Spring 5-27-2021

Impact of Warnings on Gamified Personality Assessments

Sabrina Yu
CUNY Bernard M Baruch College

Impact of Warnings on Gamified Personality Assessments

Sabrina Yu

Thesis Advisor: Dr. Charles Scherbaum

Submitted in partial fulfillment
of the requirements for the degree of
Master of Sciences Industrial-Organizational Psychology, Baruch College
The City University of New York

2021

ABSTRACT

Over the past several years, there has been a growing interest in using gamified and game-based assessments to measure individual differences (e.g., cognitive ability, personality) in both research and selection contexts. This interest, in part, stems from the research showing that gamified assessments are associated with more positive candidate reactions to selection processes and the potential to mitigate faking on personality assessments. To date, little research has examined the impact of gamification on faking overall and little research has combined gamification with a more traditional method of mitigating faking on personality assessments such as warnings. The purpose of this study is to combine warnings with gamified and traditional personality assessments to examine both the independent and combined effectiveness of warnings and gamification in deterring faking. This study also captures and evaluates the participants' attitudinal reactions to both warnings and gamified personality assessments. Overall, the results supported only one of the study's hypotheses. Implications of the findings and directions for future research are discussed.

Keywords: warnings, gamification, faking, personality, selection

Impact of Warnings on Gamified Personality Assessments

Personality assessments are frequently used in low-stakes and high-stakes employment selection contexts (Ployhart et al., 2017; Sackett et al., 2017). This high level of use is largely driven by research findings that personality assessments can predict future job performance (e.g., Barrick & Mount, 1991) and have little adverse impact based on gender or race (McFarland & Ryan, 2000). However, a major concern with personality assessments is the issue of faking from job applicants (Landers et al., 2011; McFarland & Ryan, 2000; Morgeson et al., 2007). In high-stakes employment contexts, applicants may be motivated to complete personality assessments in a way that will make them a more desirable candidate for a job in order to be hired by an organization. The concern is that applicants will knowingly artificially inflate their scores and be hired over applicants who possess the higher levels of the measured personality characteristics (Jeong et al., 2017). To prevent faking or dishonest responding in job applicants, researchers and practitioners have studied and implemented a number of preventive measures and deterrents including changing the format of personality assessments (Fisher et al., 2018; Pavlov et al., 2019), adding verification in the beginning of the assessment (Guo & Drasgow, 2010; Kantrowitz & Dainis, 2014), and including warnings that faking can be detected (Burns et al., 2015; Law et al., 2016). Warnings are the most used deterrent strategy and there is a large body of literature examining and comparing the effectiveness of warnings to prevent faking on personality assessments. The general consensus of this literature is that warnings can be effective in reducing faking (e.g., Dwight & Donovan, 2003; Fan et al., 2012; Landers et al., 2011).

Recently, organizations have increasingly become interested in moving away from traditional assessments and moving toward gamification and game-based assessments. The interest in these types of assessments is driven, in part, by the belief and research findings that

they may create a more positive applicant experience. Additionally, there is emerging research suggesting that gamified personality assessments may be a strategy to prevent or reduce applicants' motivation and ability to fake (Georgiou et al., 2019; Ramsay, 2017). It has been argued that while applicants are completing a personality assessment designed with game elements or designed like a game, they are less likely to fake because a) they are more engaged with playing the game and do not have the cognitive resources available to fake, and b) applicants may find it more difficult to find the exact desirable behaviors or traits that the assessment is measuring (Barends et al., 2019; Ihsan & Furnham, 2018; McCord et al., 2019).

Although much of recent research around gamification has recommended or encouraged the use of gamified assessments – including personality assessments – in low-stakes and high-stakes employment contexts, there is little work integrating gamified personality assessments with the more traditional methods of deterring faking. For instance, only one study using gamified personality assessments has included warnings (Barends et al., 2019). However, that study was not focused on the independent and joint effects of gamification and warnings on faking. Moreover, the use of gamification may overcome some of the concerns about using warnings. For example, Burns et al. (2015) have found that the implementation of warnings can negatively impact applicant reactions, and Converse et al. (2008) have found that warnings can lead to increased levels of applicant test anxiety. Gamification on the other hand has been found to lead to high levels of engagement from applicants (Levy et al., 2016) and more positive reactions from applicants (Georgiou & Nikolaou, 2020).

Tett et al. (2006) have also stated that warnings may not be effective on their own. They argue that warnings may not be an effective deterrent for all kinds of applicants and may need to be "supplemented" with other faking prevention measures to deter faking for all applicants. It

may well be the case that the addition of gamification, which candidates react positively to, mitigates the negative impact of the use of warnings on applicants and supplement them to increase their effectiveness. Therefore, this study will examine the impact of warnings on gamified personality assessments compared to traditional personality assessments. It is posited that the gamification of selection personality assessments combined with warnings to respond honestly will reduce participants' motivation and ability to fake, and also not elicit negative reactions.

Personality Assessments

After a hiatus of use, personality assessments have become mainstream in low-stakes and high-stakes employment contexts (Moyle & Hackston, 2018; Ployhart et al., 2017; Rothstein & Goffin, 2006, Sackett et al., 2017). Personality assessments have become frequently used and attractive because they have been found to be predictive of future job performance (Barrick & Mount, 1991; Goffin & Boyd, 2009) and they also show little adverse impact due to the lack of mean differences between gender, ethnicity, and age groups (Hough et al., 2001; McFarland & Ryan, 2000). Despite their appeal, one significant issue with the use of personality assessments in selection contexts is that applicants may be able to fake their responses.

Faking on Personality Assessments

Faking on personality assessments is defined as job applicants deliberately distorting their assessment responses, and knowingly engaging in impression management or socially desirable responding to create a positive or desirable profile of themselves (Ziegler et al., 2012). That is, job applicants select responses that will enable them to achieve a high score on these assessments to make them look like qualified or desirable job applicants to employers (Burns et al., 2015; Converse et al., 2008; Fisher et al., 2018; Goffin & Boyd, 2009; Richman et al., 1999;

Ziegler et al., 2012). Although previous initial research has debated if faking has an impact on selection decisions (Hanson et al., 2003; Hogan et al., 2007; McFarland & Ryan, 2000), there is a general consensus that faking can impact the rank order of applicants in both simulated and actual selection contexts and presents a real pressing concern in selection settings (Griffith et al., 2011; Landers et al., 2011; Robie et al., 2007; Rothstein & Goffin, 2006). Griffith et al. (2011) states that approximately 30% of applicants fake on selection assessments such as personality assessments.

While there are many theoretical models to explain faking behaviors (e.g., Ellingson & McFarland, 2011; Goffin & Boyd, 2009; Tett & Simonet, 2011; Zickar, 2000), Snell et al. (1999) provide one of the most direct models of faking with two factors of the ability to fake and the motivation to fake. The ability to fake refers to a test-taker's capacity to distort his or her answers in an assessment. Several variables that influence applicants' ability to fake may be dispositional factors, experiential factors, and test characteristics. For example, a personality assessment using a rating scale may provide individuals the opportunity to fake by choosing the most extreme answer choice. The motivation to fake – or intent to fake - refers to an individual's willingness to distort his or her answers in an assessment (Dullaghan, 2010; McFarland & Ryan, 2000; Rothstein & Goffin, 2006). Consistent with other frameworks (e.g., McFarland & Ryan, 2000), individuals can raise their score on personality assessments when they have the motivation and ability to fake. In this experimental study, a response instruction manipulation is used which induces the motivation to fake and the study design allows for the ability to fake. In line with previous research, it is hypothesized that those instructed to fake will have higher scores than those instructed to respond honestly.

Hypothesis 1: There will be a main effect for the response instructions such that those

instructed to fake their responses will have higher scores than those instructed to respond honestly.

Strategies to Deter Faking on Personality Inventories

The research has taken several different approaches to prevent or deter the ability and motivation to fake personality assessments including those focused on item format and type, and those focused on warning individuals that faking can be detected and will have consequences. As will be described later in this manuscript, the recent research has extended to the use of gamification as a deterrent strategy.

Item Types and Formats. The body of research focusing on designing personality items and formats to be faking resistant (i.e., decrease ability to fake) has focused on several techniques including item content and item format. Item content includes attempts to create items where the most desirable response is not easily identified (Kuncel & Borneman, 2007). Item format has primarily focused on using forced-choice items. Forced-choice personality items usually contain at least 2 desirable personality trait statements in which test-takers need to choose the statement that best described them or rank the given choices from what they think is the most to least descriptive (Converse et al., 2008). It is argued that forced-choice items prevent test-takers from choosing what they think is the "correct" response and instead forces test-takers to choose from the equally desirable response options. A recent meta-analysis by Cao and Drasgow (2019) has found that forced-choice items can be an effective strategy for minimizing the ability to fake. Specifically, they found that the score inflation effect across forced-choice personality measures was 0.06, which was smaller compared to the effect of single-statement personality measures.

Warnings. Another and very common technique to deter faking is the use of warnings (identification and consequence) in an assessment that faking can be detected. Warnings

generally consist of written notices in an assessment or verbal instructions from a proctor at the start or in the middle of the assessment telling respondents not to engage in faking behavior because the assessment is designed in a way that faking can be detected (Burns et al., 2015; Converse et al., 2008). For example, some warnings may include statements that the assessment has a back-end technology that can catch respondents who are dishonest. Additionally, warnings may state that there will be negative consequences for test-takers who are caught faking such as the test-taker may be removed from the assessment process or that the test-taker's answers will no longer be valid (Burns et al., 2015).

There has been an extensive amount of research on warnings, and they have been widely used. The interest results from the fact that they are a direct and easy strategy to implement to deter faking, as they do not require the modification of the personality inventory and can be used with any type of personality assessment. Most importantly, the body of research has found that warnings can lead to lower mean scores on personality assessments among those who are instructed or motivated to fake. In many studies comparing conditions with warnings and instructing participants to fake, conditions with warnings resulted in lower scores compared to conditions without warnings provided (Dwight & Donovan, 2003; Fan et al., 2012; Landers et al., 2011; Law et al., 2016; Vasilopoulos et al., 2005). For example, Dwight and Donovan's (2003) found that text-based warnings had an average effect on faking of d = .23, which is around a 30% decrease in the score inflation due to faking.

The downside of implementing warnings in a personality assessment is applicants' reactions. When completing personality inventories and being warned not to fake their answers, applicants may have negative reactions to the assessment overall. Applicants who are warned may have lower levels of organizational trust, overall justice and fairness perceptions, attraction to the

organization, and satisfaction with the selection process. For example, Burns et al. (2015) found that negatively worded warnings may induce test-taking anxiety in some applicants. Therefore, certain types of warnings may have a negative impact on how applicants complete the personality assessment and react to the assessment process (McFarland & Ryan, 2000).

In this study, the presence of a warning was manipulated with some participants receiving a traditional written warning and other participants receiving no warning. Consistent with the existing literature on the impact of warnings on faking behavior and applicant reactions, the following hypotheses are offered:

Hypothesis 2a – There will be an interaction between the response condition and warning condition such that those who were instructed to fake their responses and were warned will score lower than those who were instructed to fake their responses, but were not warned and those who were instructed to respond honestly (warned and not warned).

Hypothesis 2b – There will be an interaction between the response condition and warning condition such that those who were instructed to fake their responses and were warned will have higher levels of anxiety to the assessments than those who were instructed to fake their responses but were not warned and those who were instructed to respond honestly (warned and not warned).

Hypothesis 2c – There will be an interaction between the response condition and warning condition such that those who were instructed to fake their responses and were warned will report lower levels of test taking motivation than those who were instructed to fake their responses, but not warned and those who were instructed to respond honestly (warned and not warned).

Hypothesis 2d – There will be an interaction between the response condition and warning condition such that those who were instructed to fake their responses and were warned will have less favorable fairness perception of the assessment than those who were instructed to fake their responses, but not warned and those who were instructed to respond honestly (warned and not warned).

Despite the research support for warnings, Tett et al. (2006) have argued that warnings alone are not enough to reduce applicant faking. Warnings may reduce an applicant's motivation to fake in personality assessments, but not all applicants fake in the same way. Tett et al. (2006) contend that to create a truly effective faking deterrent strategy, warnings must be paired with another type of deterrent. Previous research has pointed to the combination of forced-choice response formats and warnings (Fisher et al., 2018; Rothstein & Goffin, 2006; Tett & Simonet, 2011). Forced-choice item response formats will reduce applicants' ability or capacity to fake while warnings will reduce applicants' motivation and willingness to fake. Therefore, a faking prevention strategy that affects motivation and opportunity to fake may be more successful. As is described in the next section, another possibility that has yet to be examined is pairing warnings with gamified personality assessments.

Gamification as a Strategy to Deter Faking

The most recent deterrent to faking is changing the structure of the personality assessment by adding game elements or re-designing the assessment as a game, otherwise known as gamification and game-based assessments (GBA). First, gamification must be differentiated from GBAs. GBAs are stand-alone games that have the built-in capabilities to capture and measure knowledge, skills, and abilities (Armstrong et al., 2016). Gamification, on the other hand, is a group of techniques that add game design elements to a non-game assessment. In other words, a

gamified assessment may not necessarily be a game itself (Armstrong et al., 2016; Barends et al., 2019; Ihsan & Furnham, 2018; Georgiou et al., 2018; Ketamo et al., 2018; Wood et al., 2013).

A "gamified assessment" can be considered in different ways because there are no standard requirements for which game elements to be added to an existing assessment or how many need to be added for it to be "gamified". One point of view is that any combination of game elements can be added to an assessment to make it "gamified" (Sailer et al., 2016). For example, although there are many game elements that can be included in an assessment, Hamari et al. (2014) found that most studies in their literature review have used points, leaderboards, and badges in their assessments. However, story narratives and avatars are also common (Bedwell et al., 2012).

Points are one of the most commonly used game elements in studies about gamification (Hamari et al., 2014). Points are given or awarded to players when they answer a question or fulfill a specific request or task asked of them in the gamified assessment. Points are generally displayed and provide the player feedback on how successfully the game is being played. Five different point systems in games are categorized as: experience points, redeemable points, skill points, karma points, and reputation points. The point system closest to what this study is implementing is skill points where players can obtain points by completing other tasks in addition to the main game task (Zichermann & Cunningham, 2011). Leaderboards are another common game element that provide feedback on how well a player is playing compared to other players. Leaderboards are generally public, so all players know their relative standing. One reason that leaderboards are popular is the belief that they motivate players to perform well. However, one negative consequence of using leaderboards is that players can be discouraged or face anxiety when seeing their scores compared to others. Lastly, badges are another common game element. Like points, badges also represent how well a player is performing and can act as

milestones or goals for a player to achieve (Sailer et al., 2016).

In the gamified personality assessment used in this study (i.e., Bubble Trip), the game elements are based on Bedwell's taxonomy of game attributes (Bedwell et al., 2012). Bedwell's taxonomy – while not a final or ultimate list – was created by subject matter experts and from previous literature to try to showcase a comprehensive taxonomy or list of serious game attributes that most or all games have (Armstrong et al., 2016). This taxonomy includes scoring, interaction, environment, immersion, sensory stimuli, and rules. In the Bubble Trip game, the test-taker controls a fish (avatar) in an underwater sea environment to swim to open shells to answer the personality inventory items. At the same time, the test-taker is to avoid the random jellyfish that are in the assessment and try to swim around to pop bubbles for points.

There has been growing interest in using gamification as part of assessments used to select talent. The research, to date, has found some support for the validity and utility of gamified assessments for measuring both cognitive and non-cognitive individual differences including personality. For example, Georgiou et al. (2018) found that gamified assessments of cognitive ability and personality can predict job-performance at similar levels to traditional assessments of these individual differences.

Because gamified assessments are delivered to applicants in a more dynamic and interactive environment, gamified assessments are believed to be more engaging and enjoyable for applicants to complete compared to traditional personality assessments (Armstrong et al., 2016; Georgiou et al., 2018; Ihsan & Furnham, 2018; Levy et al., 2016). This leads to applicants having more positive reactions to the assessment. Georgiou and Nikolaou (2020) have found that applicants responded more favorably to a gamified SJT compared to a traditional SJT; applicants also reported higher levels of satisfaction and fairness with the gamified version.

Lastly, it has been posited that gamified assessments can deter or prevent applicants' motivation and ability to fake on personality assessments (Levy et al., 2016). It has been argued that gamification reduces the level of faking because respondents are more engaged in a gamified assessment and do not have the cognitive resources to engage the game elements and fake at the same time (Ramsay, 2017). In support of this argument, cognitive load theory states that cognitive resources have limited capacity (Huang & Johnson, 2011), and engaging in faking behavior requires using those cognitive resources (McFarland & Ryan, 2000). In a gamified assessment, applicants are so immersed in the game elements built into the task that their cognitive resources are in use and not available for faking behavior (Ihsan & Furnham, 2018; Ramsay, 2017). In addition, gamification can be seen as a form of "stealth assessment" where a gamified assessment can better hide what traits the assessment is trying to measure; therefore, applicants are unable to determine what the "correct" answer is for faking their responses (Levy et al., 2016; Shute, 2011). In this study, participants were randomly assigned to complete a traditional personality assessment or a gamified version of that assessment. Based on this literature, it is hypothesized that:

> *Hypothesis 3a – There will be an interaction between the response condition and personality assessment condition such that those who were instructed to fake their responses on the traditional assessment will score higher than those who were instructed to fake their responses on the gamified assessment and those who were instructed to respond honestly on either type of personality assessment.*
>
> *Hypothesis 3b – There will be an interaction between the response condition and personality assessment condition such that those who were instructed to fake their responses on the traditional assessment will report higher levels of anxiety than those*

who were instructed to fake their responses on the gamified assessment and those who were instructed to respond honestly on either type of personality assessment.

Hypothesis 3c – There will be an interaction between the response condition and personality assessment condition such that those who were instructed to fake their responses on the traditional assessment will report lower levels of test taking motivation than those who were instructed to fake their responses on the gamified assessment and those who were instructed to respond honestly on either type of personality assessment.

Hypothesis 3d – There will be an interaction between the response condition and personality assessment condition such that those who were instructed to fake their responses on the traditional assessment will find the assessment less fair than those who were instructed to fake their responses on the gamified assessment and those who were instructed to respond honestly on either type of personality assessment.

Hypothesis 4a – There will be an interaction between the response, warning, and personality assessment conditions such that those who were instructed to fake their responses, were warned, and completed the gamified personality assessment will score lower than those who were instructed to fake, were not warned, and completed the traditional assessment and those who were instructed to respond honestly.

Hypothesis 4b – There will be an interaction between the response, warning, and personality assessment conditions such that those who were instructed to fake their responses, were warned, and completed the gamified assessment will report lower levels of anxiety than those who faked their responses, were not warned, and completed the traditional assessment and those who were instructed to respond honestly.

Hypothesis 4c – There will be an interaction between the response, warning, and personality assessment conditions such that those who were instructed to fake their responses, were warned, and completed the gamified assessment will report higher levels of motivation than those who faked their responses, were not warned, and completed the traditional assessment and those who were instructed to respond honestly.

Hypothesis 4d – There will be an interaction between the response, warning, and personality assessment conditions such that those who were instructed to fake their responses, were warned, and completed the gamified assessment will find the assessment fairer than those who faked their responses, were not warned, and completed the traditional assessment and those who were instructed to respond honestly.

Present Study

The present study uses an experimental research design to examine the joint impact of gamification of a personality assessment and warnings on faking, applicant reactions of test anxiety, test fairness, and test-taking motivation. Specifically, this study examined a gamified version of the HEXACO using the Bubble trip game and a traditional version of the HEXACO.

METHODS

Participants

Participants were recruited from undergraduate psychology and management courses at a large Northeastern college and from snowball sampling (non-probability sampling) where the surveys were posted on LinkedIn, Reddit, and sent to others via email[1]. Undergraduate participants received course credit in exchange for participating. To be eligible to participate, participants must be 18 years of age or older, reside in the United States, and speak and read

[1] A Chi Square test was performed; for gender – $X^2(2, N = 152) = 12.18$, $p = .00$. For race - $X^2(4, N = 152) = 22.47$, $p = .00$.

English fluently; there were no other restrictions. A total of 137 participants were in the final sample of this study – 113 participants were recruited from the undergraduate courses and 24 participants were recruited from snowball sampling. 55.5% of the participants identified as female, 43.1% identified as male, and 1.5% identified as gender fluid. The average age of the participants was 22.3. 38% of the participants identified as Asian, 12.4% identified as Black, 24.1% identified as White, 17.5% identified as Hispanic or Latino, and 7.3% selected other for their race. For demographics, the two samples did not differ much. The snowball sample's mean results were 2.96, 1.58, and 26.29 for race, gender, and age respectively. The student sample's mean results were 2.32, 1.45, and 21.37. The mean results were also very similar between the two samples when measuring the core dependent variables, which allowed the two sample data sets to be combined.

Design

This study used a 2x2x2, between-subjects design with three independent variables including response instructions, warnings, and the format of the personality assessment (see Table 1 below). Participants were randomly assigned to one of the eight conditions. For response instructions, participants were instructed to respond honestly or to respond dishonestly – fake their answers – on the personality assessment. For the warnings, participants were either given a warning that the assessment had several techniques that were built in to detect dishonest responding or faking from participants or were not given any warning. For the personality assessment format condition, participants either completed a traditional personality assessment or a gamified personality assessment.

Table 1. Summary of Experimental Conditions

	Traditional Assessment		**Gamified Assessment**	
	No Warning	**Warning**	**No Warning**	**Warning**
Honest	Condition 1	Condition 2	Condition 5	Condition 6
Faking	Condition 3	Condition 4	Condition 7	Condition 8

For the response instruction manipulations, participants who were told to respond honestly were told: *When completing the personality assessment, please respond honestly about yourself and in a way that accurately describes yourself.* Participants who were told to fake their responses were told: *When completing the personality assessment, please respond in a way that will make you look like an ideal job candidate for your ideal job position, but please do so without being obvious that you are not responding honestly.*

The warning condition provided manipulations to inform participants about techniques built into the assessment to catch individuals who were faking. This manipulation occurred after the response instructions and personality assessment format instructions. Participants were randomly assigned to either receive a warning or receive no warning at all. The warnings were constructed based on the language of the warnings used by Burns et al. (2015): *Please be aware that this personality assessment is designed to contain features to catch people who are not responding honestly. Research has shown that these features have been able to identify individuals who provide inaccurate information about themselves.*

Participants assigned to a personality condition completed either a traditional or gamified personality assessment. Participants who completed the traditional personality assessment completed the traditional online HEXACO-60 personality inventory. They were given the

following instructions: "*When selecting applicants for job positions, some organizations use personality assessments to evaluate job applicants. Personality assessments used for hiring employees measure your general tendency in interacting with others and your work. In this study, you will be asked to complete a standard personality assessment that is used for selecting job applicants. After you complete the personality assessment, you will answer questions about your reactions to the assessment, and then answer questions about your background."* The instructions were organized based on previous faking studies such as Kuncel and Borneman (2007).

Participants who completed the gamified personality assessment completed a gamified assessment called "Bubble Trip", which included the HEXACO-60 inventory. Participants who completed the gamified personality assessment were given the following instructions: *When selecting applicants for job positions, some organizations use personality assessments to evaluate job applicants to complete. Personality assessments used for hiring employees measure your general tendency in interacting with others and your work. In this study, you will be asked to complete a gamified personality assessment that is used for selecting job applicants. After you complete the gamified personality assessment, you will answer questions about your reactions to the assessment, and then answer questions about your background.".* Both personality assessments are described in the following section.

Measures

Traditional Personality Inventory. The HEXACO-60 Personality Inventory was used as the main selection assessment in this study for both the traditional and gamified personality assessment. This inventory measures the personality domains of Honesty-Humility, Emotionality, Extraversion, Agreeableness, Conscientiousness, and Openness to Experience

(Ashton & Lee, 2009). Each personality factor was a 10-item scale. All sixty items were rated using a 5-point Likert-type scale with the anchors of "Strongly Disagree" to "Strongly Agree". Several items throughout the inventory have also been reverse-coded. In this study, honesty-humility scores had a reliability coefficient of $\alpha = .73$. Emotionality scores had a reliability coefficient of $\alpha = .72$. Extraversion scores had a reliability coefficient of $\alpha = .78$. Agreeableness scores had a reliability coefficient of $\alpha = .66$. Conscientiousness scores had a reliability coefficient of $\alpha = .80$. Openness to experience scores had a reliability coefficient of $\alpha = .74$. See Appendix A for all HEXACO-60 items.

Gamified Personality Inventory. The gamified personality inventory was a 2D gamified HEXACO-60 inventory gamified assessment called "Bubble Trip" that was developed by Levy and her colleagues (Levy et al., 2016). In Bubble Trip, the test-taker of the gamified assessment controls a fish avatar. The test-taker is instructed to answer the HEXACO-60 personality inventory items with each item being displayed at the top of the screen. The test-taker has to direct his or her fish by using the arrow keys on the keyboard to make the fish swim up and answer the item. Answer choices are displayed as shells under different icons (see Figure 1). Each item has five shells to represent the strongly disagree to strongly agree scaling. To successfully answer the question, the fish must swim to the shell associated with the rating scale point the test-taker wants to select. This study used the "full version" of the Bubble Trip gamified assessment where bubbles and jellyfish were displayed on the screen; the test-taker was instructed while answering the items to pop the bubbles to earn points and avoid touching the jellyfish. The test-taker receives a point for every bubble the test-taker touches. The test-taker will be momentarily shocked if the test-taker touches a jellyfish; however, the score will not decrease. While the points that the test-taker receives from touching the bubbles will not

influence the assessment, the bubbles and jellyfish were added elements to the gamified assessment to make the assessment more game-like.

Figure 1

Test Anxiety. After the participants answered the HEXACO-60 items in either the traditional or gamified personality assessment, participants answered the Comparative Anxiety items from the Test Attitude Survey (Arvey et al., 1990). In this study, 2 out of 9 items were used ($\alpha = .63$; see Appendix C). The items were rated on a 5-point Likert type scale ranging from very inaccurate (1) to very accurate (5).

Test Taking Motivation. Participants' test-taking motivation was measured using items from Avery et al.'s (1990) Test Attitude Survey. In this study, 8 out of 10 items were used ($\alpha = .89$; see Appendix C). There were 2 items that were reverse-coded. The items were rated on a 5-point Likert type scale ranging from very inaccurate (1) to very accurate (5).

Perceived Test Fairness. The Perceived Test Fairness scale was included to measure test-takers' perceptions of the fairness of the personality assessment (Kluger & Rothstein, 1993). In this study, 1 of 4 items on this scale were used (see Appendix D). The item was rated on a 5-point Likert scale ranging from strongly disagree (1) to strongly agree (5).

Exploratory Measures. In addition to the main variables, an additional measure was included for exploratory purposes. No hypotheses were offered using this measure. Perceptions of the predictive validity of the personality inventory was assessed using two items from Chan et al. (1998). These items were rated on a 5-point Likert scale ranging from strongly disagree (1) to strongly agree (5). The scores of the item on this scale had a reliability coefficient of $\alpha = .69$ (see Appendix E for a complete list of the items).

Experience with games. All participants answered the Experience with Games scale and Preference for Video Games scale (Bourgonjon et al., 2010). The Experience with Games scale measures test-takers' perceptions about the use of video games in the classroom and asks for their experiences playing video games. The scale consists of 5 items and is rated from a 5-point Likert scale ranging from strongly disagree to strongly agree. The Preference for Video Games scale asks test-takers about their preference for the use of video games in selection assessments. The scale consists of 1 item. The scores on these scales overall had a combined reliability coefficient of $\alpha = .92$ (see Appendix F for a complete list of the items).

Manipulation Checks and Demographics. At the end of the survey, participants were asked demographic and manipulation check questions. Manipulation-check items were created to ask if participants paid attention to the instructions and warnings given prior to completing the assessments. These manipulation checks items included: *"How were you told to complete the*

personality assessment at the beginning of the study?"; "Were you given a warning after the instructions?"; "To what extent did you believe that the assessment was designed to have certain features that could catch people who were answering dishonestly?"

Procedure

Participants were randomly assigned to one of the eight conditions, and completed the entire study online distributed by Qualtrics. After completing the informed consent procedure, participants read the instructions and warnings (if applicable) and completed the personality inventory. Next, they completed reaction items and demographic questions. Both the traditional personality inventory and Bubble Trip game were embedded in the Qualtrics survey.

RESULTS

Descriptive statistics, correlations, and reliability estimates are shown in Table 2. A MANOVA to test for the manipulation check shows that the manipulation check did not entirely work, $F(24, 350.07) = .95$, $p = .54$. Only 43 participants answered both manipulation checks correctly. 13 participants selected "I don't remember" for the first manipulation check, and 46 participants selected "I don't remember" for the second manipulation check. None of the data was discarded.

Table 2

Means, Correlations, and Reliability between HEXACO-60 personality factors and reaction measures

Measure	*M*	*SD*	1	2	3	4	5	6	8	10	11	12	13	14
1. H	3.31	.61	(.73)											
2. E	3.29	.60	.04	(.72)										
3. X	3.22	.63	-.06	-.22	(.78)									
4. A	3.23	.54	.19	-.02	.27	(.66)								
5. C	3.66`	.62	.40	.03	.24	.00	(.80)							
6. O	3.44	.64	.25	-.07	.11	.02	.25	(.74)						
7. Anxiety	2.14	.96	-.13	.27	-.14	-.09	-.19	-.12	(.63)					
8. Motivation	3.25	.40	-.08	.19	.13	-.01	.23	.01	.16	(.20)				
9. Justice	3.27	.77	.02	.05	.12	.19	.02	-.19	-.05	.24	(.72)			
10. Fairness	2.64	1.01	-.05	-.13	.07	.09	-.14	-.25	.10	.32	.59	--		
11. Validity	2.50	.92	-.18	-.02	.12	.02	-.12	-.28	.12	.21	.42	.65	(.69)	
12. Exp with Games	2.78	1.15	-.06	-.20	.11	.19	-.03	.16	-.08	-.06	.16	.04	.16	(.94)

Tests of Hypotheses

In the following section, the tests of the hypotheses are grouped according to the dependent variables.

Faking. Hypotheses 1, 2a, 3a, and 4a focused on the different personality assessment scores among participants in the response, warning, and personality assessment conditions. For H1, a MANOVA was performed with the six personality factors of the HEXACO-60 as the dependent variables, and the response condition as the independent variable. Although personality scores were higher when participants were instructed to fake than those who were instructed to respond honestly, the multivariate effect was not statistically significant different between the response condition; Pillai's Trace = .07, $F(6, 108) = 1.33$, $p = .25$, partial $\eta^2 = .07$ and Hypothesis 1 was not supported.

For H2a, a MANOVA found an interaction between the response and warning condition with differences between those who were instructed to fake and who were warned compared to those who were instructed to fake and who were not warned; $F(6, 106) = 2.34$, $p = .04$, partial $\eta^2 = .12$. Therefore, Hypothesis 2a is supported. However, a follow-up two-way ANOVA indicated that there were no significant interactions for the individual personality factors (see Table 3).

Table 3

Personality	FW	FNW	*p*	η^2
H	2.86	2.96	0.27	0.01
E	3.19	3.22	0.46	0.00
X	3.28	3.16	0.15	0.02
A	3.10	3.00	0.29	0.01
C	2.91	2.99	0.83	0.00
O	3.22	3.05	0.04	0.04

FW = Fake and Warned, FNW = Fake and Not Warned.

Hypotheses 3a and 4a were not supported. It was also found that there were no statistically significant differences among those who completed the gamified assessment and those who completed the traditional assessment; $F(6, 106) = 1.27$, $p = .28$, partial $\eta^2 = .07$. It was also found that there were no statistically significant differences between the personality assessment and warning conditions; $F(6, 102) = 1.12$, $p = .36$, partial $\eta^2 = .06$. Table 4 provides all of the mean scores for those in each of the 8 conditions.

Table 4

Personality	Fake, Warn, Game	Fake, Warn, Traditional	Fake, Not Warn, Game	Fake, Not Warn, Traditional	p	η^2
H	2.81	2.88	2.93	2.98	0.47	0.01
E	3.03	3.26	3.44	3.13	0.14	0.02
X	3.36	3.27	3.35	3.08	0.67	0.00
A	3.11	3.08	3.05	2.98	0.64	0.00
C	2.86	2.94	2.97	3.01	0.42	0.01
O	3.39	3.16	3.15	3.01	0.41	0.01
Personality	Honest, Warn, Game	Honest, Warn Traditional	Honest, Not Warn, Game	Honest, Not Warn, Traditional	p	η^2
H	2.78	3.07	2.85	2.89	0.47	0.01
E	3.21	3.44	3.19	3.31	0.14	0.02
X	2.99	3.16	3.1	3.22	0.67	0.00
A	3.10	3.05	3.1	3.18	0.64	0.00
C	2.98	2.94	2.98	3.12	0.42	0.01
O	3.11	3.12	3.28	3.17	0.41	0.01

Comparative Anxiety. Hypotheses 2b, 3b, and 4b were all grouped together as they hypothesized about the applicants' levels of anxiety. A two-way ANOVA was performed as Comparative Anxiety as the dependent variable, and the response and warning conditions as the independent variables. Hypothesis 2b was not supported as it was found that there were no statistically significant differences among the participants' anxiety for those who were warned

and those who were not warned; $F(1, 133) = 1.52$, $p = .22$, partial $\eta^2 = .01$. Hypothesis 3b was not supported as it was found that there were no statistically significant differences among the participants' anxiety for those who completed the gamified assessment and those who were completed the traditional assessment; $F(1, 133) = .30$, $p = .58$, partial $\eta^2 = .00$. Hypothesis 4b was also not supported as it was found that there were no statistically significant differences among the participants' comparative anxiety; $F(1, 129) = .48$, $p = .39$, partial $\eta^2 = .00$.

Motivation. Hypotheses 2c, 3c, and 4c focused on participants' levels of test taking motivation. For H2c, a two-way ANOVA was performed. This hypothesis was not supported as there was no statistically significant interaction between the response and warning condition on motivation levels; $F(1, 133) = .12$, $p = .73$ partial $\eta^2 = .00$. Hypothesis 3c and 4c were not supported as there was no statistically significant interaction between effects of the response and personality assessment condition on motivation levels; $F(1, 133) = .13$, $p = .72$ partial $\eta^2 = .00$, and there was no statistically significant interaction between effects of the response, warning, and personality assessment conditions on motivation levels; $F(1, 129) = .81$, $p = .37$ partial $\eta^2 = .01$.

Fairness. Hypotheses 2d, 3d, and 4d were grouped together as they hypothesized about participants' reactions about the fairness of the assessments. A two-way ANOVA was performed for the reaction measure of test fairness as the dependent variable, and the response, warning, and personality assessment conditions as the independent variables. Hypothesis 2d was not supported as there was no statistically significant interaction between response and warning conditions on motivation levels; $F(1, 133) = .07$, $p = .80$ partial $\eta^2 = .00$. Hypothesis 3d was not supported as there was no statistically significant interaction between the response and personality conditions on motivation levels; $F(1, 133) = 1.91$, $p = .17$ partial $\eta^2 = .01$. And lastly,

Hypothesis 4d was not supported as there was no statistically significant interaction between the response, warning, and personality assessment conditions on motivation levels; $F(1, 129) = 3.48$, $p = .06$ partial $\eta^2 = .03$. See Table 5 for all mean scores for the measures of Comparative Anxiety, Motivation, and Test Fairness across all 8 conditions.

Table 5

Reaction	Fake, Warn, Game	Fake, Warn, Traditional	Fake, No Warn, Game	Fake, No Warn, Traditional	p	η^2
Anxiety	2.25	2.32	2.36	2.06	0.49	0.00
Motivation	3.29	3.61	3.84	3.74	0.37	0.01
Test Fairness	2.56	7.47	2.18	7.94	0.06	0.03
Reaction	Honest, Warn, Game	Honest, Warn Traditional	Honest, No Warn, Game	Honest, No Warn, Traditional	p	η^2
Anxiety	1.90	1.92	2.15	2.26	0.49	0.00
Motivation	3.34	3.36	3.53	3.60	0.37	0.01
Test Fairness	2.00	8.00	2.65	8.16	0.06	0.03

DISCUSSION

Applicant faking in unproctored internet tests (UIT) still poses a large issue and concern for employers when distributing online selection assessments. Therefore, research for the most effective faking deterrents – or a combination of them - is still essential (Landers et al., 2011; Robie et al., 2007). Warnings is the easiest and one of the most effective faking deterrents (Dwight & Donovan, 2003; McFarland & Ryan, 2000). While warnings are effective to prevent faking, however, they can also lead to applicants having negative reactions, such as text anxiety (Burns et al., 2015). With the recent increase in popularity with gamification, gamifying selection assessments has also become a new and effective way to prevent faking (Armstrong et al., 2016; Georgiou et al., 2018; Ihsan & Furnham, 2018). There has also been research finding that participants have more favorable reactions to gamified assessments compared to traditional

assessments (Georgiou & Nikolaou, 2020). Therefore, this study wanted to find if the combination of a gamified assessment with warnings would also be useful to prevent applicants' ability and motivation to fake; this study also measured applicant reactions to both warnings and the gamified assessment.

Theoretical and Practical Implications

Ultimately, the data failed to support most if not all hypotheses. Hypotheses 1, 2a, 3a, and 4a hypothesized that those who faked their responses, were not warned, and completed the traditional assessment would score higher than those who responded honestly, were warned, and completed the gamified assessment. Looking at Table 4, the personality inventory scores across all 3 conditions were mixed for each personality factor. The mixed results could be due to participants not answering the manipulation checks correctly and forgetting if they were told to respond honestly or fake their responses, which may have affected the data. A larger sample size for each of the conditions would also be needed. Future research is needed to see if there is a true difference in scores between those who complete gamified and traditional assessments.

The rest of the hypotheses were about the participants' reactions to completing the traditional and gamified personality assessments. Like the personality scores, the reaction measure scores were mixed across the reaction measures. One surprise was the fairness results. It was hypothesized that participants would find the gamified assessment fairer than the personality assessment. However, this was not the case as participants found the traditional assessment – those who faked their responses and responded honestly, and those who were warned or not – fairer than the gamified assessment. The mean results were mixed across all 8 conditions when measuring applicant motivation and anxiety, so it was unclear if gamified assessments decreased applicant anxiety and increased their motivation. Because the participants

who completed the gamified assessment had to answer all 60 items, they may have found the gamified assessment to be taking too long, which may have affected their reactions to the type of assessment. It would be worthwhile to further investigate other types of gamified assessments in terms of time duration and game elements, and to look at applicant reactions to other gamified assessments in terms of fairness, motivation, and anxiety.

Limitations and Future Research

There were a variety of limitations that may have impacted the results of this study. The most crucial one is the low sample size. Unfortunately, each of the 8 conditions in this study failed to contain a moderate sample size. The low sample size may have led to the study not having enough power to achieve statistical significance for any of the hypotheses. Due to the conditions of the pandemic, collecting additional data was not possible.

Another limitation that affected the quality of the data was that there were a number of participants who completed the gamified assessment but did not finish playing the Bubble Trip game. Measures were taken to encourage participants to complete the game, including explaining how many items were in the inventory, adding questions to ask if the participants completed all 60 items, and not letting the participants move on to the next page until a certain duration of time has passed. Unfortunately, this entire study was conducted and distributed online, so there was no way of enforcing participants to complete the entire game.

For the future, a replication of this study is needed. It is recommended to conduct this research study in person or in a lab setting to ensure all participants complete the gamified and traditional personality assessments. Adding additional manipulation checks may also be needed to ensure that participants respond the way they were told to in the beginning. It is also

recommended that future research studies the different types of gamified personality assessments that can be implemented. Such assessments could include gamified forced-choice assessments, or simply a differently designed gamified assessment. Future research should also look into the effects of gamified assessments on the motivation and ability to fake, and further look into applicant reactions to gamification and game-based assessments in general. Gamification is not going anywhere, and more organizations are implementing the use of GBA and gamified assessments. Therefore, it is important to look at the effects of gamification on reducing faking and study different combinations of gamification with other faking deterrents and personality assessment types.

Conclusion

While all hypotheses were not supported, there were several personality factors and reaction measures that showed promising trends and directions when comparing warnings and gamified and traditional personality assessments. There were also some surprising results where scores led to a trend that participants reported higher levels of test fairness for the traditional assessment compared to the gamified assessment (results were not statistically significant); therefore, there are more questions than answers at this point and more research is needed to better understand the trends. If warnings combined with a gamified assessment are more effective than just gamified assessments or traditional assessments with warnings, then organizations can start to move forward with implementing warnings in their gamified selection assessments to prevent faking in applicants. Overall, as gamification becomes more mainstream in selection practices, more research is needed to look at a combination of techniques to deter applicant faking, along with looking at applicant reactions to these kinds of techniques and selection processes. This study contributes to this body of research.

REFERENCES

Armstrong, M. B., Ferrell, J. Z., Collmus, A. B., & Landers, R. N. (2016). Correcting misconceptions about gamification of assessment: More than SJTs and badges. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, *9*(3), 671–677. https://doi-org.remote.baruch.cuny.edu/10.1017/iop.2016.69

Arvey, R. D., Strickland, W., Drauden, G., & Martin, C. (1990). Motivational components of test taking. *Personnel Psychology*, *43*(4), 695-716.

Ashton, M. C., & Lee, K. (2009). The HEXACO–60: A short measure of the major dimensions of personality. *Journal of personality assessment*, *91*(4), 340-345.

Barends, A. J., de Vries, R. E., & van Vugt, M. (2019). Gamified Personality Assessment. *Zeitschrift für Psychologie*.

Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: a meta-analysis. *Personnel psychology*, *44*(1), 1-26.

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.

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.

Bourgonjon, J., Valcke, M., Soetaert, R., & Schellens, T. (2010). Students' perceptions about the use of video games in the classroom. *Computers & Education*, *54*(4), 1145-1156.

Burns, G. N., Filipkowski, J. N., Morris, M. B., & Shoda, E. A. (2015). Impact of electronic warnings on online personality scores and test-taker reactions in an applicant

simulation. *Computers in Human Behavior*, *48*, 163-172.

Cao, M., & Drasgow, F. (2019). Does forcing reduce faking? A meta-analytic review of forced-choice personality measures in high-stakes situations. *Journal of Applied Psychology*, *104*(11), 1347.

Chan, D., Schmitt, N., Sacco, J. M., & DeShon, R. P. (1998). Understanding pretest and posttest reactions to cognitive ability and personality tests. *Journal of Applied Psychology, 83*(3), 471–485. https://doi-org.remote.baruch.cuny.edu/10.1037/0021-9010.83.3.471

Converse, P. D., Oswald, F. L., Imus, A., Hedricks, C., Roy, R., & Butera, H. (2008). Comparing personality test formats and warnings: Effects on criterion-related validity and test-taker reactions. *International Journal of Selection and Assessment, 16*(2), 155-169.

DeSimone, J. A., & Harms, P. D. (2018). Dirty data: The effects of screening respondents who provide low-quality data in survey research. *Journal of Business and Psychology, 33*(5), 559–577. https://doi-org.remote.baruch.cuny.edu/10.1007/s10869-017-9514-9

Donovan, J. J., Dwight, S. A., & Hurtz, G. M. (2003). An assessment of the prevalence, severity, and verifiability of entry-level applicant faking using the randomized response technique. *Human Performance*, *16*(1), 81-106.

Dullaghan, T. R. (2010). The effect of a reasoning warning on faking in personality testing for selection and the perception of procedural justice.

Dwight, S. A., & Donovan, J. J. (2003). Do warnings not to fake reduce faking?. *Human Performance, 16*(1), 1-23.

Ellingson, J. E., & McFarland, L. A. (2011). Understanding faking behavior through the lens of motivation: An application of VIE theory. *Human Performance*, *24*(4), 322-337.

Fan, J., Gao, D., Carroll, S. A., Lopez, F. J., Tian, T. S., & Meng, H. (2012). Testing the efficacy

of a new procedure for reducing faking on personality tests within selection contexts. *Journal of Applied Psychology*, *97*(4), 866.

Fisher, P. A., Robie, C., Christiansen, N. D., & Komar, S. (2018). The impact of psychopathy and warnings on faking behavior: A multisaturation perspective. *Personality and Individual Differences*, *127*, 39-43.

Georgiou, K., & Nikolaou, I. (2020). Are applicants in favor of traditional or gamified assessment methods? Exploring applicant reactions towards a gamified selection method. *Computers in Human Behavior*, 106356.

Georgiou, K., Gouras, A., & Nikolaou, I. (2019). Gamification in employee selection: The development of a gamified assessment. *International journal of selection and assessment*, *27*(2), 91-103.

Goffin, R. D., & Boyd, A. C. (2009). Faking and personality assessment in personnel selection: Advancing models of faking. *Canadian Psychology/Psychologie canadienne*, *50*(3), 151.

Griffith, R. L., Lee, L. M., Peterson, M. H., & Zickar, M. J. (2011). First dates and little white lies: A trait contract classification theory of applicant faking behavior. *Human Performance*, *24*(4), 338-357.

Guo, J., & Drasgow, F. (2010). Identifying Cheating on Unproctored Internet Tests: The Z-test and the likelihood ratio test. *International Journal of Selection and Assessment*, *18*(4), 351-364.

Hamari, J., Koivisto, J., & Sarsa, H. (2014, January). Does gamification work?--a literature review of empirical studies on gamification. In *2014 47th Hawaii international conference on system sciences* (pp. 3025-3034). Ieee.

Hanson, Rose & Heggestad, Eric & Thornton, George. (2003). Faking and selection:

Considering the use of personality from select-in and select-out perspectives. The Journal of applied psychology. 88. 348-55. 10.1037/0021-9010.88.2.348.

Hogan, J., Barrett, P., & Hogan, R. (2007). Personality measurement, faking, and employment selection. *Journal of Applied Psychology*, *92*(5), 1270.

Hough, L. M., Oswald, F. L., & Ployhart, R. E. (2001). Determinants, detection and amelioration of adverse impact in personnel selection procedures: Issues, evidence and lessons learned. *International Journal of Selection and Assessment*, *9*(1-2), 152-194.

Huang, W. D., & Johnson, T. (2011). Instructional game design using cognitive load theory. In *Instructional Design: Concepts, Methodologies, Tools and Applications* (pp. 1586-1606). IGI Global.

Ihsan, Z., & Furnham, A. (2018). The new technologies in personality assessment: A review. *Consulting Psychology Journal: Practice and Research*, *70*(2), 147.

Jeong, Y. R., Christiansen, N. D., Robie, C., Kung, M. C., & Kinney, T. B. (2017). Comparing applicants and incumbents: Effects of response distortion on mean scores and validity of personality measures. International Journal of Selection and Assessment, 25(3), 311-315.

Kantrowitz, T. M., & Dainis, A. M. (2014). How secure are unproctored pre-employment tests? Analysis of inconsistent test scores. *Journal of Business and Psychology*, *29*(4), 605-616.

Ketamo, H., Devlin, K., & Kiili, K. (2018). Gamifying assessment: extending performance measures with gaming data. In *Proceedings of American Educational Researcher Association's Annual Conference*.

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. https://doi-org.remote.baruch.cuny.edu/10.1007/BF02230391

Kuncel, N. R., & Borneman, M. J. (2007). Toward a new method of detecting deliberately faked personality tests: The use of idiosyncratic item responses. *International Journal of Selection and Assessment*, *15*(2), 220-231.

Landers, R. N., Sackett, P. R., & Tuzinski, K. A. (2011). Retesting after initial failure, coaching rumors, and warnings against faking in online personality measures for selection. *Journal of Applied Psychology*, *96*(1), 202.

Law, S. J., Bourdage, J., & O'Neill, T. A. (2016). To fake or not to fake: Antecedents to interview faking, warning instructions, and its impact on applicant reactions. *Frontiers in psychology*, *7*, 1771.

Levy, L. M., Solomon, R., Johnson, J., Wilson, J., Lambeth, A. J., Gandy, M., ... & Liu, R. (2016). Grouches, Extraverts, and Jellyfish: Assessment Validity and Game Mechanics in a Gamified Assessment. In *DiGRA/FDG*.

McCord, J. L., Harman, J. L., & Purl, J. (2019). Game-like personality testing: An emerging mode of personality assessment. *Personality and Individual Differences*, *143*, 95-102.

McFarland, L. A., & Ryan, A. M. (2000). Variance in faking across noncognitive measures. *Journal of Applied Psychology*, *85*(5), 812–821. https://doi-org.remote.baruch.cuny.edu/10.1037/0021-9010.85.5.812

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological methods*, *17*(3), 437.

Morgeson, F. P., Campion, M. A., Dipbooye, R. L., Holenbeck, J.R., Murphy, K., & Schmitt, N. (2007). Reconsidering the use of personality tests in personnel selection contexts. *Personnel psychology*, *60*(3), 683-729.

Moyle, P., & Hackston, J. (2018). Personality assessment for employee development: Ivory

tower or real world?. *Journal of personality assessment, 100*(5), 507-517.

Pavlov, G., Maydeu-Olivares, A., & Fairchild, A. J. (2019). Effects of applicant faking on forced-choice and Likert scores. *Organizational Research Methods*, *22*(3), 710-739.

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. https://doi.org/10.1037/apl0000081

points of contention. In M. Ziegler, C. MacCann, & R. Roberts (Eds.) *New perspectives on faking in personality assessment*. Oxford University Press.

Ramsay, P. S. (2017). Can Selection Tests Administered via Video Games Reduce Faking?

Richman, W. L., Kiesler, S., Weisband, S., & Drasgow, F. (1999). A meta-analytic study of social desirability distortion in computer-administered questionnaires, traditional questionnaires, and interviews. *Journal of applied psychology*, *84*(5), 754.

Robie, C., Brown, D. J., & Beaty, J. C. (2007). Do people fake on personality inventories? A verbal protocol analysis. *Journal of Business and Psychology*, *21*(4), 489-509.

Rothstein, M. G., & Goffin, R. D. (2006). The use of personality measures in personnel selection: What does current research support?. *Human resource management review*, *16*(2), 155-180.

Sackett, P. R., Lievens, F., Van Iddekinge, C. H., & Kuncel, N. R. (2017). Individual differences and their measurement: A review of 100 years of research. *Journal of Applied Psychology, 102*(3), 254–273. https://doi.org/10.1037/apl0000151

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.

Shute, V. J. (2011). Stealth assessment in computer-based games to support learning. *Computer games and instruction*, *55*(2), 503-524.

Smither, J. W., Reilly, R. R., Millsap, R. E., & Pearlman, K. (1993). Applicant reactions to selection procedures. *Personnel Psychology*, *46*(1), 49–76. https://doi-org.remote.baruch.cuny.edu/10.1111/j.1744-6570.1993.tb00867.x

Snell, A. F., Sydell, E. J., & Lueke, S. B. (1999). Towards a theory of applicant faking: Integrating studies of deception. *Human Resource Management Review*, *9*(2), 219-242.

Su, C. H., & Cheng, C. H. (2013). 3D Game-Based Learning System for Improving Learning Achievement in Software Engineering Curriculum. *Turkish Online Journal of Educational Technology-TOJET*, *12*(2), 1-12.

Tett, R. P., & Simonet, D. V. (2011). Faking in personality assessment: A "multisaturation" perspective on faking as performance. *Human Performance*, *24*(4), 302–321. https://doi-org.remote.baruch.cuny.edu/10.1080/08959285.2011.597472

Tett, R. P., Anderson, M. G., Ho, C., Yang, T. S., Huang, L., & Hanvongse, A. (2006). Seven nested questions about faking on personality tests: An overview and interactionist model of item-level response distortion. *A closer examination of applicant faking behavior*, 43-83.

Vasilopoulos, N, L., Cucina, J. M., & McElreath, J. M. (2005). Do warnings of response verification moderate the relationship between personality and cognitive ability?. *Journal of Applied Psychology*, *90*(2), 306.

Zickar, M. J. (2000). Modeling faking on personality tests. In D. R. Ilgen & C. L. Hulin (Eds.), *Computational modeling of behavior in organizations: The third scientific*

discipline (pp. 95–113). American Psychological Association. https://doi.org/10.1037/10375-005

Ziegler, M., MacCann, C., & Roberts, R. D. (Eds.). (2012). Faking: Knowns, unknowns, and points of contention. In M. Ziegler, C. MacCann, & R. D. Roberts (Eds.), *New perspectives on faking in personality assessment* (pp. 3–16). Oxford University Press.

Zichermann, G., & Cunningham, C. (2011). *Gamification by design: Implementing game mechanics in web and mobile apps*. " O'Reilly Media, Inc.".

Appendix A

HEXACO-60 Items

1. I would be quite bored by a visit to an art gallery. (r)
2. I plan ahead and organize things, to avoid scrambling at the last minute.
3. I rarely hold a grudge, even against people who have badly wronged me.
4. I feel reasonably satisfied with myself overall.
5. I would feel afraid if I had to travel in bad weather conditions.
6. I wouldn't use flattery to get a raise or promotion at work, even if I thought it would succeed.
7. I'm interested in learning about the history and politics of other countries.
8. I often push myself very hard when trying to achieve a goal.
9. People sometimes tell me that I am too critical of others. (r)
10. I rarely express my opinions in group meetings. (r)
11. I sometimes can't help worrying about little things.
12. If I knew that I could never get caught, I would be willing to steal a million dollars. (r)
13. I would enjoy creating a work of art, such as a novel, a song, or a painting.
14. When working on something, I don't pay much attention to small details. (r)
15. People sometimes tell me that I'm too stubborn. (r)
16. I prefer jobs that involve active social interaction to those that involve working alone.
17. When I suffer from a painful experience, I need someone to make me feel comfortable.
18. Having a lot of money is not especially important to me.
19. I think that paying attention to radical ideas is a waste of time. (r)
20. I make decisions based on the feeling of the moment rather than on careful thought. (r)
21. People think of me as someone who has a quick temper. (r)
22. On most days, I feel cheerful and optimistic.
23. I feel like crying when I see other people crying.
24. I think that I am entitled to more respect than the average person is. (r)
25. If I had the opportunity, I would like to attend a classical music concert.
26. When working, I sometimes have difficulties due to being disorganized. (r)
27. My attitude toward people who have treated me badly is "forgive and forget."
28. I feel that I am an unpopular person. (r)
29. When it comes to physical danger, I am very fearful.
30. If I want something from someone, I will laugh at that person's worst jokes. (r)
31. I've never really enjoyed looking through an encyclopedia. (r)
32. I do only the minimum amount of work needed to get by. (r)
33. I tend to be lenient in judging other people.
34. In social situations, I'm usually the one who makes the first move.
35. I worry a lot less than most people do. (r)
36. I would never accept a bribe, even if it were very large.
37. People have often told me that I have a good imagination.
38. I always try to be accurate in my work, even at the expense of time.
39. I am usually quite flexible in my opinions when people disagree with me.
40. The first thing that I always do in a new place is to make friends.
41. I can handle difficult situations without needing emotional support from anyone else. (r)
42. I would get a lot of pleasure from owning expensive luxury goods. (r)

43. I like people who have unconventional views.
44. I make a lot of mistakes because I don't think before I act. (r)
45. Most people tend to get angry more quickly than I do.
46. Most people are more upbeat and dynamic than I generally am. (r)
47. I feel strong emotions when someone close to me is going away for a long time.
48. I want people to know that I am an important person of high status. (r)
49. I don't think of myself as the artistic or creative type. (r)
50. People often call me a perfectionist.
51. Even when people make a lot of mistakes, I rarely say anything negative.
52. I sometimes feel that I am a worthless person. (r)
53. Even in an emergency I wouldn't feel like panicking. (r)
54. I wouldn't pretend to like someone just to get that person to do favors for me.
55. I find it boring to discuss philosophy. (r)
56. I prefer to do whatever comes to mind, rather than stick to a plan. (r)
57. When people tell me that I'm wrong, my first reaction is to argue with them. (r)
58. When I'm in a group of people, I'm often the one who speaks on behalf of the group.
59. I remain unemotional even in situations where most people get very sentimental. (r)
60. I'd be tempted to use counterfeit money, if I were sure I could get away with it. (r)

Items with (r) are reverse-coded.

Appendix B

HEXAC-60 Scoring Key

Honesty-Humility	
Sincerity	6, 30R, 54
Fairness	12R, 36, 60R
Greed-Avoidance	18, 42R
Modesty	24R, 48R
Emotionality	
Fearfulness	5, 29, 53R
Anxiety	11, 35R
Dependence	17, 41R
Sentimentality	23, 47, 59R
Extraversion	
Social Self-Esteem	4, 28R, 52R
Social Boldness	10R, 34, 58
Sociability	16, 40
Liveliness	22, 46R
Agreeableness	
Forgiveness	3, 27
Gentleness	9R, 33, 51
Flexibility	15R, 39, 57R
Patience	21R, 45
Conscientiousness	

Organization	2, 26R
Diligence	8, 32R
Perfectionism	14R, 38, 50
Prudence	20R, 44R, 56R
Openness to Experience	
Aesthetic Appreciation	1R, 25
Inquisitiveness	7, 31R
Creativity	13, 37, 49R
Unconventionality	19R, 43, 55R

Appendix C

Test Attitude Survey

Comparative Anxiety (2 out of 9 items used)

3. During the testing, I often thought about how poorly I was doing.
4. During the testing, I got so nervous I could not do as well as I should have.

Items with a * are reverse-coded.

Motivation (8 out of 10 items used)

8. Doing well on this test (or these tests) is important to me.
9. I wanted to do well on this test or tests.
10. I tried my best on this test or tests.
11. While taking this test or tests, I concentrated and tried to do well.
12. I pushed myself to work hard on this test or these tests.
13. I was extremely motivated to do well on this test or tests.
14. I just didn't care how I did on this test or tests.
15. I didn't put much effort into this test or tests.

Items with a * are reverse-coded.

Appendix D

Perceived Test Fairness

1. I believe that this test can predict whether I will be a successful employee.

Appendix E

Predictive Validity Perceptions

1. An employer can tell a lot about the applicant's ability to do the job based on the results of the test.
2. Failing to perform well on the test indicates that the applicant cannot perform well on the job.

Appendix F

Game Experience and Preference

1. I like playing video games.
2. I often play video games.
3. Compared to people of my age, I play a lot of video games
4. I would describe myself as a gamer.
5. I play different types of video games.

Preference for Video Games Items

1. If I had the choice, I would choose to complete selection assessments in which video games are used.