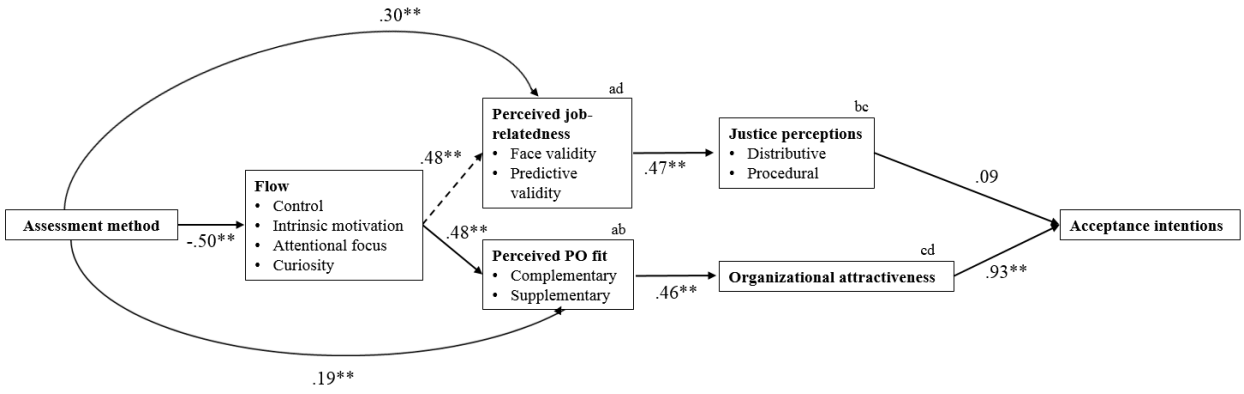
While many of the hypothesized relationships were supported in the path coefficients of the model, the model did not achieve recommended levels of fit, $\chi^2 = 611.92$, $p < .00$, SRMR = .11, RMSEA = .17, CFI = .60.

I then tested a few additional models based on the modification indices. In the first alternative model, I removed selection decision as a moderator. The fit for this model was improved, $\chi^2 = 280.17$, $p < .00$, SRMR = .08, RMSEA = .13, CFI = .80. I then also removed the three personality moderators (technology self-efficacy, openness to experience, and preference for games) and saw higher values for SRMR, RMSEA, and CFI indicating similar fit to when just the selection decision moderator was removed, $\chi^2 = 145.81$, $p < .00$, SRMR = .10, RMSEA = .19, CFI = .88. Then I allowed a few of the variables to correlate: organizational attractiveness and perceived PO fit each with justice perceptions and perceived job-relatedness. This improved the fit indices to nearly acceptable levels, $\chi^2 = 26.35$, $p < .00$, SRMR = .05, RMSEA = .09, CFI = .98.

Finally, I tested two versions in which the "top half" (perceived job-relatedness and justice perceptions) or the "bottom half" (perceive PO fit and organizational attractiveness) were removed, and this brought my fit indices closer to accepted ranges, though not as good as the model in which both moderators were removed and variables were allowed to correlate. The improvement in fit was possibly because of the penalty for model complexity inbuilt to the CFI and RMSEA indices (Kenny, 2015).

The best fitting model was the one shown below in Figure 9 where all moderators were removed and in which variables with the same letter above their box were allowed to correlate.

Figure 9. *Path coefficients for revised model without moderators and with intercorrelations*



Exploratory analyses on age differences

Because of the ongoing study of whether age affects the experience of game-based tools (Bittner & Schipper, 2014; Koivisto & Hamari, 2014), I looked at a few exploratory relationships to understand whether age accounted for differences in my sample.

I first looked at whether age was significantly related to technology self-efficacy, as some researchers have made the argument that comfort with technology may be the root cause of older users' disengagement (Chung, Park, Wang, Fulk, & McLaughlin, 2010; Venkatesh, Morris, Davis, & Davis, 2003). Only 20.7% of my total sample was over 40, but there was still a positive

correlation between technology self-efficacy and age ($r = .17$, $p < .001$, using age as a continuous variable). My sample was unique in that almost all of my older participants' data was collected through MTurk, indicating that they as a group may be more comfortable with technology than the average population.

I then looked to see whether there was a difference in preference for games by age. This relationship was insignificant age ($r = .02$, $p = .76$).

Next I was interested in whether there would be a difference in awareness of game elements by age. I dichotomized my participants as under and over 40, and conducted an ANOVA (still controlling for sample source and perceived time to take the assessment). None of the game elements (focused goals/rules, feedback, challenging tasks, and sensory stimuli) were perceived significantly differently by those under 40 as those over 40 as can be seen in Table 35.

Table 35. *ANOVA results for awareness of game elements by age*

	Under 40		**Over 40**		***F***	***df1***	***df2***	***p***
	M	***SD***	***M***	***SD***				
Goals/Rules	4.26	.95	4.21	.92	.19	1	364	.67
Feedback	2.79	1.35	2.51	1.47	2.40	1	364	.12
Challenging Tasks	4.17	.87	4.26	.73	.07	1	364	.80
Sensory Stimuli	3.87	1.33	3.60	1.55	2.77	1	364	.10

Finally, I examined whether the experience of flow differed by age, again using an ANOVA. There was no significant difference $F(1,364) = 1.46$, $p = .23$. These findings as a whole indicated that if anything there was slightly more comfort with technology for older participants, and no significant impact of age on experience with game-based tools.

DISCUSSION

The purpose of this paper was to examine applicant reactions to game-based assessments in a controlled and theoretically-grounded manner. This is one of the first studies that uses a diverse sample to understand how GBAs in a selection context affect important reactions such as perceived job-relatedness, PO fit, justice perceptions, organizational attractiveness, and intentions to accept. Importantly, this study used an experimental context to not just look at GBAs, but compare applicant reactions on these novel tools to more traditional ones like cognitive and personality assessments. This study also introduced a well-known but previously unused concept into the selection literature: flow. I looked at whether GBAs induce a state of flow in which the applicant feels focused, in control, interested, and curious during his or her experience. Lastly, this study examined whether a number of moderators were important in understanding the relationship between assessment method and reactions. Specifically, I examined whether technology self-efficacy, openness to experience, and preference for games were important from an individual difference perspective, and whether a manipulated selection decision had an impact on reactions. At a time when the use of game elements is expanding rapidly, this work is one of the first empirical studies that can help further our understanding of how applicants respond to the new trend in a hiring environment.

Summary of Findings

In general there was support for the hypothesized model. One of the most interesting findings was that participants who took the GBA experienced higher rates of flow than those who took the non-GBA. This suggests that, as advertised, GBAs and the game elements that constitute them have the ability to captivate applicants in a way that more traditional measures are unable to do.

Further, this relationship between assessment method and flow was moderated by technology self-efficacy, where when a participant took the GBA he or she was significantly more likely to experience flow if he or she expressed comfort and feelings of efficacy with technology. Participants who felt confident and capable in their ability to navigate technology were able to more fully engage with the task at hand. In accordance with cognitive load theory, when applicants use fewer cognitive resources navigating technology, they are able to devote more attention to the task (Landers & Callan, 2012), facilitating a flow state. Interestingly, my exploratory analyses showed that in this study age was positively related to technology self-efficacy. This runs counter to work that shows older populations are less comfortable with technology (Venkatesh et al., 2003) but was promising in that it means incorporating game elements into assessment tools may not frustrate older applicants as some might think it would. One possible rationale is that games satisfy basic human needs, which span generational cohorts (Sailer, Hense, Mayr, & Mandl, 2017). Nevertheless, there are differences in comfort with technology within applicant populations and this moderation indicates that a lack of efficacy can cause lower levels of flow. This is one of likely several boundary conditions to the relationship between GBAs and flow.

While technology self-efficacy was a significant moderator, openness to experience and preference for games were not even though both of these constructs seemed to have sufficient range and variability in the sample. As mentioned in the procedure for the main study, I asked participants if they had any general comments about the study. In these qualitative comments, quite a few people mentioned that the game was "fun" but that if it was used for a job they really wanted, they would be displeased. Because the openness and preference for games measures capture receptivity and comfort with games generally, it is possible that these measures are not

capturing the true source of resistance. It may matter less whether people like new experiences or game-based tools, and depend more on their ideas of what a selection process "should" look like or the circumstances under which they believe it is appropriate to employ gamed tools. A person may enjoy playing video games and having a fitness tracker that awards points for every mile walked, but may balk at the idea of this kind of frivolity being applied in a work context. In this case, a more appropriate measure may have been one that assessed preference for gamification or comfort with serious games. These kinds of measures may better capture people's proclivity toward the use of games in non-game contexts, as was done in this study.

The next relationship was between flow and perceived job-relatedness. For this exploratory question I had offered supporting arguments for both a negative and positive correlation. The relationship was positive, indicating that entering a state of flow was related to seeing the assessment as job-related. In a supplemental analysis I looked at this finding by assessment method condition, and found it was the case in both conditions. This means there may be some support for the feelings-as-information perspective that if a participant is in a positive state (Schwarz & Clore, 1983; Wyer & Carlston, 1979), that some of those attributions can be ascribed to the assessment's relevance to the job. GBAs are some of the first hiring tools that are built to be engaging without necessarily being face valid, so it is certainly of interest that the engagement trickles into other aspects. However, the non-GBA assessments used in this study (cognitive and personality tests) also suffer from low job-relatedness perceptions (Macan et al., 1994; Pulakos, 2005), so one may get a different perspective in comparing a GBA to a straightforward but highly job-related tool like a work sample.

While the non-GBA assessments used here do not have as favorable of perceived job-relatedness ratings as simulations or work samples, they were still rated as more job-related,

notably face valid, than their GBA counterpart when flow was controlled for. This means that GBAs may only be seen as job-related when they are able to induce flow, but that even then increased perceptions of job-relatedness are not guaranteed.

I looked at the link between perceived job-relatedness and justice in the next hypothesis, and in keeping with the bulk of the literature (Hausknecht et al., 2004), all composites and dimensions of each of the constructs were positively related. I then looked at whether selection decision – whether the participant was offered the job or not – moderated this relationship. The moderations were significant using both the composite and the two dimensions of perceived job-relatedness, driven mainly by distributive justice (the regressions for the justice composite were significant, but those for procedural justice were not). In some ways this made sense as I was looking at the difference in offer conditions, which has a larger impact on distributive justice. Interestingly, the simple slopes analysis was significant for both those who were made an offer and those who were not, but contrary to my hypothesizing the slope was steeper for those who were made an offer. This meant that when a participant felt the assessment was job-related *and* was given an offer, his or her perceptions of distributive justice were at their highest (higher than when the participant simply received an offer but did not view the assessment as job-related). The rationale here may be that feeling the assessment was job-relevant added to participants' feelings of justice; it was not just that they got the job, but that they got it by showcasing skills that were pertinent to the role.

While the results of the moderation were at odds with some other work that has tested the same relationship and found rejected applicants to be more affected by perceived job-relatedness (i.e., Gilliland, 2004), the work on outcome favorability and justice perceptions has often shown that when an applicant is not made an offer, he or she will experience less justice (Bauer et al.,

1998; Kluger & Rothstein, 1993). In these cases, other factors may matter less, and the failure to receive an offer may serve as a dampener. The self-serving attribution bias can also help explain the results of this study because it also supports the idea that those who perceive the assessment to be fair and are selected for the role will experience heightened justice perceptions (Chan, Schmitt, DeShon, Clause, & Delbridge, 1997). Further in this study distributive justice was measured on Day 2; at that point, the offer decision was likely much more salient than the perceived job-relatedness of the assessment as it often would be in a real-life scenario too.

Moving on from perceived job-relatedness, I then looked at whether assessment method had a direct relationship with person-organization fit. This relationship was not significant, indicating either that the values signaled by the two assessment methods were not much different from one another, or that these signals were not being attributed back to the organization and used to evaluate fit. While one could argue that an experiment in which the participant has no other information to go on to make PO fit judgments *except* the assessment is the only way to get a pure measure of the relationship, we should also acknowledge that an experimental environment may be too superficial for a participant to really reflect on what the assessment says about the values of the organization.

Taking this analysis a step further, I did find support for the idea that flow mediates the relationship between assessment method and perceived PO fit. For the flow composite as well as the dimensions of curiosity and interest, flow partially mediated the relationship between assessment method and perceived PO fit. For the flow dimensions of control and attention, flow fully mediated the relationship between the assessment method and perceived PO fit. Interestingly, while all of the indirect effects were positive (indicating those in the GBA condition had higher perceived PO fit as a result of their increased rates of flow), for the flow

composite, curiosity dimension, and interest dimension, the direct effects were negative. This is what Zhao and colleagues have called competitive mediation (2010), and may indicate that there is another significant mediator that is not included in the model.

In the next analysis, another robust relationship was supported. Perceived PO fit was positively correlated with organizational attractiveness, in keeping with the idea that people will opt to work in organizations with which they feel they fit (Byrne, 1971; Schneider, 1987). However, contrary to my hypothesizing, selection decision did not moderate this relationship. Some work shows that if the selection process is perceived to be fair and/or applicants have a positive experience, they may be willing to reapply to the company or position in the case that they are rejected (Konradt, Warstza, & Ellwart, 2013; LaHuis, MacLane, & Schlessman, 2007). To further explore this I conducted a supplemental analysis with my data and looked at a combined and dichotomized version of the items that asked about intentions to accept (coding one through three on the Likert scale as "no" and four and five as "yes"). Almost 48% of participants who were not offered the job said they would accept it if it was offered at a later point. In this case, failing to receive an offer may not have as big of an impact on organizational attractiveness as I suggested. Applicants may not see this as their one chance to enter the organization, especially in today's world of increased mobility (Giang, 2016).

Finally in the last of the individual hypotheses I looked to see if justice perceptions and organizational attractiveness were related to intentions to accept the offer. These relationships were both positive and significant, as expected. When applicants feel they are able to have input to the process, and experience fair outcomes, they are more likely to join the organization that provided the positive experience (Hausknecht et al., 2004). Similarly, if an applicant is attracted to an organization and is able to join, he or she is likely to do so.

I believed it was important to test the individual links of this model as some of these linkages are new to the selection literature, but I tested the overall model as well. The model with the best fit was one that excluded all moderators (technology self-efficacy, openness to experience, preference for games and selection decision), and allowed many of the reactions measures to correlate. This was logical; many of the moderators (with the exception of technology self-efficacy, and selection decision for the perceived job-relatedness to justice perceptions relationship) were not significant, and many of the reactions measures were moderately correlated. This model met the recommended values almost across the board, offering support for this modified version of the overall model. This meant that GBAs did induce higher rates of flow than non-GBAs, as well as positively relate to a host of applicant reactions. With flow included in the model, GBAs garnered higher ratings of perceived job-relatedness and justice perceptions than non-GBAs meaning that in this scenario were the GBA induced higher rates of flow than the non-GBAs, engaging tools were viewed as more valid and fair predictors. GBAs were also positively associated with PO fit and organizational attractiveness, supporting the often-touted benefit that GBAs can promote a positive brand image, and help applicants assess their fit (Shergill,, n.d.). Lastly, GBAs positively related to acceptance intentions, which is often organization's ultimate goal when designing applicant-centric selection systems.

In addition to the hypothesized findings, there were a few interesting conclusions or implications of the measures used as manipulation or alternate explanation checks. The first is around the preference for games measure. This measure was created for this study, had a reasonable alpha (.87), and should be useful in other studies where the effect of game elements or gamification is examined. The second discussion is around the awareness of game elements measure. It was also designed for this study, around the dimensions of focused goals/rules,

feedback, challenging tasks, and sensory stimulation. While the alphas in the pilot were acceptable (α = .88 for the GBA, α = .71 for the non-GBA), they were quite a bit lower in the main study (α = .69 for the GBA, α = .45 for the non-GBA). There may have been a few reasons for this. While all of the items asked about game elements, during the literature review I illustrated how enormously varied game elements, and their uses, are. Therefore we may not expect items that ask about different game elements to be highly internally consistent. To further explore this, I provide the reliability for each of the four dimensions in Table 36 below.

Table 36. *Dimensional reliabilities for awareness of game elements*

Facet	**α for GBA**	**α for non-GBA**
Focused goals/rules	.79	.79
Feedback	.36	.56
Challenging tasks	.69	.77
Sensory stimulation	.89	.90

These reliabilities were generally better than the overall. Noticeably, the feedback dimension had by far the worst reliabilities; this was the only dimension for which there was a reverse-coded item. While the data were cleaned based on attention checks, participants may not have used the scale points correctly.

I also included a number of measures to serve as alternative explanation checks. By examining whether there were differences between my two groups on constructs like ease of faking and perceived performance, I hoped to better understand what effect the experience had on the participants' reactions more broadly. One of the measures I created to use as an alternative explanation check was perceived time to take the assessment. This two-item measure had a lower alpha than considered acceptable (α = .55 across method conditions). Upon further examination, this was not much different when calculated by condition (α = .46 for the GBA, α = .59 for the non-GBA). In part, this may be due to the role of flow. Flow engrosses the participant, and can

remove the awareness of time. People who were in varying states of flow while taking the assessment may have responded to the scale differently, adding error to the measure.

Limitations

There were a few limitations to this study. While it would have been nearly impossible to compare two assessment methods in a true selection context, the first limitation is that in the experimental setting of this study, participants may have responded and had different reactions to the assessments than if they were truly applying for a job. Participants may have reacted more negatively to the GBA, or found it harder to enter a state of flow if they were more concerned about their assessment performance.

A second limitation was the almost exclusive use of self-report in the study. The dangers of common-method bias are well-known (Lindell & Whitney, 2001). Other than the assessments, my measures were all self-report which could have contributed to shared variance among my constructs. This limitation was somewhat mitigated by the fact that my CFAs showed dimensionality within and some support for differentiation between my constructs. In the future, researchers exploring this kind of work could use more behavioral measures of reactions measures. For example, to measure flow researchers could ask participants how much time they think it took to complete the assessment, and compare it to actual time passed to serve as an assessment of attentional focus.

The third limitation was the difference between my GBA and non-GBA. I worked with an existing GBA because it was more realistic than one I would be able to create, but the compromise was that the constructs it evaluated were not completely mirrored in the non-GBA (meaning I did not find as strong of convergent validity as I would have liked). Small differences in what was measured could have contaminated some of my findings by introducing variance

that was not accounted for. Another difference between the assessments was the platform of administration- one could speculate that the difference in mobile vs. computerized administration accounted for some of my results. However in the pilot, I collected an attitude toward mobile testing measure to understand differences in participants' preferences, and found that there was a slight preference toward computerized assessments. Despite this preference, the GBA was able to induce flow and a host of other positive reactions even though it was administered on the less-endorsed of the two platforms. This may indicate that the difference in platform had a minimal effect. Regardless, in future studies that are comparing assessment methods, researchers should strive to reduce these contaminants as much as possible. This may mean building assessments from scratch. Future studies should also consider using diverse GBAs; here I used one GBA but each GBA differs hugely, from what constructs it is measuring to what game elements it incorporates. This makes GBA research difficult, as each tool within one method can look and feel so different from the next. The goal should be to expand this repertoire of studies, and see if similar trends emerge across GBAs.

One consideration is that gamification and game-based tool research is increasingly moving toward studying game elements as opposed to gamed tools. By isolating game elements and their underlying motivational properties and effects on behavior, researchers and practitioners alike gain more information on how and under what circumstances particular game elements should be leveraged (Mekler, Bruhlmann, Tuch, & Opwis, 2017; Seaborn & Fels, 2015). In this study, I was not able to do that as 1) I wanted to pit two assessment methods against each other as a foundation for more nuanced examinations and 2) I was interested in applicants' reactions to a GBA they would be more likely to encounter in a real setting, meaning a GBA that incorporates multiple game elements for full effect. However, there is value in

understanding how game elements affect reactions, especially relating to which elements are better able to induce flow. One of the learnings from the measure I used for awareness of game elements is diverse elements may not "fall together" neatly in one measure, so researchers should be mindful of assessing elements in a way that is psychometrically sound.

Theoretical Implications

There are three primary theoretical implications from this study. The first is introducing flow into the assessment literature to understand the unique state that some assessment methods elicit. Flow, while widely studied in psychology and other literatures, has rarely been used to understand the hiring process. This study serves as an initial explanation of how GBAs are able to captivate audiences: flow serves as one of the mediating mechanism through which GBAs elicit positive applicant reactions. Future work should continue to look at the role of flow in different assessment methods, especially as assessments move toward increasingly high-fidelity technology like augmented or virtual reality. Researchers could also look at how much equity flow can buy an assessment; here I hypothesized in my overall model that if a GBA could induce flow, the positive feelings or high fidelity of this state could result in increased perceived job-relatedness. It would be interesting to see how far this extends. If applicants are put in an enthralling and hyper-realistic scenario, will they have positive reactions regardless of what the test seems to be measuring? Lastly, more work should be done to identify the specific game elements that are able to induce flow so that assessment designers can employ elements in meaningful ways.

The next theoretical contribution of this study is successfully situating GBAs in an applicant reactions framework, something that has been suggested by other researchers in this space (Armstrong, Landers, and Collmus, 2016). Here I replicated relationships that are well-

established in the literature to compare the two assessment methods. However, because the reactions contained in this study are by no means comprehensive, it serves as only a starting point for examining how GBAs affect applicant's perceptions. Future research can explore a host of other concepts in this nomological net (Hausknecht et al., 2004). Do GBAs fare better if they are used within particular industries or for types of jobs? Does it matter to applicants where in the selection process a GBA is placed? How do GBAs affect withdrawal, reapplication, or recommendation intentions? These and many other questions are ripe for exploration.

Lastly, this study opens up questions about the kinds of individual differences and attitudes that are important in a GBA context. Technology self-efficacy was a significant moderator, but openness to experience and preference for games were not. As I mentioned in the summary of findings, it may matter less whether an applicant is open to new experiences in games *in general* and more how appropriate they feel it is to apply this level of novelty in a high stakes situation like hiring. Put differently, the moderating effect may actually be the environment itself. Individual differences and attitudes have been shown to affect applicant reactions (Ryan & Ployhart, 2000) and have been found to have an effect in the gamed instruction literature (Landers & Armstrong, 2015), so more should be done to understanding how and under what conditions they come into play for GBAs.

Practical Implications

The need to engage applicants during the hiring process is more pressing than ever (Ployhart, Schmitt, & Tippins, 2017), reflected in the fact that organizations and applicants continue to ask for entertaining or absorbing tools (Meister, 2013; Zielinski, 2015). One of the biggest practical implications of this study is empirical and theoretically-grounded evidence that game-based tools can elicit a fundamentally different reaction than their traditional counterparts.

While vendors have operated under this assumption and have conducted their own internal studies to support this point, there has been little objective or transparent research that organizations can draw on to inform their decision-making.

The finding that GBAs induce higher rates of flow has a few positive implications: organizations who are interested in incorporating novel tools into their process to attract young or innovative talent may want to use GBAs, as long as they are as psychometrically sound and legally-defensible as the non-GBA options. GBAs are often ideally suited to be administered on a smartphone or tablet; while this is good for ease of access, mobile assessments are more likely to be taken in distracting environments (Chang, Lawrence, Kinney, & O'Connell, 2016). If the assessment itself induces an absorptive state, the applicant may be better able to tune out external disruptions, and possibly perform better than if he or she was disengaged. Further, some work has shown initial promise that gamified tools may reduce test-taking anxiety (Collmus et al., 2016; Smits & Charlier, 2011). GBAs are optimally positioned to assess ability (because many existing cognitive tasks are already game-like), so using this method for anxiety-inducing constructs such as mathematical reasoning could allow for purer measurement across individuals. However, research should be conducted to verify these assumptions and see if GBAs can be used to minimize the effects of distraction and anxiety.

Through the mediating effect of flow, GBAs were found to relate to a number of applicant reactions. When applicants are in flow, they are more likely to see the tool as job-related. This is helpful information, as cognitive ability and personality assessments have suffered from poor ratings of job relevance relative to other tools (Pulakos, 2005). If organizations are able to leverage this finding, these predictive constructs can be measured in a less cumbersome way for the applicant. More work is needed to verify that the relationship

between flow and perceived job-relatedness is positive in other GBAs, as well as the parameters to this relationship.

The positive links from the GBA to PO fit and organizational attractiveness indicate that the positive feelings from flow can carry over to the organization and the applicant's feelings of oneness with it. Therefore organizations that want to boost their brand can use a GBA to do so, but must keep in mind that some components of the GBA may attract certain applicants and deter others. In my study, the average age of the student sample was 20.53 years and of my MTurk sample, 36.32 years. This is a fairly young group, and one could hypothesize that they would find the mobile technology, stimulating graphics, and gamified feel of the assessment attractive. However, my exploratory analyses found that age made little difference in perceived awareness of game elements, or the experience of flow. Additionally, the student sample had lower levels of flow, justice perceptions, organizational attractiveness, and acceptance intentions than the MTurk sample. This is promising and indicates game elements may be useful more broadly than with just a young audience, but organizations need to confirm that each tool they use is attractive to the desired candidate population.

Lastly, the moderator results offer a word of caution to practitioners who want to use GBAs. I found that technology self-efficacy moderated the relationship between assessment method and flow, especially for GBAs. This means not all applicants will respond to the GBAs the same way. Other work has supported this, by showing that attitudes towards game-based learning and experience with video games have implications for attitudes towards gamified instruction (Landers & Armstrong, 2015). Organizations should pay attention to their applicant pool's comfort with technology, and implement GBAs with caution to avoid a hiring process in which applicants feel they are prevented from demonstrating their skills because the mechanics

of the assessment are too daunting. There has not been any work showing that there are segments of the population who have consistently unfavorable reactions to gamed technology, but more work needs to be done to understand if there are certain demographics with which GBAs have limited utility.

Conclusion

This study integrates work from several fields to offer a theoretical framework for the effect of GBAs on applicant reactions. It offers evidence for the idea that GBAs can induce positive applicant reactions, in part through the effect of a flow state where the applicant's attention and interest are focused on the assessment. GBAs, through the flow state, positively relate to perceived job-relatedness, justice perceptions, perceived PO fit, and organizational attractiveness. Importantly, they can also play an influential role in persuading applicants to accept offers from organization. I provide support for the idea that GBAs as an assessment method offer a unique applicant experience, opening the door to future studies that may explore how and under what conditions researchers and practitioners can leverage these novel tools.

APPENDICES

Appendix A

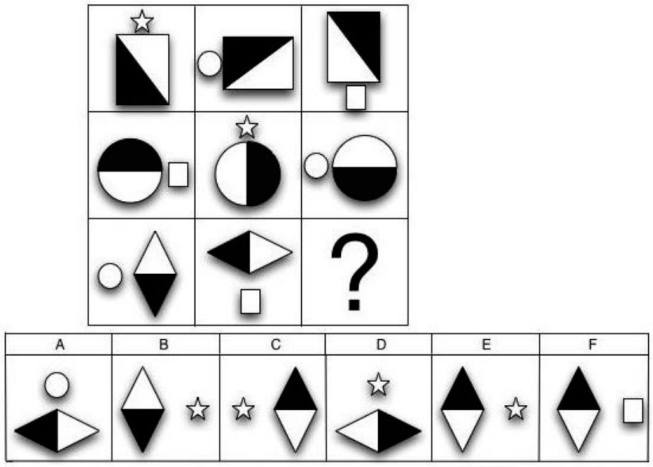
Non-GBA Measures

Cognitive ability sample items (ICAR, 2014)

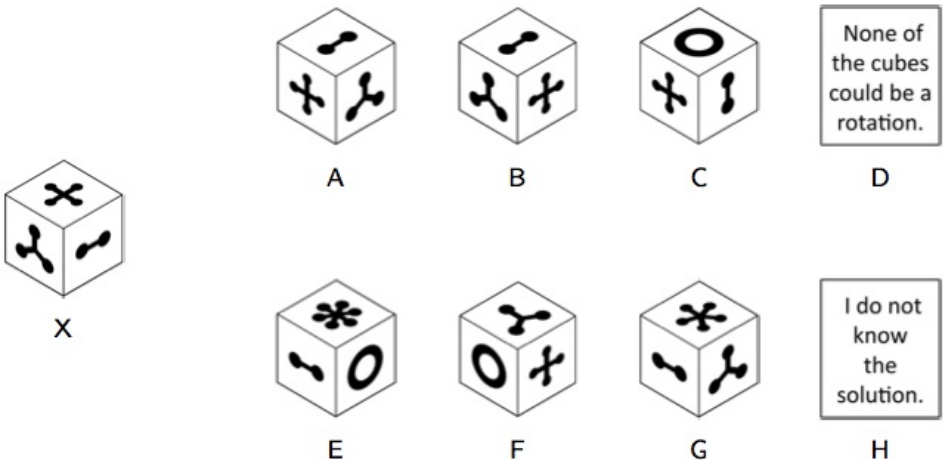
Letter and number series

In the following number series, what number comes next? 64, 81, 100, 121, 144, ...
(1) 154 (2) 156 (3) 162 (4) 169 (5) 178 (6) 196 (7) None of these (8) I don't know

Matrix reasoning



3D Rotation



Digit span

592

Please enter the numbers you just saw, in order. _____

Connor-Davidson Resilience Scale (Connor & Davidson, 2003) – *on a 1 (Rarely true) to 5 (True nearly all of the time) scale*

1. I am able to adapt to change.
2. I like close and secure relationships.
3. I can deal with whatever comes.
4. Past success gives me confidence for new challenge.
5. I see the humorous side of things.
6. Coping with stress strengthens me.
7. I tend to bounce back after illness or hardship.
8. Things happen for a reason.
9. I give my best effort no matter what.
10. I can achieve your goals.
11. When things look hopeless, I do not give up.
12. I know where to turn for help.
13. Under pressure, I focus and think clearly.
14. I prefer to take the lead in problem solving.
15. I am not easily discouraged by failure.
16. I think of myself as strong person.
17. I make unpopular or difficult decisions.
18. I can handle unpleasant feelings.
19. I have to act on a hunch.
20. I have a strong sense of purpose.
21. I am in control of my life.
22. I like challenges.
23. I work to attain my goals.
24. I take pride in my achievements.

IPIP-VIA Industry/Perseverance/Persistence (Peterson & Seligman, 2004) - *on a 1 (Very inaccurate) to 5 (Very accurate) scale*

1. I do not quit a task before it is finished.
2. I am a goal-oriented person.
3. I finish things despite obstacles in the way.
4. I am a hard worker.
5. I do not get sidetracked when I work.
6. I do not finish what I start. (R)
7. I give up easily. (R)

8. I do not tend to stick with what I decide to do. (R)

IPIP Social Confidence (Jackson, 2004) – *on a 1 (Very inaccurate) to 5 (Very accurate) scale*

1. I feel comfortable around people.
2. I do not mind being the center of attention.
3. I am good at making impromptu speeches.
4. I express myself easily.
5. I have a natural talent for influencing people.
6. I hate being the center of attention. (R)
7. I lack the talent for influencing people. (R)
8. I often feel uncomfortable around others. (R)
9. I do not like to draw attention to myself. (R)
10. I have little to say. (R)

Bomb Risk Elicitation Task (Crosetto & Filippin, 2012)

> **Instructions**: On a sheet of paper on your desk you see a square composed of 100 numbered boxes. Behind one of these boxes hides a mine; all the other 99 boxes are free from mines. You do not know where this mine lies. You only know that the mine can be in any place with equal probability.
>
> Your task is to choose how many boxes to collect. Boxes will be collected in numerical order. So you will be asked to choose a number between 1 and 100. At the end of the experiment we will randomly determine the number of the box containing the mine by means of a bag containing 100 numbered tokens.
>
> If you happen to have harvested the box where the mine is located – i.e. if your chosen number is greater than or equal to the drawn number – you will earn zero. If the mine is located in a box that you did not harvest – i.e. if your chosen number is smaller than the drawn number – you will earn in euro an amount equivalent to the number you have chosen divided by ten.
>
> In the next screen we will ask you to indicate how many boxes you would like to collect. You confirm your choice by hitting 'OK'.

1	2	3	4	5	6	7	8	9	10
11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30
31	32	33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48	49	50
51	52	53	54	55	56	57	58	59	60
61	62	63	64	65	66	67	68	69	70
71	72	73	74	75	76	77	78	79	80
81	82	83	84	85	86	87	88	89	90
91	92	93	94	95	96	97	98	99	100

General Risk (Nye, personal communication)- *on a 1 (Strong disagree) to 5 (Strongly agree) scale*

1. Taking risks makes life more fun.
2. My friends would say that I am a risk-taker.
3. I enjoy taking risks in most aspects of my life.
4. I would take a risk even if it meant I might get hurt.
5. Taking risks is an important part of my life.
6. In general I avoid taking risks. (R)
7. I generally like to "play it safe". (R)
8. I commonly make risky decisions.
9. I am a believer of taking chances.
10. When taking a chance, I focus more on winning than on possibly losing.
11. I am attracted, rather than scared, by risk.
12. I generally avoid risky situations. (R)
13. I focus more on the positive outcomes of risk, rather than negative ones.
14. You never get anywhere without taking chances.

Propensity to Innovate (Burningham & West, 1995)- *on a 1 (Strong disagree) to 5 (Strongly agree) scale*

1. I try to introduce improved methods of doing things at work.
2. I have ideas which would significantly improve the way the job is done.
3. If there is a problem or challenge at work I will ignore the rules in order to find a new solution.
4. I tend to change existing policies relating to my area of work.
5. I suggest new working methods to the people I work with.
6. I try to avoid introducing changes for change's sake. (R)
7. I tend to improve methods for solving problems when the answer is not apparent.
8. I change the specification of what my job entails so as to improve my functioning in it.
9. I use different methods than other individuals performing the same type of job.
10. I look for novel approaches to dealing with my work.
11. I contribute to changes in the way my department works.
12. I am receptive to new ideas which I can use to improve things at work.

Appendix B

Pilot Measures

Measures below were included in pilot but not retained for the full study due to low correlations with the GBA constructs.

Brief Resilience Scale (Smith et al., 2008) - *on a 1 (Strong disagree) to 5 (Strongly agree) scale*

1. I tend to bounce back quickly after hard times.
2. I have a hard time making it through stressful events. (R)
3. It does not take me long to recover from a stressful event.
4. It is hard for me to snap back when something bad happens. (R)
5. I usually come through difficult times with little trouble.
6. I tend to take a long time to get over set-backs in my life. (R)

IPIP Sociability - *on a 1 (Very inaccurate) to 5 (Very accurate) scale*

1. Usually like to spend my free time with people.
2. Talk to a lot of different people at parties.
3. Love to chat.
4. Make friends easily.
5. Enjoy being part of a group.
6. Seem to derive less enjoyment from interacting with people than others do. (R)
7. Rarely enjoy being with people.(R)
8. Would not enjoy a job that involves a lot of social interaction. (R)
9. Am hard to get to know. (R)
10. Keep others at a distance. (R)

Affiliative Tendency- *on a 1 (Very inaccurate) to 5 (Very accurate) scale*

1. When I am introduced to someone new, I don't make much effort to be liked.
2. I prefer a leader who is friendly and easy to talk to over one who is more aloof and respected by his or her followers.
3. When I am not feeling well, I would rather be with others than alone.
4. If I had to choose between the two, I would rather be considered intelligent than sociable.
5. Having friends is very important to me.
6. I would rather express open appreciation to others most of the time than reserve such feelings for special occasions.
7. I enjoy a good movie more than a big party.
8. I like to make as many friends as I can.
9. I would rather travel abroad starting my trip alone than with one or two friends.
10. After I meet someone I did not get along with, I spend time thinking about arranging another, more pleasant meeting.
11. I think that fame is more rewarding than friendship.
12. I think that any experience is more significant when shared with a friend.

13. When I see someone I know walking down the street, I am usually the first one to say hello.
14. I prefer the independence that comes from lack of attachments to the good and warm feelings associated with close ties.
15. I join clubs because it is such a good way of making friends.
16. I would rather serve in a position to which my friends had nominated me than be appointed to an office by a distant national headquarters.
17. I don't believe in showing overt affection toward friends.
18. I would rather go right to sleep at night than talk to someone else about the days' activities.
19. I have very few close friends.
20. When I am with people I don't know, it doesn't matter much to me if they like me or not.
21. If I had to choose, I would rather have strong attachments to my friends than have them regard me as witty and clever.
22. I prefer individual activities such as crossword puzzles to group ones such as bridge or canasta.
23. I am much more attracted to warm, open people than I am to stand-offish ones.
24. I would rather read an interesting book or go to the movies than spend time with friends.
25. When traveling, I prefer meeting people to simply enjoying the scenery or going places alone.

Multiple Stimulus Types Ambiguity Tolerance Scale-II (McLain, 2009) - *on a 1 (Strong disagree) to 5 (Strongly agree) scale*

1. I don't tolerate ambiguous situations well. (R)
2. I would rather avoid solving a problem that must be viewed from several different perspectives. (R)
3. I try to avoid situations that are ambiguous. (R)
4. I prefer familiar situations to new ones. (R)
5. Problems that cannot be considered from just one point of view are a little threatening. (R)
6. I avoid situations that are too complicated for me to easily understand. (R)
7. I am tolerant of ambiguous situations.
8. I enjoy tackling problems that are complex enough to be ambiguous.
9. I try to avoid problems that don't seem to have only one "best" solution. (R)
10. I generally prefer novelty over familiarity.
11. I dislike ambiguous situations. (R)
12. I find it hard to make a choice when the outcome is uncertain. (R)
13. I prefer a situation in which there is some ambiguity.

Creative Personality Scale (Gough, 1979) – *order will be randomized, participants will be asked to check items with which they identify*

1. Capable (P)
2. Clever (P)
3. Confident (P)

4. Egotistical (P)
5. Humorous (P)
6. Individualistic (P)
7. Informal (P)
8. Insightful (P)
9. Intelligent (P)
10. Interests wide (P)
11. Inventive (P)
12. Original (P)
13. Reflective (P)
14. Resourceful (P)
15. Self-confident (P)
16. Sexy (P)
17. Snobbish (P)
18. Unconventional (P)
19. Affected (N)
20. Cautious (N)
21. Commonplace (N)
22. Conservative (N)
23. Conventional (N)
24. Dissatisfied (N)
25. Honest (N)
26. Interests narrow (N)
27. Mannerly (N)
28. Sincere (N)
29. Submission(N)
30. Suspicious (N)

The following are not part of the assessment manipulation, but are reactions and individual differences measures that were used **ONLY** in the pilot study.

Attitudes toward Mobile Assessment (King, Ryan, Kantrowitz, Grelle, & Dainis, 2015) - *on a 1 (Strongly disagree) to 5 (Strongly agree) scale other than item 1*

1. In general, how would you compare completing a test on a mobile device versus a computer.
    a. Better on a mobile device than on a computer
    b. As good on a mobile device as on a computer
    c. Worse on a mobile device than on a computer
    d. No opinion
2. I would prefer to complete tests on a mobile device versus completing them on a computer.
3. I would be more likely to apply for a job at a company that allowed me to complete a 'pre-employment' test on my mobile device, such as a smart phone, tablet, or iPad versus a company that allowed taking the same test only on a computer.
4. It is equally fair to use a test given on a mobile device as it is to use the same one given on a computer to make a hiring or promotion decision for a job.
5. I believe a company that allows me to take its test on my mobile device would be a better place to work compared to a company that only allows its test to be taken on a computer.
6. I would be more likely to accept a job offer from a company that allows me to take its test on my mobile device versus a company that only allows its test to be taken on a computer.
7. Having the option to complete this assessment on a mobile device positively represents the company's brand image.

Appendix C

Main Study Measures

Unless otherwise notes, responses will be scored on a 1 (Strongly disagree) to 5 (Strongly agree) scale

Belief in Tests (Arvey et al., 1990)

1. Employment selection tests are a good way of selecting people into jobs.
2. Employment selection tests should be eliminated. (R)
3. I don't believe that employment selection tests are valid. (R)

Flow (Webster et al., 1993)

Control

1. When taking the assessment, I felt in control.
2. I felt that I had no control over my interaction with the assessment. (R)
3. The assessment allowed to me to control the interaction.

Attention Focus

4. When taking the assessment, I thought about other things. (R)
5. When taking the assessment, I was aware of distraction. (R)
6. When taking the assessment, I was totally absorbed in what I was doing.

Curiosity

7. Taking the assessment excited my curiosity.
8. Interacting with the assessment made me curious.
9. Using the assessment aroused my imagination.

Intrinsic Interest

10. Taking the assessment bored me. (R)
11. Taking the assessment was intrinsically interesting.
12. Taking the assessment was fun for me.

Technology Self-Efficacy (Shortened from Cassidy & Eachus, 2002)

Computers

1. I find working with computers very easy.
2. I am very confident in my abilities with computers.
3. I find it difficult to get computers to do what I want. (R)
4. I usually find it easy to learn how to use a new program on a computer.
5. I seem to waste a lot of time struggling with computers. (R)
6. As far as computers go, I don't consider myself to be very competent. (R)
7. Computers help me save a lot of time.
8. I find working with computers very frustrating. (R)

Smartphones

9. I find working smartphone apps very easy.
10. I am very confident in my abilities to make use of smartphone apps.
11. I find it difficult to get smartphone apps to do what I want. (R)

12. I usually find it easy to learn how to use a new smartphone app.
13. I seem to waste a lot of time struggling with smartphone apps. (R)
14. As far as smartphone apps go, I don't consider myself to be very competent. (R)
15. Smartphone apps help me save a lot of time.
16. I find working with smartphone apps very frustrating. (R)

Preference for Games (created for this study)

1. I would be eager to use a new tool or system to which game elements have been added.
2. It is unnecessary to make everyday tasks or actions feel like a game. (R)
3. I prefer to feel like I am playing a game when I complete ordinary tasks.
4. The trend toward "gamification" is a positive one.
5. I do not like it when game elements are added to existing tools or systems. (R)
6. Adding game elements to tasks makes them more enjoyable.

Job-relatedness (Smither et al., 1993)

Face validity

1. I did not understand what the assessment had to do with the job. (R)
2. I could not see any relationship between the assessment and what is required on the job. (R)
3. It would be obvious to anyone that the assessment is related to the job.
4. The actual content of the assessment was clearly related to the job.
5. There was no real connection between the assessment that I went through and the job. (R)

Predictive validity

6. Failing to pass the assessment clearly indicates that you can't do the job.
7. I am confident that the assessment can predict how well an applicant will perform on the job.
8. My performance on the assessment was a good indicator of my ability to do the job.
9. Applicants who perform well on this kind of assessment are more likely to perform well on the job than applicants who perform poorly.
10. The employer can tell a lot about the applicant's ability to do the job from the results of the assessment.

Perceived PO Fit (Cable & Judge, 1996) – *on a 1 (Not at all) to 5 (Completely) scale*

Supplementary

Instructions: Please think about your current values and personality. This would be the values and personality you hold today. Now answer the questions below according to your current self.

1. To what degree do you feel your current values match or fit the organization for which you are applying for this job?
2. My current values match those of current employees in the organization in which I am applying for this job?
3. Do you think the values and personality of the hiring organization reflect your own current values and personality?

Complementary

Instructions: Please think about your preferred values and personality. This would be the values and personality you would have in an idealized world; the "you" that you strive to be. Now answer the questions below according to your preferred self.

1. To what degree do you feel your preferred values match or fit the organization for which you are applying for this job?
2. My preferred values match those of current employees in the organization in which I am applying for this job.
3. Do you think the values and personality of this organization reflect your own preferred values and personality?

Justice Perceptions (Smither et al., 1993)

Procedural

Instructions: The following items refer to the procedures used to decide if you get an offer from the company.

1. Overall, I believe that the assessment process was fair.
2. I felt good about the way the assessment was conducted and administered.

Distributive

Instructions: The following items refer to your offer from the company.

3. The offer decision accurately reflected how well I performed on the assessment.
4. I deserved the offer decision that I received on the assessment.
5. The assessment fairly reflected my ability to do the job.

Organizational Attractiveness (Highhouse, Lievens, & Sinar, 2003)

1. For me, this company would be a good place to work.
2. I would not be interested in this company except as a last resort. (R)
3. This company is attractive to me as a place for employment.
4. I am interested in learning more about this company.
5. A job at this company is very appealing to me.

Acceptance intentions (Ployhart & Ryan, 1998)

Instructions: For the following item, please do not think about other constraints you would normally consider when accepting a job (i.e. labor market, salary, location, benefits). Simply reflect on the interactions you have had with the organization, and base your decision off them.

Hired

1. I would accept the job at this organization.

Not hired

2. Even if I was now offered the job, I would not accept it. (R)

Awareness of Game Elements

Focused rules/goals

1. During the assessment, there were clear goals I needed to meet to be successful.
2. During the assessment, there were clear rules I needed to follow to be successful.

Feedback

3. The assessment provided me with feedback while I was taking it.
4. I did not have an idea of how well I was performing on the assessment while taking it. (R)

Challenging tasks

5. The assessment provided me with challenging tasks and situations that I had to overcome.
6. I had to perform well in the face of challenge in order to complete the assessment.

Sensory stimuli

7. The "look" of the assessment gamified.
8. I felt like I was playing a game while I took the assessment.

Alternative Explanation Checks

Time

1. This assessment took a reasonable amount of time.
2. I felt like the assessment was too long. (R)

Ease of faking (Gilliland & Honig, 1994)

3. I think some people would distort their responses during the selection process to try to make themselves look better.
4. It would be easy for people to be dishonest when answering questions and make themselves look good.
5. I thought you could beat the tests if you were smart and gave the answers they were looking for.
6. It was obvious how you should respond to some of the questions if you wanted the job.

Opportunity to perform (Bauer et al., 2001)

7. I could really show my skills and abilities through this test.
8. This test allowed me to show what my job skills are.
9. This test gives applicants the opportunity to show what they can really do.
10. I was able to show what I can do on this test.

Perceived performance (Smither et al., 1993)

11. After I finished the assessment it was clear to me how well I performed.
12. I knew exactly what aspects of the assessment I performed well and poorly.
13. Anyone who went through the assessment would know clearly how well or poorly they did.

Demographics

1. What is your age (in years)? _____
2. What is your gender?
   - Male
   - Female
   - Transgender

- Other _____
- Prefer not to say

3. Is your ethnicity Hispanic/Latino?
   - Yes
   - No
4. Please select your race below (may select more than one):
   - American Indian or Alaska Native
   - Asian
   - Black or African American
   - Native Hawaiian or Other Pacific Islander
   - White
   - Other
5. How many years of work experience do you have? (Part-time work counts) _____
6. How would you rate your job performance in the job you currently hold relative to others who are performing the same job? *(MTurk only)*
   - I could use improvement
   - 2
   - 3
   - I am a satisfactory performer
   - 5
   - 6
   - I am a top performer
7. How quickly do you learn new tasks or assignments relative to other employees? *(MTurk only)*
   - It takes me a little longer than most employees
   - 2
   - 3
   - I learn at the same pace as most other employees
   - 5
   - 6
   - I learn more quickly than most other employees
8. How many times have you used or interacted with a game-based assessment (i.e. assessments that are game-like or include game elements such as points systems, badges, or leaderboards)?
   - No experience
   - 0-1 times
   - 2-5 times
   - 6-10 times
   - 10+ times

9. What is your current cumulative college GPA? *(Student only)*
   - less than 2.00
   - 2.00 to 2.29
   - 2.30 to 2.59

- 2.60 to 2.89
- 2.90 to 3.19
- 3.20 to 3.49
- 3.50 to 3.79
- 3.80 to 3.99
- 4.00
- I am in my first semester

10. What is your intended major? _____ *(Student only)*
11. In which of the following industries do you work? *(MTurk only)*
    - Retired
    - Unemployed
    - Office and Administrative Support
    - Sales and Related
    - Architecture and Engineering
    - Arts, Design, Entertainment, Sports, and Media
    - Legal
    - Construction and Extraction
    - Installation, Maintenance, and Repair
    - Building and Grounds Cleaning and Maintenance
    - Production
    - Business and Financial Operations
    - Personal Care and Service
    - Healthcare Support
    - Protective Service
    - Food Preparation and Serving Related
    - Computer and Mathematical
    - Farming, Fishing, and Forestry
    - Management
    - Healthcare Practitioners and Technical
    - Life, Physical, and Social Science
    - Education, Training, and Library
    - Community and Social Service
    - Transportation and Materials Moving
    - Other (please specify)
12. How many times have you participated in the hiring process? *(MTurk only)*
    - No experience
    - 0-1 times
    - 2-5 times
    - 6-10 times
    - 10+ times

REFERENCES

REFERENCES

Agarwal, R., & Karahanna, E. (2000). Time flies when you're having fun: Cognitive absorption and beliefs about information technology usage. *MIS Quarterly, 24*(4), 665-694.

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*(2), 179-211.

Alicke, M. D. (1985). Global self-evaluation as determined by the desirability and controllability of trait adjectives. *Journal of Personality and Social Psychology, 49*(6), 1621-1630.

Altman, I. (2015, Dec 1). Top 10 business trends that will drive success in 2016. *Forbes*. Retrieved from http://www.forbes.com/sites/ianaltman/2015/12/01/top-10-business-trends-that-will-drive-success-in-2016/#2f5529355719

Ambrose, M. L., Arnaud, A., & Schminke, M. (2007). Individual moral development and ethical climate: The influence of person–organization fit on job attitudes. *Journal of Business Ethics*, *77*(3), 323–333.

Anderson, J. Q., & Rainie, H. (2012). Gamification: Experts expect 'game layers' to expand in the future, with positive and negative results. *Pew Internet & American Life Project*. Retrieved from http://www.pewinternet.org/files/old-media/Files/Reports/2012/PIP_Future_of_Internet_2012_Gamification.pdf

Anderson, N. (2003). Applicant and recruiter reactions to new technology in selection: A critical review and agenda for future research. *International Journal of Selection and Assessment, 11*(2-3), 121-136.

Armstrong, M. B., Landers, R. N., & Collmus, A. B. (2015). Gamifying recruitment, selection, training, and performance management: Game-thinking in human resource management. In D. Davis & H. Gangadharbatla (Eds.), *Handbook of research on trends in gamification* (pp.140-165). Hershey, PA: Information Science Reference.

Arthur Jr, W., & 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*(2), 435.

Artic Shores (n.d.). *People insights from game technology*. Retrieved from https://www.arcticshores.com/

Arvey, R. D., Strickland, W., Drauden, G., & Martin, C. (1990). Motivational components of test taking. *Personnel Psychology*, *43*(4), 695-711.

Ashton, M. C., Lee, K., & Goldberg, L. R. (2007). The IPIP–HEXACO scales: An alternative, public-domain measure of the personality constructs in the HEXACO model. *Personality and Individual Differences*, *42*(8), 1515-1526.

Bakker, A. B. (2008). The work-related flow inventory: Construction and initial validation of the WOLF. *Journal of Vocational Behavior, 72*(3), 400–414.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191-215.

Barbeite, F. G., & Weiss, E. M. (2004). Computer self-efficacy and anxiety scales for an Internet sample: testing measurement equivalence of existing measures and development of new scales. *Computers in Human Behavior*, *20*(1), 1-15.

Bauer, T. N., Maertz Jr, C. P., Dolen, M. R., & Campion, M. A. (1998). Longitudinal assessment of applicant reactions to employment testing and test outcome feedback. *Journal of Applied Psychology*, *83*(6), 892-903.

Bauer, T. N., Truxillo, D. M., Paronto, M. E., Weekley, J. A., & Campion, M. A. (2004). Applicant reactions to different selection technology: Face-to-face, interactive voice response, and computer-assisted telephone screening interviews. *International Journal of Selection and Assessment, 12*(1-2), 135-148.

Bauer, T. N., Truxillo, D. M., Sanchez, R. J., Craig, J. M., Ferrara, P., & Campion, M. A. (2001). Applicant reactions to selection: Development of the selection procedural justice scale (SPJS). *Personnel Psychology*, *54*(2), 387–419.

Baumeister, R. F., Muraven, M., & Tice, D. M. (2000). Ego depletion: A resource model of volition, self-regulation, and controlled processing. *Social Cognition, 18*(2), 130-150.

Bedwell, W. L., Pavlas, D., Heyne, K., Lazzara, E. H., & Salas, E. (2012). Toward a taxonomy linking game attributes to learning: An empirical study. *Simulation & Gaming, 43*(6), 729–760.

Bernerth, J. B., Feild, H. S., Giles, W. F., & Cole, M. S. (2006). Perceived fairness in employee selection: The role of applicant personality. *Journal of Business and Psychology*, *20*(4), 545-563.

Bies, R. J., & Moag, J. S. (1986). Interactional justice: Communication criteria of fairness. *Research on Negotiation in Organizations, 1*, 43-55.

Bittner, J. V., & Schipper, J. (2014). Motivational effects and age differences of gamification in product advertising. *Journal of Consumer Marketing, 31*(5), 391-400.

Blieszner, R., Willis, S. L., & Baltes, P. B. (1981). Training research in aging on the fluid ability of inductive reasoning. *Journal of Applied Developmental Psychology, 2*(3), 247-265.

Bond, M., Leung, K. & Wan, K. (1982). How does cultural collectivism operate? The impact of task and maintenance contributions on reward distribution. *Journal of Cross-Cultural Psychology 13,* 186-200.

Borkenau, P. & Ostendorf, F. (1993). *NEO-Fünf-Faktoren Inventar (NEO-FFI) nach Costa und McCrae* (NEO-five-factor inventory following Costa and McCrae). Göttingen: Hogrefe.

Bowman, R. F. (1982). A "Pac-Man" theory of motivation: Tactical implications for classroom instruction. *Educational Technology, 22*(9), 14-17.

Bretz, R. D., Jr., & Judge, T. A. (1994a). Person–organization fit and the theory of work adjustment: Implications for satisfaction, tenure, and career success. *Journal of Vocational Behavior*, *44*(1), 32–54.

Bretz, R. D., Jr., & Judge, T. A. (1994b). The role of human resource systems in job applicant decision processes. *Journal of Management*, *20*(3), 531–551.

Bretz, R. D., Jr., Ash, R. A., & Dreher, G. F. (1989). Do people make the place? An examination of the attraction-selection-attrition hypothesis. *Personnel Psychology*, *42*(3), 561-581.

Brown, K. W., & Ryan, R. M. (2003). The benefits of being present: Mindfulness and its role in psychological well-being. *Journal of Personality and Social Psychology, 84*(4), 822-848.

Bruk-Lee, V., Lanz, J., Drew, E. N., Coughlin, C., Levine, P., Tuzinski, K., & Wrenn, K. (2016). Examining applicant reactions to different media types in character-based simulations for employee selection. *International Journal of Selection and Assessment*, *24*(1), 77-91.

Bunce, D., & West, M. A. (1995). Self-perceptions and perceptions of group climate as predictors of individual innovation at work. *Applied Psychology, 44*(3), 199-215.

Bureau of Labor Statistics. (2016). Labor force statistics from the current population survey, *Bureau of Labor Statistics, Washington, D.C*. Retrieved from https://www.bls.gov/cps/cpsaat09.htm

Burningham, C., & West, M. A. (1995). Individual, climate, and group interaction processes as predictors of work team innovation. *Small Group Research, 26*(1), 106–117.

Byrne, D. (1971).*The attraction paradigm*. Orlando, FL: Academic Press.

Cable, D. M., & DeRue, D. S. (2002). The convergent and discriminant validity of subjective fit perceptions. *Journal of Applied Psychology*, *87*(5), 875–884.

Cable, D. M., & Edwards, J. R. (2004). Complementary and supplementary fit: A theoretical and empirical integration. *Journal of Applied Psychology*, *89*(5), 822–834.

Cable, D. M., & Judge, T. A. (1994). Pay preferences and job search decisions: A person-organization fit perspective. *Personnel Psychology*, *47*(2), 317–348.

Cable, D. M., & Judge, T. A. (1996). Person–organization fit, job choice decisions, and organizational entry. *Organizational Behavior and Human Decision Processes*, *67*(3), 294–311.

Cable, D. M., & Parsons, C. K. (2001). Socialization tactics and person-organization fit. *Personnel Psychology*, *54*(1), 1–23.

Caplan, R. D. (1987). Person-environment fit theory and organizations: Commensurate dimensions, time perspectives, and mechanisms. *Journal of Vocational Behavior*, *31*(3), 248-267.

Carless, S. A. (2005). Person–job fit versus person–organization fit as predictors of organizational attraction and job acceptance intentions: A longitudinal study. *Journal of Occupational and Organizational Psychology*, *78*(3), 411–429.

Cassidy, S., & Eachus, P. (2002). Developing the computer user self-efficacy (CUSE) scale: Investigating the relationship between computer self-efficacy, gender and experience with computers. *Journal of Educational Computing Research*, *26*(2), 133–153.

Chan, D. (1997). Racial subgroup differences in predictive validity perceptions on personality and cognitive ability tests. *Journal of Applied Psychology*, *82*(2), 311–320.

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.

Chan, D., & Schmitt, N. (2004). An agenda for future research on applicant reactions to selection procedures: A construct-oriented approach. *International Journal of Selection and Assessment*, *12*(1–2), 9–23.

Chan, D., Schmitt, N., DeShon, R. P., Clause, C. S., & Delbridge, K. (1997). Reactions to cognitive ability tests: The relationships between race, test performance, face validity perceptions, and test-taking motivation. *Journal of Applied Psychology*, *82*(2), 300–310.

Chan, D., Schmitt, N., Jennings, D., Clause, C. S., & Delbridge, K. (1998a). Applicant perceptions of test fairness: Integrating justice and self-serving bias perspectives. *International Journal of Selection and Assessment*, *6*(4), 232–239.

Chan, D., Schmitt, N., Sacco, J. M., & DeShon, R. P. (1998b). Understanding pretest and posttest reactions to cognitive ability and personality tests. *Journal of Applied Psychology*, *83*(3), 471–485.

Chang, L. C., Lawrence, A. D., Kinney, T. B. & O'Connell, M. S. (2016). *What was that? Investigating distractions on test performance and reactions.* Paper presented at the 31st annual conference of the Society for Industrial and Organizational Psychology, Anaheim, CA.

Chapman, D. S., Uggerslev, K. L., Carroll, S. A., Piasentin, K. A., & Jones, D. A. (2005). Applicant attraction to organizations and job choice: A meta-analytic review of the correlates of recruiting outcomes. *Journal of Applied Psychology*, *90*(5), 928–944.

Charness, G., & Gneezy, U. (2010). Portfolio choice and risk attitudes: An experiment. *Economic Inquiry, 48*(1), 133-146.

Charness, G., & Gneezy, U. (2011). Strong evidence for gender differences in risk taking. *Journal of Economic Behavior & Organization*, *83*(1), 50-58.

Chatman, J. A. (1989). Improving interactional organizational research: A model of person-organization fit. *Academy of Management Review*, *14*(3), 333–349.

Chatman, J. A. (1991). Matching people and organizations: Selection and socialization in public accounting firms. *Administrative Science Quarterly*, *36*(3), 459-484.

Chatman, J. A., & Jehn, K. A. (1994). Assessing the relationship between industry characteristics and organizational culture: How different can you be? *Academy of Management Journal*, *37*(3), 522–553.

Chen, J. (2007). Flow in games (and everything else). *Communications of the ACM, 50*(4), 31–34.

Chiang, Y. T., Lin, S. S. J., Cheng, C. Y., & Liu, E. Z. F. (2011). Exploring online game players' flow experiences and positive affect. *Turkish Online Journal of Educational Technology, 10*(1), 106–114.

Chin, J., Dukes, R., & Gamson, W. (2009). Assessment in simulation and gaming: A review of the last 40 years. *Simulation & Gaming, 40*(4), 553–568.

Choi, D. H., Kim, J., & Kim, S. H. (2007). ERP training with a web-based electronic learning system: The flow theory perspective. *International Journal of Human-Computer Studies, 65*(3), 223–243.

Chou, T. J., & Ting, C. C. (2003). The role of flow experience in cyber-game addiction. C*yberPsychology & Behavior, 6*(6), 663–675.

Chou, Y. K. (2015). *Actionable gamification: Beyond points, badges, and leaderboards*. Octalysis Media.

Chou, Y. K. (2017). A comprehensive list of 90+ gamification cases with ROI stats. *Yu-Kai Chou & Gamification*. Retrieved from http://www.yukaichou.com/gamification-examples/gamification-stats-figures/#.UzYtafldWVp

Chung, J. E., Park, N., Wang, H., Fulk, J., & McLaughlin, M. (2010). Age differences in perceptions of online community participation among non-users: An extension of the technology acceptance model. *Computers in Human Behavior, 26*(6), 1674–1684.

Codol, J. P. (1975). On the so-called 'superior conformity of the self' behavior: Twenty experimental investigations. *European Journal of Social Psychology, 5*(4), 457-501.

Cohen-Charash, Y., & Spector, P. E. (2001). The role of justice in organizations: A meta-analysis. *Organizational Behavior and Human Decision Processes, 86*(2), 278-321.

Collmus, A.B. (2016). *Game-framing cognitive assessments to improve applicant perceptions*. Retrieved from ODU Digital Commons Psychology Theses & Dissertations.

Colquitt, J. A., Conlon, D. E., Wesson, M. J., Porter, C. O., & Ng, K. Y. (2001). Justice at the millennium: A meta-analytic review of 25 years of organizational justice research. *Journal of Applied Psychology*, *86*(3), 425-445.

Colquitt, J. A., Hollenbeck, J. R., Ilgen, D. R., LePine, J. A., & Sheppard, L. (2002). Computer-assisted communication and team decision-making performance: The moderating effect of openness to experience. *Journal of Applied Psychology, 87*(2), 402-410.

Connelly, B. L., Certo, S. T., Ireland, R. D., & Reutzel, C. R. (2011). Signaling theory: A review and assessment. *Journal of Management*, *37*(1), 39–67.

Connor, K. M., & Davidson, J. R. T. (2003). Development of a new resilience scale: The Connor-Davidson Resilience Scale (CD-RISC). *Depression and Anxiety, 18*(2), 76–82.

Cook, K. S., & Hegtvedt, K. A. (1983). Distributive justice, equity, and equality. *Annual Review of Sociology*, *9*(1), 217-241.

Costa, P. T. & McCrae, R. R. (1985). *The NEO Personality Inventory*. Odessa, FL: Psychological Assessment Resources.

Costa, P. T., & McCrae, R. R. (1992). Normal personality assessment in clinical practice: The NEO Personality Inventory. *Psychological Assessment, 4*(1), 5-13.

Cowley, B., Charles, D., Black, M., & Hickey, R. (2008). Toward an understanding of flow in video games. *Computers in Entertainment, 6*(2), 1–27.

Crawford, C. (1984). *The art of computer design*. Berkeley, CA: Osborne/McGraw-Hill.

Cropanzano, R., & Ambrose, M. L. (2001). Procedural and distributive justice are more similar than you think: A monistic perspective and a research agenda. In J. Greenberg & R. Cropanzano (Eds.), *Advances in organizational justice* (pp. 119–151). Stanford, CA: Stanford University Press.

Csikszentmihalyi, M. (1975a). *Beyond boredom and anxiety*. San Francisco, CA: Jossey-Bass.

Csikszentmihalyi, M. (1975b). Play and intrinsic rewards. *Journal of Humanistic Psychology, 15*(3), 41–63.

Csikszentmihalyi, M. (1979). Intrinsic rewards and emergent motivation. In M. R. Lepper & D. Greene (Eds). *The hidden costs of reward* (pp. 205-216). Morristown, NJ: Erlbaum.

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. New York: Harper & Row.

Csikszentmihalyi, M., & Larson, R. (1978). Intrinsic rewards in school crime. *Crime and Delinquency, 24*, 322-335.

Csikszentmihalyi, M., & LeFevre, J. (1989). Optimal experience in work and leisure. *Journal of Personality and Social Psychology*, *56*(5), 815-822.

Daft, R. L., & Lengel, R. H. (1986). Organizational information requirements, media richness and structural design. *Management Science, 32*(5), 554-571.

Dawis, R. V., & Lofquist, L. H. (1984). *A psychological theory of work adjustment: An individual-differences model and its applications*. Minneapolis, MN: University of Minnesota Press.

DeSilver, D. (2017, Jan 4). 5 facts about the minimum wage. *Pew Research Center*. Retrieved from http://www.pewresearch.org/fact-tank/2017/01/04/5-facts-about-the-minimum-wage/

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: Defining gamification. Proceedings from the 15th international academic MindTrek conference: *Envisioning future media environments*. Tampere: Finland. Retrieved from http://dl.acm.org/citation.cfm?id=2181040.

Dickey, M. D. (2005). Engaging by design: How engagement strategies in popular computer and video games can inform instructional design. *Educational Technology Research and Development, 53*(2), 67–83.

Dineen, B. R., Ash, S. R., & Noe, R. A. (2002). A web of applicant attraction: Person-organization fit in the context of Web-based recruitment. *Journal of Applied Psychology*, *87*(4), 723–734.

Drew, E. N., Lamer, J. J., Bruk-Lee, V., LeVine, P. J., & Wrenn, K. A. (2012, April). *Keeping up with the Joneses: Applicant reactions to multimedia SJTs*. Poster presented at the annual meeting of the Society for Industrial and Organizational Psychology, San Diego, CA.

DuVernet, A. M., & Popp, E. (2014). Gamification of workplace practices. *The Industrial-Organizational Psychologist, 52*, 39-42.

Edwards, J. (1991). Person-job fit: A conceptual integration, literature review, and methodological critique. In C. L. Cooper & I. T. Robertson (Eds.), *International review of industrial and organizational psychology* (Vol. 6, pp. 283–357). Chichester, UK: Wiley.

Elkins, T. J., & Phillips, J. S. (2000). Job context, selection decision outcome, and the perceived fairness of selection tests: Biodata as an illustrative case. *Journal of Applied Psychology, 85*(3), 479-484.

England, G. W. (1967). Personal value systems of American managers. *Academy of Management Journal, 10*(1), 53-68.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11*(1), 19-23.

Engle, R. W., & Kane, M. J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. In B. Ross (Ed.). *The psychology of learning and motivation* (Vol. 44, pp. 145–199). NY: Elsevier

Erdogan, B., & Bauer, T. N. (2005). Enhancing career benefits of employee proactive personality: The role of fit with jobs and organizations. *Personnel Psychology, 58*(4), 859-891.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175-191.

Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, *7*(2), 117-140.

Fetzer, M. (2013). Future directions. In M. Fetzer & K. Tuzinski (Eds.), *Simulations for personnel selection* (pp. 259-264). New York: Springer Science + Business Media.

Fetzer, M. (2015). Serious games for talent selection and development. *TIP: The Industrial-Organizational Psychologist*, *52*, 117-125.

Floryan, M. (2009). A literature review of the field of serious games. *Computer Science Department of University of Massachusetts, Amherst.*

Folger, R., & Konovsky, M. A. (1989). Effects of procedural and distributive justice on reactions to pay raise decisions. *Academy of Management Journal, 32*(1), 115-130.

Fortune (2016). 100 best companies to work for. *Fortune*. Retrieved from http://fortune.com/best-companies/.

Fredrickson, B. L., Tugade, M. M., Waugh, C. E., & Larkin, G. R. (2003). What good are positive emotions in crisis? A prospective study of resilience and emotions following the terrorist attacks on the United States on September 11th, 2001. *Journal of Personality and Social Psychology, 84*(2), 365-376.

Fu, F. L., Su, R. C., & Yu, S. C. (2008). EGameFlow: A scale to measure learners' enjoyment of e-learning games. *Computers & Education, 52*(1), 101–112.

Funder, D. C., & Colvin, C. R. (1991). Explorations in behavioral consistency: Properties of persons, situations, and behaviors. *Journal of Personality and Social Psychology, 60*(5), 773-794.

Garris, R., Ahlers, R. & Driskell, J. E. (2002). Games, motivation and learning: A research and practice model. *Simulation & Gaming, 33*(4), 441-467.

Gee, J. P. (2009). Deep learning properties of good digital games: How far can they go? In U. Ritterfeld, M. Cody, & P. Vorderer (Eds.), *Serious games: Mechanisms and effects* (pp. 65–80). New York: Routledge.

Ghani, J. A., & Deshpande, S. P. (1994). Task characteristics and the experience of optimal flow in human - computer interaction. *Journal of Psychology, 128*(4), 381–391.

Ghani, J. A., Supnick, R., & Rooney, P. (1991). The experience of flow in computer-mediated and in face-to-face groups. In J. I. DeGross, I. Benbasat, G. DeSanctis, & C. M. Beath (Eds.), *Proceedings of the Twelfth International Conference on Information Systems* (pp. 229-237). New York: ICIS.

Ghani, J.A. (1995). Flow in human-computer interactions: Test of a model. In J. Carey (Ed.), *Human factors in information systems: Emerging theoretical bases* (pp. 291- 312). Greenwich, CT: Ablex Publishing Corporation.

Giang, V. (2016, Jan 7). You should plan on switching jobs every three years for the rest of your life. *Fast Company*. Retrieved from https://www.fastcompany.com/3055035/you-should-plan-on-switching-jobs-every-three-years-for-the-rest-of-your-

Gilliland, S. W. (1993). The perceived fairness of selection systems: An organizational justice perspective. *The Academy of Management Review*, *18*(4), 694-734.

Gilliland, S. W. (1994). Effects of procedural and distributive justice on reactions to a selection system. *Journal of Applied Psychology*, *79*(5), 691–701.

Gilliland, S. W., & Honig, H. (1994). *Development of the selection fairness survey*. Paper presented at the 9th Annual Conference of the Society for Industrial and Organizational Psychology, Nashville, TN.

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. C. (2006). The International Personality Item Pool and the future of public-domain personality measures. *Journal of Research in Personality, 40*, 84-96.

Gomez-Mejia, L. R. and Welbourne, T. (1994). Compensation strategies in a global context. In R.A. Noe, J.R. Hollenbeck, B. Gerhart and P.M. Wright (Eds.) *Readings in human resource management* (pp. 561-562), Homewood, IL: Irwin.

Goodman, S. A., & Svyantek, D. J. (1999). Person–organization fit and contextual performance: Do shared values matter. *Journal of Vocational Behavior*, *55*(2), 254–275.

Google. (2017, February 9). In *Wikipedia, The Free Encyclopedia*. Retrieved from https://en.wikipedia.org/w/index.php?title=Google&oldid=764547394

Gough, H. G. (1979). A creative personality scale for the Adjective Check List. *Journal of Personality and Social Psychology, 37*(8), 1398.

Gough, H., & Heilbrun, A. (1965). *Manual for the adjective checklist*. Palo Alto, CA: Consulting Psychologists Press.

Greenberg, J. (1990). Organizational justice: Yesterday, today, and tomorrow. *Journal of Management, 16*(2), 399-432.

Guo, Y. M., & Poole, M. S. (2009). Antecedents of flow in online shopping: A test of alternative models. *Information Systems Journal, 19*(4), 369-390.

Hackman, J. R., & Oldham, G. R. (1975). Development of the job diagnostic survey. *Journal of Applied Psychology, 60*(2), 159-170.

Hamari, J., Shernoff, D. J., Rowe, E., Coller, B., Asbell-Clarke, J., & Edwards, T. (2016). Challenging games help students learn: An empirical study on engagement, flow and immersion in game-based learning. *Computers in Human Behavior, 54*, 170–179.

Hartmann, T., & Klimmt, C. (2006). Gender and computer games: Exploring females' dislikes. *Journal of Computer-Mediated Communication, 11*(4), 910-931.

Hausknecht, J. P., Day, D. V., & Thomas, S. C. (2004). Applicant reactions to selection procedures: An updated model and meta-analysis. *Personnel Psychology, 57*(3), 639–683.

Hayes, A. F. (2013). *An introduction to mediation, moderation, and conditional process analysis*. New York: The Guilford Press.

Heider, F. (1958). *The psychology of interpersonal relations*. New York, NY: Wiley.

Herriot, P. (1989). Selection as a social process. In M. Smith & I.T. Robertson (Eds.), *Advances in selection and assessment* (pp. 171-187). Chichester: John Wiley & Sons.

Hertel, G., Naumann, S., Konradt, U., & Batinic, B. (2002). Person assessment via Internet: Comparing online and paper-and-pencil questionnaires. In B. Batinic, U. Reips, & M. Bosnjak (Eds.), *Online social sciences* (pp. 115-133). Berlin: Hogrefe.

Highhouse, S., Lievens, F., & Sinar, E. F. (2003). Measuring attraction to organizations. *Educational and Psychological Measurement*, *63*(6), 986-1001.

Hoffman, D. L., & Novak, T. P. (1996). Marketing in hypermedia computer-mediated environments: Conceptual foundations. *Journal of Marketing, 60*(3), 50–68.

Hoffman, D. L., & Novak, T. P. (2009). Flow online: Lessons learned and future prospects. *Journal of Interactive Marketing*, *23*(1), 23–34.

Hoffman, D. L., Novak, T. P., & Yung, Y. F. (1999). Measuring the flow construct in online environments: A structural modeling approach. *Vanderbilt University Working Paper*. Retrieved from https://pdfs.semanticscholar.org/3960/994a1570b994e2f485bdcd37268799f70dc2.pdf

Holland, J. L. (1959). A theory of vocational choice. *Journal of Counseling Psychology, 6*(1), 35-45.

Holland, J. L. (1985). *Making vocational choices: A theory of careers*. Englewood Cliffs, NJ: Prentice-Hall.

Holland, J. L. (1997). *Making vocational choices: A theory of vocational personalities and work environments.* Odessa, FL: Psychological Assessment Resources.

Hoorens, V. (1993). Self-enhancement and superiority biases in social comparison. *European Review of Social Psychology, 4*(1), 113-139.

Horn, R., & Cleaves, A. (Eds.). (1980). *The guide to simulations/games for education and training (4th ed.)*. Beverly Hills, CA: Sage.

Hsu, C. L., & Lu, H. P. (2004). Why do people play on-line games? An extended TAM with social influences and flow experience. *Information & Management, 41*(7), 853–868.

Hurley, R. F., & Hult, G. T. M. (1998). Innovation, market orientation, and organizational learning: An integration and empirical examination. *The Journal of Marketing*, *62*(3), 42-54.

ICAR: The International Cognitive Ability Resource Team (2014). http://icar-project.com

Jackson, D. N. (2004). *Jackson Personality Inventory, Revised: Manual*. Port Huron, MI: Sigma Assessment Systems, Incorporated.

Jackson, S. A. & Csikszentmihalyi, M. (1999). *Flow in sports*. Champaign, IL: Human Kinetics.

Jackson, S. A. & Marsh, H. W. (1996). Development and validation of a scale to measure optimal experience: The Flow State Scale. *Journal of Sport and Exercise Psychology, 18*, 17–35.

Jackson, S. A. (1995). Factors influencing the occurrence of flow state in elite athletes. *Journal of Applied Sport Psychology, 7*(2), 138-166.

Jackson, S. A., & Eklund, R. C. (2002). Assessing flow in physical activity: The Flow State Scale–2 and Dispositional Flow Scale–2. *Journal of Sport and Exercise Psychology, 24*(2), 133-150.

Jones, D. A., Willness, C. R., & Madey, S. (2014). Why are job seekers attracted by corporate social performance? Experimental and field tests of three signal-based mechanisms. *Academy of Management Journal, 57*(2), 383-404.

Judge, T. A., & Bono, J. E. (2000). Five-factor model of personality and transformational leadership. *Journal of Applied Psychology, 85*(5), 751-765.

Judge, T. A., & Bretz, R. D., Jr. (1992). Effects of work values on job choice decisions. *Journal of Applied Psychology*, *77*(3), 261–271.

Judge, T. A., & Cable, D. M. (1997). Applicant personality, organizational culture, and organization attraction. *Personnel Psychology*, *50*(2), 359-394.

Judge, T. A., Erez, A., Bono, J. E., & Thoresen, C. J. (2002). Are measures of self-esteem, neuroticism, locus of control, and generalized self-efficacy indicators of a common core construct?. *Journal of Personality and Social Psychology, 83*(3), 693-710.

Kanning, U. P., Grewe, K., Hollenberg, S., & Hadouch, M. (2006). From the subjects' point of view. *European Journal of Psychological Assessment, 22*(3), 168-176.

Kantrowitz, T. M. (2014) Global assessment trends of 2014. *CEB Report*. Retrieved from https://www.cebglobal.com/talent-management/talent-assessment/resources/forms/global-assessment-trends-report.html

Kapp, K. M. (2012). *The gamification of learning and instruction: Game-based methods and strategies for training and education*. San Francisco: Pfeiffer.

Kiili, K. (2005). Digital game-based learning: Towards an experiential gaming model. *The Internet and Higher Education, 8*(1), 13–24.

Kim, T. Y., Cable, D. M., & Kim, S. P. (2005). Socialization tactics, employee proactivity, and person-organization fit. *Journal of Applied Psychology*, *90*(2), 232–241.

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.

King, D., Delfabbro, P., & Griffiths, M. (2010). Video game structural characteristics: A new psychological taxonomy. *International Journal of Mental Health & Addiction, 8*(1), 90–106.

Kluger, A. N., & Rothstein, H. R. (1993). The influence of selection test type on applicant reactions to employment testing. *Journal of Business and Psychology*, *8*(1), 3–25.

Koivisto, J., & Hamari, J. (2014). Demographic differences in perceived benefits from gamification. *Computers in Human Behavior, 35*, 179-188.

Konradt, U., Filip, R., & Hoffmann, S. (2003). Flow experience and positive affect during hypermedia learning. *British Journal of Educational Technology*, *34*(3), 309–327.

Konradt, U., Warszta, T., & Ellwart, T. (2013). Fairness perceptions in web-based selection: Impact on applicants' pursuit intentions, recommendation intentions, and intentions to reapply. *International Journal of Selection and Assessment, 21*(2), 155-169.

Koufaris, M. (2002). Applying the technology acceptance model and flow theory to online consumer behavior. *Information Systems Research*, *13*(2), 205–223.

Kristof, A. L. (1996). Person-organization fit: An integrative review of its conceptualizations, measurement, and implications. *Personnel Psychology*, *49*(1), 1-49.

Kristof-Brown, A. L., Zimmerman, R. D., & Johnson, E. C. (2005). Consequences of individuals' fit at work: A meta-analysis of person–job, person–organization, person–group, and person–supervisor fit. *Personnel Psychology*, *58*(2), 281–342.

LaHuis, D. M., MacLane, C. N., & Schlessman, B. R. (2007). Do applicants' perceptions matter? Investigating reapplication behavior using fairness theory. *International Journal of Selection and Assessment, 15*(4), 383-393.

Landers, R. N. (2014). Developing a theory of gamified learning linking serious games and gamification of learning. *Simulation & Gaming, 45*(6), 752–768.

Landers, R. N., & Armstrong, M. B. (2015). Enhancing instructional outcomes with gamification: An empirical test of the Technology-Enhanced Training Effectiveness Model. *Computers in Human Behavior*, *71*, 499-507.

Landers, R. N., & Landers, A. K. (2015). An empirical test of the theory of gamified learning: The effect of leaderboards on time-on-task and academic performance. *Simulation & Gaming*, *45*(6), 769-785.

Lauver, K. J., & Kristof-Brown, A. (2001). Distinguishing between employees' perceptions of person–job and person–organization fit. *Journal of Vocational Behavior*, *59*(3), 454–470.

Leeper, T. J. (2017). MTurkR: R Client for the MTurk Requester API. *R package version 0.8.0.*

Leventhal, G. S. (1980). What should be done with equity theory? New approaches to the study of fairness in social relationship. In K. J. Gergen, M. S. Greenberg, & R. H. Willis (Eds.), *Social exchange: Advances in theory and research* (pp. 27-55). New York: Plenum.

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.

Lievens, F., Hoye, G., & Schreurs, B. (2005). Examining the relationship between employer knowledge dimensions and organizational attractiveness: An application in a military context. *Journal of Occupational and Organizational Psychology, 78*(4), 553-572.

Lindell, M. K., & Whitney, D. J. (2001). Accounting for common method variance in cross-sectional research designs. *Journal of Applied Psychology, 86*(1), 114-121.

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist*, *57*(9), 705-717.

Lounsbury, J. W., Bobrow, W., & Jensen, J. B. (1989). Attitudes toward employment testing: Scale development, correlates, and "known-group" validation. *Professional Psychology: Research and Practice, 20*, 340-349.

Macan, T. H., Avedon, M. J., Paese, M., & Smith, D. E. (1994). The effects of applicants' reactions to cognitive ability tests and an assessment center. *Personnel Psychology*, *47*(4), 715-738.

Maertz, C. P., Jr., Bauer, T. N., Mosley, D. C., Jr., Posthuma, R. A., & Campion, M. A. (2005). Predictors of self-efficacy for cognitive ability employment testing. *Journal of Business Research, 58*(2), 160-167.

Maertz, C. P., Mosley, D. C., Bauer, T. N., Posthuma, R. A., & Campion, M. A. (2004). Do procedural justice perceptions in a selection testing context predict applicant attraction and intention toward the organization?. *Journal of Applied Social Psychology, 34*(1), 125-145.

Malone, T. W. (1981). Toward a theory of intrinsically motivating instruction. *Cognitive Science*, *5*(4), 333–369.

Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods, 44*(1), 1-23.

McCarthy, J. M., Bauer, T. N., Truxillo, D. M., Anderson, N. R., Costa, A. C., & Ahmed, S. M. (2017). Applicant perspectives during selection: A review addressing "so what?,""what's new?," and "where to next?". *Journal of Management*, *43*(6), 1693-1725.

McCrae, R. R. (2002). NEO-PI-R data from 36 cultures: Further intercultural comparisons. In R. R. McCrae & J. Allik. (Eds.), *The Five-Factor Model of personality across cultures* (pp. 105-125). New York: Kluwer Academic/Plenum Publishers.

McCrae, R. R., & Costa, P. T., Jr. (2010). *NEO Inventories professional manual.* Odessa, FL: Psychological Assessment Resources.

McCrae, R. R., Martin, T. A., & Costa Jr, P. T. (2005). Age trends and age norms for the NEO Personality Inventory-3 in adolescents and adults. *Assessment, 12*(4), 363-373.

McDonald, P., & Gandz, J. (1991). Identification of values relevant to business research. *Human Resource Management*, *30*(2), 217-236.

McFarlin, D. B., & Sweeney, P. D. (1992). Distributive and procedural justice as predictors of satisfaction with personal and organizational outcomes. *The Academy of Management Journal*, *35*(3), 626–637.

McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence, 37*(1), 1–10.

McLain, D. L. (2009). Evidence of the properties of an ambiguity tolerance measure: The Multiple Stimulus Types Ambiguity Tolerance Scale-II (MSTAT-II). *Psychological Reports, 105*(3), 975–988.

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.

Mehrabian, A. (2000). *Manual for the Affiliative Tendency Scale (MAFF).* (Available from Albert Mehrabian, 1130 Alta Mesa Road, Monterey, CA, 93940).

Meister, J. C. (2013). How Deloitte made learning a game. *Harvard Business Review*. Retrieved from https://hbr.org/2013/01/how-deloitte-made-learning-a-g

Mekler, E. D., Brühlmann, F., Tuch, A. N., & Opwis, K. (2017). Towards understanding the effects of individual gamification elements on intrinsic motivation and performance. *Computers in Human Behavior*, *71*, 525-534.

Michael, D. R., & Chen, S. L. (2005). *Serious games: Games that educate, train, and inform*. Boston, MA: Thomson Course Technology.

Mislevy, R. J., Oranje, A., Bauer, M., von Davier, A. A., Hao, J., Corrigan, S., et al. (2014). *Psychometric considerations in game-based assessment*. New York, NY: Institute of Play.

Morford, Z. H., Witts, B. N., Killingsworth, K. J., & Alavosius, M. P. (2014). Gamification: The intersection between behavior analysis and game design technologies. *The Behavior Analyst, 37*(1), 25-40.

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

Muchinsky, P. M., & Monahan, C. J. (1987). What is person-environment congruence? Supplementary versus complementary models of fit. *Journal of Vocational Behavior*, *31*(3), 268-277.

Muraven, M., & Baumeister, R. F. (2000). Self-regulation and depletion of limited resources: Does self-control resemble a muscle?. *Psychological Bulletin, 126*(2), 247-259.

Muthén, L. K., & Muthén, B. O. (1998-2017). *Mplus User's Guide. Sixth Edition*. Los Angeles, CA: Muthén & Muthén.

Nakamura, J., & Csikszentmihalyi, M. (2002). The concept of flow. In C. R. Snyder & S. J. Lopez (Eds.), *Handbook of positive psychology* (pp. 89–105). New York: Oxford University Press.

Nisbett, R. E., & Wilson, T. D. (1977). The halo effect: Evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology*, *35*(4), 250-256.

Novak, T. P., Hoffman, D. L., & Duhachek, A. (2003). The influence of goal-directed and experiential activities on online flow experiences. *Journal of Consumer Psychology*, *13*(1-2), 3-16.

O'Reilly, C. A., Chatman, J., & Caldwell, D. F. (1991). People and organizational culture: A profile comparison approach to assessing person-organization fit. *The Academy of Management Journal*, *34*(3), 487–516.

Ostberg, D., Truxillo, D. M., & Bauer, T. N. (2001, April). Effects of pre-test information on applicants' perceptions of selection fairness. In Ployhart RE (Chair), *Presenting selection information to applicants: Theoretical and practical implications*. Symposium conducted at the 16th Annual Conference of the Society for Industrial and Organizational Psychology, San Diego, CA.

Ostroff, C. (2012). Person-environment fit in organizational settings. In S. W. J. Kozlowski (Ed.), *The Oxford handbook of organizational psychology* (Vol. 1, pp. 373-408). New York: Oxford University Press.

Parsons, F. (1909). *Choosing a vocation*. Boston: Houghton-Mifflin.

Peters, M. (2005). Sex differences and the factor of time in solving Vandenberg and Kuse mental rotation problems. *Brain and Cognition, 57*(2), 176-184.

Peterson, C., & Seligman, M. E. P. (2004). *Character strengths and virtues: A handbook and classification*. Washington, DC: American Psychological Association.

Pilke, E. M. (2004). Flow experiences in information technology use. *International Journal of Human-Computer Studies*, *61*(3), 347–357.

Ployhart, R. E., & Harold, C. M. (2004). The applicant attribution-reaction theory (AART): An integrative theory of applicant attributional processing. *International Journal of Selection and Assessment*, *12*(1–2), 84–98.

Ployhart, R. E., & Ryan, A. M. (1997). Toward and explanation of applicant reactions: An examination of organizational justice and attribution frameworks. *Organizational Behavior and Human Decision Processes*, *72*(3), 308–335.

Ployhart, R. E., & Ryan, A. M. (1998). Applicants' reactions to the fairness of selection procedures: The effects of positive rule violations and time of measurement. *Journal of Applied Psychology*, *83*(1), 3–16.

Ployhart, R. E., Ehrhart, K. H., & Hayes, S. C. (2005). Using attributions to understand the effects of explanations on applicant reactions: Are reactions consistent with the covariation principle?. *Journal of Applied Social Psychology, 35*(2), 259-296.

Ployhart, R. E., Schmitt, N., & Tippins, N. T. (2017). Solving the supreme problem: 100 years of selection and recruitment at the Journal of Applied Psychology. *Journal of Applied Psychology*, *102*(3), 291-304.

Posner, B. Z. (1992). Person-organization values congruence: No support for individual differences as a moderating influence. *Human Relations*, *45*(4), 351–361.

Posner, B. Z., Kouzes, J. M., & Schmidt, W. H. (1985). Shared values make a difference: An empirical test of corporate culture. *Human Resource Management, 24*(3), 293–309.

Potosky, D., & Bobko, P. (2004). Selection testing via the internet: Practical considerations and exploratory empirical findings. *Personnel Psychology*, *57*(4), 1003-1034.

Prensky, M. (2001). *Digital game-based learning*. New York: McGraw-Hill.

Pulakos, E. D. (2005). *Selection assessment methods: A guide to implementing formal assessments to build a high-quality workforce*. SHRM Foundations. Retrieved from https://www.shrm.org/about/foundation/research/Documents/assessment_methods.pdf

Ratan, R. A., & Dawson, M. (2015). When mii is me: A psychophysiological examination of avatar self-relevance. *Communication Research, 43*(8), 1065–1093.

Raven, J., Raven, C., & Court, J. H. (1998). *Manual for Raven's progressive matrices and vocabulary scales*. Oxford: Oxford Psychologists Press Ltd.

Ravlin, E. C., & Meglino, B. M. (1989). The transitivity of work values: Hierarchical preference ordering of socially desirable stimuli. *Organizational Behavior and Human Decision Processes, 44*, 494-508.

Reeves, B., & Read, L. J. (2009). *Total engagement. Using games and virtual worlds to change the way people work and businesses compete*. Boston, MA: Harvard Business School Press.

Resick, C. J., Baltes, B. B., & Shantz, C. W. (2007). Person-organization fit and work-related attitudes and decisions: Examining interactive effects with job fit and conscientiousness. *Journal of Applied Psychology*, *92*(5), 1446–1455.

Revelian. (n.d.). *Game-based assessments.* Retrieved from http://www.revelian.com/game-based-assessments/.

Richman-Hirsch, W. L., Olson-Buchanan, J. B., & Drasgow, F. (2000). Examining the impact of administration medium on examinee perceptions and attitudes. *Journal of Applied Psychology*, *85*(6), 880-887.

Rogatko, T. P. (2009). The influence of flow on positive affect in college students. *Journal of Happiness Studies*, *10*(2), 133-148.

Rokeach, M. (1973). *The nature of human values*. New York: Free Press.

Ryan, A. M., & Chan, D. (1999). Perceptions of the EPPP: How do licensure candidates view the process? *Professional Psychology: Research and Practice*, *30*(5), 519–530.

Ryan, A. M., & Huth, M. (2008). Not much more than platitudes? A critical look at the utility of applicant reactions research. *Human Resource Management Review, 18*(3), 119-132.

Ryan, A. M., & Ployhart, R. E. (2000). Applicants' perceptions of selection procedures and decisions: A critical review and agenda for the future. *Journal of Management, 26*(3), 565–606.

Ryan, A. M., & Ployhart, R. E. (2014). A century of selection. *Annual Review of Psychology*, *65*, 693-717.

Ryan, A. M., Greguras, G. J., & Ployhart, R. E. (1996). Perceived job relatedness of physical ability testing for firefighters: Exploring variations in reactions. *Human Performance*, *9*(3), 219–240.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*(1), 68-78.

Rynes, S. L. (1993). Who's selecting whom? Effects of selection practices on applicant attitudes and behavior. In N. Schmitt, W. C. Borman, & associates (Eds.), *Personnel selection in organizations* (pp.240-274). San Francisco: Jossey-Bass.

Rynes, S. L., & Barber, A. E. (1990). Applicant attraction strategies: An organizational perspective. *Academy of Management Review, 15*(2), 286-310.

Rynes, S. L., & Connerley, M. L. (1993). Applicant reactions to alternative selection procedures. *Journal of Business and Psychology*, *7*(3), 261–277.

Saadé, R. G., & Kira, D. (2009). Computer anxiety in e-learning: The effect of computer self-efficacy. *Journal of Information Technology Education*, *8*(1), 177-191.

Sailer, M., Hense, J. U., Mayr, S. K., & Mandl, H. (2017). How gamification motivates: An experimental study of the effects of specific game design elements on psychological need satisfaction. *Computers in Human Behavior, 69*, 371-380.

Salgado, J. F., & Moscoso, S. (2003). Internet-based personality testing: Equivalence of measures and assesses' perceptions and reactions. *International Journal of Selection and Assessment*, *11*(2-3), 194-205.

Salgado, J. F. (1998). *Technical manual for the IP/5F*. Santiago de Compostela, Spain: To´rculo.

Salthouse, T. A., Atkinson, T. M., & Berish, D. E. (2003). Executive functioning as a potential mediator of age-related cognitive decline in normal adults. *Journal of Experimental Psychology: General*, *132*(4), 566-594.

Schaufeli, W. B., Salanova, M., González-Romá, V., & Bakker, A. B. (2002). The measurement of engagement and burnout: A two sample confirmatory factor analytic approach. *Journal of Happiness Studies, 3*(1), 71-92.

Schmitt, N., Oswald, F. L., Kim, B. H., Gillespie, M. A., & Ramsay, L. J. (2004). The impact of justice and self-serving bias explanations of the perceived fairness of different types of selection tests. *International Journal of Selection and Assessment*, *12*(1–2), 160–171.

Schneider, B. (1987). The people make the place. *Personnel Psychology*, *40*(3), 437-453.

Schwartz, S. H. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 25, pp. 1-65). New York: Academic Press.

Schwartz, S. H. (1994). Are there universal aspects in the content and structure of values? *Journal of Social Issues, 50*, 19-45.

Schwartz, S. H., & Bardi, A. (2001). Value hierarchies across cultures taking a similarities perspective. *Journal of Cross-Cultural Psychology*, *32*(3), 268–290.

Schwartz, T., & Porath, C. (2014, May 30). Why you hate work. *The New York Times*. Retrieved from https://www.nytimes.com/2014/06/01/opinion/sunday/why-you-hate-work.html.

Schwarz, N., & Clore, G. L. (1983). Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology, 45*(3), 513-523.

Seaborn, K., & Fels, D. I. (2015). Gamification in theory and action: A survey. *International Journal of Human-Computer Studies*, *74*, 14-31.

Sears, D. O. (1986). College sophomores in the laboratory: Influences of a narrow data base on social psychology's view of human nature. *Journal of Personality and Social Psychology*, *51*(3), 515-530.

Shaffer, J. A., & Postlethwaite, B. E. (2012). A matter of context: A meta-analytic investigation of the relative validity of contextualized and noncontextualized personality measures. *Personnel Psychology, 65*(3), 445-494.

Shank, N. C. (2011). Behavioral health providers and electronic health records: An exploratory beliefs elicitation and segmentation study. Retrieved from http://digitalcommons.unl.edu/cehsdiss/92/

Shen, L. (2016, March 29). These companies are among the hardest to get hired into. *Fortune*. Retrieved from http://fortune.com/2016/03/29/these-companies-are-among-the-hardest-to-get-hired-into/

Shergill, P. (n.d.). Winning the talent game: How gamification is impacting business and HR. *Wired*. Retrieved November 7, 2016, from https://www.wired.com/insights/2014/01/winning-talent-game-gamification-impacting-business-hr/

Shipley, W. C., Gruber, C. P., Martin, T. A., & Klein, A. M. (2009). *Shipley-2*. Los Angeles, CA: Western Psychological Services.

Shute, V. J. (2011). Stealth assessment in computer-based games to support learning. In S. Tobias & J. D. Fletcher (Eds.), *Computer games and instruction* (pp. 503–524). Charlotte, NC: Information Age.

Shute, V. J. & Ke, F. (2012). Games, learning, and assessment. In D. Ifenthaler, D. Eseryel, & X. Ge (Eds.), *Assessment in game-based learning: Foundations, innovations, and perspectives* (pp. 43-58). New York, NY: Springer.

Silverthorne, S. (2011). The most important management trends of the (still young) twenty-first century. *Harvard Business School Working Knowledge: Business Research for Business Leaders.* Retrieved from http://hbswk.hbs.edu/item/the-most-important-management-trends-of-the-still-young-twenty-first-century

Simon, H. A. (1951). A formal theory of the employment relationship. *Econometrica, 19*, 293-305.

Smith, B. W., Dalen, J., Wiggins, K., Tooley, E., Christopher, P., & Bernard, J. (2008). The brief resilience scale: Assessing the ability to bounce back. *International Journal of Behavioral Medicine, 15*(3), 194–200.

Smith, D. N., & Sivakumar, K. (2004). Flow and Internet shopping behavior: A conceptual model and research propositions. *Journal of Business Research, 57*(10), 1199–1208.

Smith, S. M. (2001). The four sources of influence of computer self-efficacy. *The Delta Pi Epsilon, 45*(1), 27–39.

Smither, J. W., Reilly, R. R., Millsap, R. E., Pearlman, K., & Stoffey, R. W. (1993). Applicant reactions to selection procedures. *Personnel Psychology*, *46*(1), 49-76.

Smits, J., & Charlier, N. (2011). Game-based assessment and the effect on test anxiety: A case study. *In European Conference on Games Based Learning* (p. 562). Academic Conferences International Limited.

Spence, M. (1973). Job market signaling. *The Quarterly Journal of Economics*, *87*(3), 355-374.

Steiner, D. D., & Gilliland, S. W. (1996). Fairness reactions to personnel selection techniques in France and the United States. *Journal of Applied Psychology*, *81*(2), 134-141.

Suazo, M. M., Martínez, P. G., & Sandoval, R. (2009). Creating psychological and legal contracts through human resource practices: A signaling theory perspective. *Human Resource Management Review*, *19*(2), 154-166.

Sweeney, P. D. (1990). Distributive justice and pay satisfaction: A field test of an equity theory prediction. *Journal of Business and Psychology, 4*(3), 329-341.

Sweetser, P., & Wyeth, P. (2005). GameFlow: A model for evaluating player enjoyment in games. *ACM Computers in Entertainment, 3*(3), 1-24.

Tajfel, H. (1974). Social identity and intergroup behavior. *Social Science Information, 13*(2), 65-93.

Tanaka, J. S. (1987). "How big is big enough?": Sample size and goodness of fit in structural equation models with latent variables. *Child Development*, *58*, 134-146.

Teng, C. I. (2008). Personality differences between online game players and nonplayers in a student sample. *CyberPsychology & Behavior, 11*(2), 232-234.

Thibaut, J., & Walker, L. (1975). *Procedural justice: A psychological analysis*. Hillsdale, NJ: Eribaum.

Thurstone, T. G., & Thurstone, L. L. (1962). *Primary mental abilities tests*. Science Research Associates.

Tippins, N. T. (2015). Technology and assessment in selection. *Annual Review of Organizational Psychology and Organizational Behavior, 2*, 551-582.

Toldi, N. L. (2011). Job applicants favor video interviewing in the candidate-selection process. *Employment Relations Today, 38*(3), 19-27.

Tom, V. R. (1971). The role of personality and organizational images in the recruiting process. *Organizational Behavior and Human Performance, 6*(5), 573-592.

Trevino, L. K., & Webster, J. (1992). Flow in computer-mediated communication electronic mail and voice mail evaluation and impacts. *Communication Research, 19*(5), 539–573.

Truxillo, D. M., & Hunthausen, J. M. (1999). Reactions of African-American and White applicants to written and video-based police selection tests. *Journal of Social Behavior and Personality*, *14*(1), 101-112.

Truxillo, D. M., Bauer, T. N., & Sanchez, R. J. (2001). Multiple dimensions of procedural justice: Longitudinal effects on selection system fairness and test-taking self-efficacy. *International Journal of Selection and Assessment*, *9*(4), 336–349.

Truxillo, D. M., Bauer, T. N., Campion, M. A., & Paronto, M. E. (2002). Selection fairness information and applicant reactions: A longitudinal field study. *Journal of Applied Psychology*, *87*(6), 1020–1031.

Truxillo, D. M., Bauer, T. N., Campion, M. A., & Paronto, M. E. (2006). A field study of the role of big five personality in applicant perceptions of selection fairness, self, and the hiring organization. *International Journal of Selection and Assessment, 14*(3), 269-277.

Truxillo, D. M., Steiner, D. D., & Gilliland, S. W. (2004). The importance of organizational justice in personnel selection: Defining when selection fairness really matters. *International Journal of Selection and Assessment, 12*(1-2), 39-53.

Tsacoumis, S. (2015). Rich-media interactive simulations: Lessons learned. In Y. Rosen, S. Ferrara, & M. Mosharaff (Eds.), *Handbook of research on technology tools for real-world skill development* (pp. 261-283). Hershey, PA: IGI Global.

Turner, J. C., (with Hogg, M. A., Oakes, P. J., Reicher, S. D. & Wetherell, M. S.). (1987). *Rediscovering the social group: A self-categorization theory.* Oxford, England: Basil Blackwell.

Uggerslev, K. L., Fassina, N. E., & Kraichy, D. (2012). Recruiting through the stages: A meta-analytic test of predictors of applicant attraction at different stages of the recruiting process. *Personnel Psychology*, *65*(3), 597–660.

Ullén, F., de Manzano, Ö., Almeida, R., Magnusson, P. K., Pedersen, N. L., Nakamura, J., ... & Madison, G. (2012). Proneness for psychological flow in everyday life: Associations with personality and intelligence. *Personality and Individual Differences, 52*(2), 167-172.

Valentine, S., Godkin, L., & Lucero, M. (2002). Ethical context, organizational commitment, and person-organization fit. *Journal of Business Ethics, 41*(4), 349–360.

Van Vianen, A. E. M. (2000). Person–organization fit: The match between newcomers' and recruiters' preferences for organizational cultures. *Personnel Psychology, 53*, 113–149.

Vancouver, J., & Schmitt, N. (1991). An exploratory examination of person-organization fit: Organizational goal congruence. *Personnel Psychology, 44*(2), 333.

Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills*, *47*(2), 599-604.

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly, 27*(3), 425–478.

Voyer, D. (2011). Time limits and gender differences on paper-and-pencil tests of mental rotation: a meta-analysis. *Psychonomic Bulletin & Review*, *18*(2), 267-277.

Webster, J., Trevino, L. K., & Ryan, L. (1993). The dimensionality and correlates of flow in human-computer interactions. *Computers in Human Behavior, 9*(4), 411–426.

Wechsler, D. (1939). *The measurement of adult intelligence*. Baltimore: Williams & Witkins.

Weibel, D., Wissmath, B., Habegger, S., Steiner, Y., & Groner, R. (2008). Playing online games against computer-vs. human-controlled opponents: Effects on presence, flow, and enjoyment. *Computers in Human Behavior, 24*(5), 2274-2291.

Weiner, B. (1986). Cognition, emotion, and action. In R. M. Sorrentino & E. T. Higgins (Eds.), *Handbook of motivation and cognition: Foundations of social behavior* (pp. 281-312). New York: Guilford Press.

Weinstein, N. D. (1980). Unrealistic optimism about future life events. *Journal of Personality and Social Psychology, 39*(5), 806-820.

West, M. A. & Farr, J. L. (Eds) (1990). *Innovation and creativity at work: Psychological and organizational strategies.* Chicheste: Wiley.

Westerman, J. W., & Cyr, L. A. (2004). An integrative analysis of person–organization fit theories. *International Journal of Selection and Assessment*, *12*(3), 252–261.

Whiteside, S. P., & Lynam, D. R. (2001). The five factor model and impulsivity: Using a structural model of personality to understand impulsivity. *Personality and Individual Differences, 30*(4), 669-689.

Wiechmann, D., & Ryan, A. M. (2003). Reactions to computerized testing in selection contexts. *International Journal of Selection and Assessment*, *11*(2-3), 215-229.

Wilson, K. A., Bedwell, W. L., Lazzara, E. H., Salas, E., Burke, C. S., Estock, J. L., … Conkey, C. (2009). Relationships between game attributes and learning outcomes: Review and research proposals. *Simulation & Gaming, 40*(2), 217–266.

Wittgenstein, L. (1953). *Philosophical investigations*. New York: Macmillan.

Wood, R. T., Griffiths, M. D., Chappell, D., & Davies, M. N. (2004). The structural characteristics of video games: A psycho-structural analysis. *CyberPsychology & Behavior, 7*(1), 1–10.

Wyer, R. S., & Carlston, D. (1979). *Social cognition, inference, and attribution*. Hillsdale, NJ: Erlbaum.

Zibarras, L. D., & Patterson, F. (2015). The role of job relatedness and self-efficacy in applicant perceptions of fairness in a high-stakes selection setting. *International Journal of Selection and Assessment*, *23*(4), 332–344.