

# Gamified Assessments: A Literature Review

WHITEPAPER



**Author:** Lisette Guy MSc

**Published:** 25 February 2019

This whitepaper examines the rationale of gamification in psychometrics and considers the current state of the art.



## Current Psychometric Testing

Over the years there have been many improvements in the field of commercial and occupational psychometric testing (Cripps, 2017), such as movement away from “paper and pencil” administration in favour of computerised formats, which not only benefit psychometrics by bringing them into a medium better suited to the digital age, but also allows for automated scoring, remote testing which affords greater flexibility, and reduction in the administrative burden to organisations.

Although this provides benefits in the form of saving time and cost, the greatest advantages are those that improve the psychometric properties of the tests themselves, thus making them inherently more useful. Implementing methods such as Computer Adaptive Testing (CAT) and Item Response Theory (IRT) enables greater reliability at all levels of ability (Weiss, 2004), superior precision of measurement than afforded by assessments that utilise Classical Test Theory (CTT; Colwell, 2013; Davey, 2011; Martin, & Lazendic, 2017; Weiss, 1982; Weiss, 2004), improved candidate experience (Martin, & Lazendic, 2017), reduction in test length (Linacre, 2000), protection against cheating attempts (Davey, 2011; Rudner, 1998) and thus the overall validity of the assessment, supporting the addition of updated test content in the form of new items.

Despite the efficacy of psychometric tests and the advances that have been made, there are still areas in which current psychometric assessments have been suggested to fall short. In a world where psychometric testing is now an important part of most job selection processes (Cripps, 2017), it is paramount that we continually look for ways to improve them.



## Problems with Current Psychometric Testing

Current psychometric tools are suggested to encounter a variety of problems, albeit not all psychometric tests are likely to suffer the same disadvantages, as this would depend on the design and development of the test itself.

Nevertheless, common issues associated with psychometric tests in general include: increased anxiety for candidates completing high stakes tests and the effect this could have on performance measures, self-report measures have been suggested to lead to selection of responses that are considered “socially desirable”, multiple choice questionnaires (MCQ) could encourage or be susceptible to “guessing behaviours”, candidates may experience fatigue if tests are too long, some tests can be quite time intensive, some tests may suffer from “ceiling effects” or “basement effects”, in which case the candidate will have reached the upper or lower limit of measurement that the test will allow.

Limitations regarding the measurement of certain dimensions could be restricted by the test itself. Some assessments could be considered “unengaging” due to their serious nature (Shute, & Rahimi, 2017), which could lead to decreased motivation from candidates. In some cases, completing assessments as part of a recruitment process could be seen as a barrier by some candidates (Moscoso, Salgado, & Anderson, 2017), which deters them from completing their applications, and potentially leads to loss of quality talent and increased candidate attrition.

The majority of these issues can be combated with tried and tested psychometric methods, for example, Computer Adaptive Tests (CATs) allow for shorter tests, greatly decreasing the time intensive burden on candidates and reducing the likelihood of candidate fatigue. The likelihood of encountering ceiling effects can be greatly reduced by creating large items banks which span a broad range of difficulties, allowing for suitable measurement at each level.

Social desirability or guessing behaviours when using self-report or MCQ measures can be identified through reverse keyed items, or by using methods such as CUTLO and CUTHI in tests that use more advanced models, these methods can be adopted in order to set a cut-off point at which examinees who score much higher or lower than predicted by the model are eliminated, as these are considered to be a product of guessing behaviours or lack of examinee motivation to complete the test to their true potential. Such behaviours can sometimes be accounted for depending on the choice of parameter logistic model adopted, such as the 3

Parameter Logistic model (3PL) which parameterises guessing behaviours automatically.

However, despite the efficacy of these methods, issues related to candidate engagement, motivation and attraction remain. Such issues are likely to have been contributing factors to the rise in popularity of “Gamification” or Game-Based Assessments (GBAs), but what is gamification?



## What is Gamification?

Although the term “Gamification” may have mixed meanings depending on the medium it is used to denote, we refer to Gamification exclusively in the context of assessment tools.

Gamification in terms of assessment, generally refers to psychometric assessments that incorporate elements of traditional game play. This has led to the development of psychometric tools, often referred to by a variety of different names such as: “Gamified Assessments”, “Serious Games” and “Game-Based Assessments”. There is a separate debate on the semantics of these terms, but for brevity, these terms are used interchangeably throughout this article.

It is difficult to clearly define what it is that constitutes a “game” in this context, although Deterding, Dixon, Khaled and Nacke (2011), offered the following definition of Gamification: “the use of game design elements in non-game

contexts". However, within this definition there is a broad range of assessments that could meet such criteria. For example, this could include assessments that are completed on a mobile device, traditional psychometric tests that happen to use game-like interfaces, those that incorporate gamified features, or full games that have been specifically developed to record candidates' interactions, responses and behaviours during gameplay, to infer attribute measures without the use of self-report or MCQ response formats.

Research suggests that incorporating game design elements into tasks or procedures that would otherwise be considered unenjoyable, could have the potential to increase user engagement, enjoyment and motivation (Flatla, Gutwin, Nacke, Bateman, & Mandryk; 2011). In recent years, many companies have incorporated elements of gameplay into training and assessment tools, to provide such benefits to candidates and users alike (Lumsden, Edwards, Lawrence, Coyle, & Munafò, 2016). Therefore, for the purpose of clarity, we refer to Gamification in the context of assessment - as the process of applying such elements to assessment processes, in a deliberate attempt to either make them more appealing, enjoyable, engaging or less onerous to candidates.



## **How Can Gamification Benefit Psychometric Assessments?**

### **Candidate Attrition, Engagement and Motivation**

Traditional psychometric tests have been suggested to be "repetitive" or "unengaging" to test takers (Shute, & Rahimi, 2017), which could be problematic considering that a certain amount of motivation is required to complete a psychometric assessment.

Interestingly, test-taker motivation has previously been shown to correlate positively with test performance (Wise & DeMars, 2005), supporting the prior notion that attitudes towards assessments may interact with overall performance (Schmit & Ryan, 1992). It may be important to consider that such a relationship between motivation and test performance could reflect differences in cognitive ability, wherein those with higher cognitive ability may find it easier to engage with assessments than those with lower cognitive ability. Although, similar studies into the relationship between test-taker motivation and ability, produced correlations that were consistently non-significant (Wise & DeMars, 2005).

Test-takers who are not engaged by their assessment material, or those who lack motivation to perform well, may provide self-report responses or select answers that are not truly indicative of their

attributes or abilities (Wise, 2007). This presents a challenge for a psychometric tool, as it could lead to reduced measurement validity, and thus have the potential to undermine the utility of the assessment itself (Finn, 2015; Wise, 2007).

Despite the important implications that test-taker motivation and engagement could have upon assessment performance, it should be noted that most of the psychometric assessments administered in the aforementioned studies would have been considered “low-stakes” assessments. When used in an occupational or recruitment context, psychometric tests are more likely to be perceived as “high-stakes” tests due to their associated outcome, i.e. the opportunity of employment, and thus may inherently inspire a certain amount of engagement or motivation from the test-taker (Finn, 2015). With this in mind, it could be suggested that psychometric tests may be less susceptible to the effects of test-taker motivation and engagement, when used in high-stakes settings such as recruitment, than when used in a low-stakes setting.

Nevertheless, when used in high-stakes settings, psychometric tests could be perceived as a barrier by certain test-takers (Moscoso, Salgado, & Anderson, 2017). In some cases, this could have the potential to decrease candidates’ motivation if they believe their goal is no longer attainable (Ames, 1992), and perhaps deter them from pursuing the next stage in an application process. Candidate attrition often comes into play after invitations to complete psychometric assessments have been sent to the most

desirable applicants (Hardy, Gibson, Sloan, & Carr, 2017). Although this may be a reflection of a candidate’s eagerness to pursue that particular role to begin with, incorporating tools to encourage candidate attraction, motivation and engagement could still prove to be beneficial at this stage.

Interestingly, research into the use of gamified assessments suggests that incorporating elements of gameplay into psychometrics may not only promote candidate or user engagement (Kato & Klerk, 2017; Shute & Rahimi, 2017), but also preserve engagement throughout the entirety of the assessment (Levy et al., 2015).

Previous studies also provided evidence that engagement could be increased through gamified assessments, whether in terms of cognitive or non-cognitive assessment (Spronck, Baemans, & Van Lankveld, 2012; Tekofsky et al., 2013). This could indicate that the successful use of gamification to promote candidate engagement may be applicable to multiple areas of assessment, such as but not limited to: intellect and cognitive ability, competency evaluation, judgement and decision-making, and personality or behavioural assessments.

## **Test Anxiety**

Another important factor to consider is test anxiety, as this issue is often associated with high-stakes psychometric assessments (Cassady, & Johnson, 2002; Martin, & Lazendic, 2017; McCarthy, & Goffin, 2005; Spielberger, & Vagg, 1995).

Ziedner (1998) suggested that test anxiety referred to physiological, psychological and phenomenological responses prompted by an evaluative situation or assessment, due to concerns regarding failure or resultant negative consequences. Although many have been offered, there is not a universally accepted definition of test anxiety. Instead, there are various definitions that share similarities in their approach, which tend to involve an anxiety provoking evaluative situation and the subsequent impact this has upon an individual (Buck, 2017). Test anxiety is generally associated with multiple dimensions, although it was originally considered to comprise a singular dimension throughout initial study (Cassady, & Johnson, 2002).

Early research produced findings to suggest that anxieties encountered during evaluative situations were largely due to a combination of "self-deprecating ruminations" and a state of "heightened physiological responses" (Sarason, 1961). Following this, Liebert and Morris (1967) proposed a two-dimensional model of test anxiety comprising emotional and cognitive components, referred to as "Emotionality" and "Worry", respectively. Emotionality is considered to represent the individual's subjective awareness of experiencing such heightened physiological responses, some examples of which are considered to include increased galvanic skin response, increased heart rate and bodily tension. Whereas "Worry", also referred to as cognitive test anxiety, is thought to refer to the related cognitions and personal beliefs manifested in response to such autonomic arousal i.e. "I feel that I will perform poorly", "I am not prepared for

this exam and will likely fail" etc. Despite the contribution of additional multidimensional models in more recent years, Liebert and Morris's (1967) two-dimensional model of test anxiety maintained its popularity among researchers (Cassady, & Johnson, 2002; McCarthy, & Goffin, 2005; Spielberger, & Vagg, 1995).

Research has suggested that individuals with high levels of test anxiety may not perform as well as those with low levels of test anxiety (Cassady, & Johnson, 2002), with some studies reporting an associated reduction in academic test scores by twelve percentile points for those with high test anxiety, when compared to low test anxiety peers (Hembree, 1988). Similar findings have also been reported in regard to selection assessments, with research demonstrating moderate, negative correlations between test anxiety and cognitive ability measures (Arvey, Strickland, Drauden, & Martin, 1990; McCarthy, & Goffin, 2005).

This has important implications for psychometric testing during applicant selection, as the high-stakes nature of such an assessment could induce physiological and psychological responses in some applicants that may have the potential to inhibit their performance, thus putting these individuals at a disadvantage. A number of factors are purportedly associated with test anxiety evoked during psychometric assessment, such as but not limited to: test difficulty, perceived performance (Tonidandel, Quiñones, & Adams, 2002), associated outcome (i.e. high-low stakes), and time limits (Morris, & Liebert, 1969).

Some methods to reduce the level of test anxiety provoked by psychometric assessments have yielded mixed findings concerning their efficacy. For example, Computer Adaptive Testing (CAT) has been suggested as a means to reduce test anxiety, as only test items that are deemed to be within the candidate's ability range are deployed and thus should not "overwhelm" candidates (Davey, 2011). Contrary to this, some findings suggested that the use of CAT may be more likely to increase test anxiety in some cases, than compared to traditional fixed-form tests (Colwell, 2013). Research conducted into performance perception and test anxiety when experimentally manipulating item difficulty via CAT, reported that the number of items answered correctly during an assessment significantly predicted candidates' perceived performance, even once actual performance had been accounted for (Tonidandel, Quiñones, & Adams, 2002). Interestingly, the same study reported that perceived performance was not a significant predictor of test anxiety, thus suggesting that the occurrence of test anxiety may in some cases be independent of whether a candidate believes they have performed well or not. Based on the available evidence, it is not unreasonable to suggest that additional methods to target test anxiety reduction should be investigated.

Suggestions to reduce test anxiety include conducting assessments within a gaming context (Shute, Hansen, & Almond, 2008). With some studies reporting a reduction in the amount of test anxiety provoked in candidates who completed a game-based assessment (Kocadere, & Çağlar, 2015), including when compared to traditional

methods of assessment (Turan, & Meral, 2018). One explanation offered to account for this, is that of "Stealth Assessment". Stealth Assessment refers to an assessment during which candidates are either unaware of what the target variables are or how they are being measured (Shute et al., 2009; Shute, Hansen, & Almond, 2008; Shute, & Ventura, 2013) whether assessed through provoking particular behaviours, responses to certain stimuli or the way candidates might interact with elements of the assessment itself. Games are considered to be a medium particularly suited to this kind of assessment, as they award greater flexibility in terms of the types of data that can be recorded, have the potential to take advantage of modern technological advances and may be better suited to disguising target variables (de Klerk, Veldkamp, & Eggen, 2015; Shute, Hansen, & Almond, 2008).

Similarly, it is possible that the inclusion of gamified elements may help to alter candidates' attitudes towards the assessment itself, as they may be less likely to perceive a gamified assessment in the same way they would a traditional assessment, thus reducing the likelihood of encountering some of the physiological and psychological responses associated with test anxiety. That is to say, gamified assessments may reduce candidate anxiety due to the inclusion of elements that are likely to make the assessment experience more immersive, enjoyable, and thus less onerous than traditional assessments (Levy, Lambeth, Solomon, & Gandy, 2018).

Although some of the aforementioned studies refer to educational assessment,

the use of gamification as a method to reduce test anxiety may also be applicable to psychometric assessments, particularly those encountered during recruitment and selection procedures.

## **Cognitive Complexity**

There is a vast amount of evidence to suggest that cognitive ability, or “mental horsepower”, is the single best predictor of job performance (Schmidt, & Hunter, 1998; Schmidt, Oh, & Shaffer, 2016), as there will generally always be cognitive requirements in any role. In fact, it has been suggested that the modern workplace is becoming increasingly cognitively complex through the requirement of additional skills to complete everyday tasks (Gottfredson, 1997b), for example, technological literacy is a basic requirement of most roles today, as the majority of organisations operate within a digital setting to a degree. Knowledge of specific programs, statistical packages, computational software and databases are commonplace among many roles.

Although specific needs may differ depending on job function, there will ordinarily be a technological element to a role regardless of job level. Employees will generally be expected to operate effectively within a companywide internal network, and to adapt to any updates or improvements that are made in the future. In this sense, employees must have a certain level of adaptability and ability to continuously learn on the job, no matter the duration of their tenure. This increased demand for learning has been suggested to place a greater cognitive

burden upon employees, not to mention that a lot of these skills will be used simultaneously while multitasking. We even see evidence of this with the number of devices used while performing a role, with computers and smartphones often being provided to employees as a necessity.

With this in mind, it makes sense to develop psychometric assessments that capture these increasing cognitive demands. Gamified assessments may benefit psychometric testing in this area, as they often incorporate additional cognitive elements such as spatial reasoning, response time and working memory, due to the nature of the interface or the game itself. As opposed to traditional psychometric formats, gamified assessments may provide a greater amount of simultaneous cognitive stimuli (such as moving objects, increased attentional requirements, sounds, and changing colour patterns), meaning that candidates are required to respond to a greater amount of information within a short time frame. Although stimuli such as flashing lights and changing colour patterns may not typically be associated with job performance, it is in fact the cognitive function required to process these multiple streams of information that is of importance, as candidates would likely be required to do so in the workplace (Gottfredson, 1997b).

By increasing cognitive complexity and including the need for various “mental processes” to be utilised, it is possible that this could increase the validity of a psychometric tool (Schmidt, 2012), as they are able to incorporate multiple cognitive constructs and thus provide a



more robust measure of cognitive ability as a whole.

### **Predominantly Language Free**

Some gamified selection assessments have been suggested to reduce, or possibly eliminate inherent biases associated with traditional cognitive pre-employment assessments, through mechanisms such as increased motivation and self-efficacy (Godollei, & Chapman, 2016). It is possible that the reduced verbal loading present in gamified assessments could also promote a “level playing field” for candidates who may not have the same level of language fluency. Some articles suggest that language is not a requirement when identifying gifted learners, and in fact suggests that the inclusion of language fluency requirements could mask nonverbal problem-solving skills (Lewis, 2001). Similar effects have been reported for “culture fair”, nonverbal cognitive assessments such as logical, inductive or diagrammatic reasoning tests (Cattell, 1940, 1949).

By reducing the in-item verbal loading of both numerical and verbal assessments, which are traditionally considered to incorporate a strong linguistic element, gamification may aid in transferring these benefits across a variety of assessments, as opposed to reserving this for assessments that focus on image-based problem solving.

However, it is important to note that the notion of “culture fair” assessment has received some criticism, with some culture fair selection methods being considered

“inadequate” for their task, or requiring more complex analysis (Petersen and Novick, 1976).

Therefore the limitations of such methods should be considered and further study into implementing culture fair assessment via gamification should be undertaken in the future.

### **Speed of Administration**

In terms of assessing cognitive ability, traditional aptitude tests often focus on a single facet of cognitive ability at a time, for example, verbal reasoning, numerical reasoning, and logical reasoning tests each provide a measure of a specific sub-facet of cognitive ability, with a combination of all three providing a measure of general cognitive ability. The time limits of traditional aptitude tests tend to require that candidates complete a question, on average, once every minute. This means that in order to administer enough items (or questions) to meet sufficient levels of measurement reliability, the tests can range from around 12 to 25 minutes in length, depending on the test developer and the chosen psychometric model. This may not sound too onerous, but when we consider that measures of general cognitive ability often require the completion of multiple aptitude tests, this time limit can triple in order to provide the most effective measure.

Although some gamified assessments may incorporate similar time limits to those of traditional aptitude tests, it is also possible for gamified items to be administered much faster than traditional question items. By providing a greater number of

questions within a short time frame, sufficient measurement reliability can be achieved much faster, allowing for robust measurement equivalent to that produced by a traditional aptitude test, in a fraction of the time. Thus making it easier to gain a holistic overview of general cognitive ability, as multiple assessments can be administered in the same time taken to complete one traditional aptitude test.

Simply put, if implemented effectively, gamified assessments may have the potential to produce superior measures of cognitive ability than traditional aptitude tests, in a greatly reduced time frame.



## What Have Games Been Used to Measure?

The literature suggests that gamified assessments have been used to measure an assortment of variables, such as but not limited to: cognitive ability (both broad and narrow traits), specific personality traits, decision-making, critical thinking, business management skills, non-verbal indices and real world competencies (Lumsden, Edwards, Lawrence, Coyle, & Munafò, 2016; Spronck, Balemans, & Van Lankveld, 2012; Shute, & Ventura, 2013; Vos, 2015).

Although the particular methods and applications of gamification may differ throughout these studies, they provide some evidence that gamification may be a useful means of assessment across an array of target measures. It should also

be noted that the majority of variables assessed via gamification during these studies are often included as part of job selection assessments, further supporting the potential application and utility of gamification as a method of psychometric assessment.

However, as the use of gamified assessments during job selection is a relatively new phenomena and the development of which is still somewhat in its infancy, it is vital that we consider not only their potential use and applicability to the domain of psychometric assessment, but that we also address their current limitations.



## Gamification: Limitations and Caveats Thus Far

Despite the potential benefits that gamification could afford the realm of assessment, we must also consider the additional impact such a method could have upon candidates and the assessment process itself, some of which may yet to be observed. The available literature proposes the notion that by incorporating gamified elements, such as irrelevant colours, objects and machinations, we could unintentionally introduce new opportunities for measurement bias to prevail (de Klerk et al., 2015; Kraemer, 1992). With this in mind, it is suggested that further understanding is required regarding the effect, if any, of game mechanics and how they could alter choices in response making, upon measured variables (Levy et al., 2016).

As discussed previously, the fundamental utility of any assessment tool is its ability to reliably measure its intended variable with adequate precision, that is to say, the quality of the assessment. Although it is unlikely that any tool is able to achieve this with perfection, having the necessary psychometric properties is paramount to a psychometric tool's intended purpose. Therefore, we must ensure that gamified assessments do not sacrifice measurement quality, for the sake of other afforded benefits.

Although some research regarding the validity of gamified assessments has been positive, comments regarding the methodology of such studies have prompted calls for increased scientific rigour (Kato et al., 2012). This is not intended to discount previous findings, but merely to question and advance our understanding of the effects of gamification. Especially as its use among businesses and organisations is on the rise and may be misused by those eager to adopt a new trend (Callan, Bauer, & Landers, 2015). Similarly, it has also been suggested that the current research body lacks adequate understanding of the reliability and validity of GBA measures, in terms of their comparison to measures obtained via traditional formats of assessment (Levy et al., 2016). With some gamified tools suggested to lack supporting evidence from any kind of validation research at all (Deterding, 2015).

Of the available research regarding gamification validity studies, the majority of published studies are considered to report small sample sizes (Lumsden,

Edwards, Lawrence, Coyle, & Munafò, 2016). This could have implications for the findings of such studies, as the avoidance of type i and type ii error during analysis requires adequate statistical power, with small sample sizes often cited as a contributing factor in terms of underpowered analysis (Cohen, 1992). By ensuring there is sufficient statistical power to perform the analysis, type i and type ii error become far less likely, that is to say, the probability of producing a significant result that is in actuality a false positive, and the likelihood of producing a non-significant result that is a false negative, becomes greatly reduced (Cohen, 1988).

In order to produce meaningful results from validity studies, it is recommended that research with larger sample sizes is undertaken, or that statistical power analyses are conducted and reported in conjunction with their findings. If this is not possible within the constraints of the research itself, or such information is not reported, the magnitude of effect size should be considered, as this may be more informative than statistical significance in some cases. However, this responsibility does not fall solely upon researchers, as those who opt to employ gamified assessments should request information regarding the psychometric properties of such tools, prior to their deployment.



## Conclusions

Having reviewed the available literature on the benefits gamification may offer to psychometric assessments, their use to date, research regarding their efficacy and their current limitations, it appears that through the application of robust psychometric methodologies to Game-based Assessments (GBAs), tests publishers may be better equipped to develop gamified assessments that meet the needs of modern recruitment and selection assessment processes.

It is clear that despite benefits afforded by traditional forms of psychometric assessment, there is room for improvement in regard to candidate experience. Conversely, gamified assessments at present appear to offer benefits in terms of candidate experience, engagement and motivation, yet are generally considered to lack the robust properties of their scientifically established counterparts. Gamified assessments appear to offer the potential to resolve current issues faced by traditional psychometrics, but are yet to be held to the same standard, or applied in the way that allows them to be most effective. As a result of this, gamification may have been somewhat overlooked as a viable, robust means of assessment, and instead seen merely as a trend or tool to improve candidate attraction. Similar views have been reflected in the literature, as a systematic review of cognitive assessment via gamification suggested that further development of gamified assessments could produce scientifically valid, engaging measures of their target variables, when applied correctly (Lumsden, Edwards,

Lawrence, Coyle, & Munafò, 2016).

In order to develop effective gamified assessments, it has been suggested that such tools may be improved through thorough definition of their target measures and understanding of the correct way to trigger them (Bellotti et al., 2013). This research also recommended that target measures should be seamless and modular, in order to be easily applicable across different GBAs. For this reason, cognitive ability may be particularly suited to gamification, based on the wealth of quality publications regarding cognitive ability theory, and years of substantial, replicable evidence concerning its undoubted ability to predict job performance (Gottfredson, 1997a; Gottfredson, 1997b; Schmidt, 2012; Schmidt, & Hunter, 1998; Schmidt, & Hunter, 2004; Schmidt, Oh, & Shaffer, 2016).

It is hoped that by refining the process of applying gamification to psychometric assessments, test publishers can bridge the gap between quality measurement and candidate experience, achieving the proposed benefits of gamified assessments (i.e. increased engagement, candidate motivation, reduced anxiety, and reduced administration time) without sacrificing the scientific rigour of tried and tested psychometric methodology.



## References

- Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of educational psychology*, 84(3), 261.
- 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*, 9(3), 671-677.
- Arvey, R. D., Strickland, W., Drauden, G., & Martin, C. (1990). Motivational components of test taking. *Personnel Psychology*, 43(4), 695-716.
- Bellotti, F., Kapralos, B., Lee, K., Moreno-Ger, P., & Berta, R. (2013). Assessment in and of serious games: an overview. *Advances in Human-Computer Interaction*, 2013, 1.
- Buck, R. J. (2017). *An Investigation of Attentional Bias in Test Anxiety* (Doctoral dissertation, The University of Manchester).
- Callan, R. C., Bauer, K. N., & Landers, R. N. (2015). How to avoid the dark side of gamification: Ten business scenarios and their unintended consequences. In *Gamification in education and business* (pp. 553-568). Springer, Cham.
- Cassady, J. C., & Johnson, R. E. (2002). Cognitive test anxiety and academic performance. *Contemporary educational psychology*, 27(2), 270-295.
- Colwell, N. M. (2013). Test anxiety, computer-adaptive testing, and the common core. *Journal of Education and Training Studies*, 1(2), 50-60.
- Cripps, B. (Ed.). (2017). *Psychometric testing: Critical perspectives*. John Wiley & Sons.
- Davey, T. (2011). *A Guide to Computer Adaptive Testing Systems*. Council of Chief State School Officers.
- de Klerk, S., Veldkamp, B. P., & Eggen, T. J. (2015). Psychometric analysis of the performance data of simulation-based assessment: A systematic review and a Bayesian network example. *Computers & education*, 85, 23-34.
- Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011, September). From game design elements to gamefulness: defining gamification. In *Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments* (pp. 9-15). ACM.
- Deterding, S. (2015). The lens of intrinsic skill atoms: A method for gameful design. *Human-Computer Interaction*, 30(3-4), 294-335.
- Flatla, D. R., Gutwin, C., Nacke, L. E., Bateman, S., & Mandryk, R. L. (2011, October). Calibration games: making calibration tasks enjoyable by adding motivating game elements. In *Proceedings of the 24th annual ACM symposium on User interface software and technology* (pp. 403-412). ACM.
- Finn, B. (2015). *Measuring Motivation in Low-Stakes Assessments*. ETS Research Report Series, 2015(2), 1-17.

Godollei, A., & Chapman, D. S. (2016). Gamified cognitive assessments in selection: Validity, discrimination and applicant reactions. In *Academy of Management Proceedings* (Vol. 2016, No. 1, p. 17438). Briarcliff Manor, NY 10510: Academy of Management.

Hardy III, J. H., Gibson, C., Sloan, M., & Carr, A. (2017). Are applicants more likely to quit longer assessments? Examining the effect of assessment length on applicant attrition behavior. *Journal of Applied Psychology*, 102(7), 1148.

Hembree, R. (1988). Correlates, causes, effects, and treatment of test anxiety. *Review of educational research*, 58(1), 47-77.

Kato, P. M. (2012). Evaluating efficacy and validating games for health. *GAMES FOR HEALTH: Research, Development, and Clinical Applications*, 1(1), 74-76.

Kato, P. M., & de Klerk, S. (2017). Serious games for assessment: welcome to the jungle. *Journal of Applied Testing Technology*, 18(S1), 1-6.

Kocadere, S. A., & Çağlar, Ş. (2015). The design and implementation of a gamified assessment. *Journal of e-Learning and Knowledge Society*, 11(3).

Kraemer, H. C. (1992). *Evaluating medical tests: Objective and quantitative guidelines* (Vol. 26). Newbury Park, CA:: Sage publications.

Levy, L., Lambeth, A., Solomon, R., & Gandy, M. (2018, August). Method in the Madness: The design of games as valid and reliable scientific tools. In *Proceedings*

of the 13th International Conference on the Foundations of Digital Games(p. 9). ACM.

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*.

Levy, L. M., Solomon, R., Moore, J., Way, J., Liu, R., & Gandy, M. (2015). Actions Speak Louder Than Words: An exploration of game play behavior and results from traditional assessments of individual differences. In *FDG*.

Lewis, J. D. (2001). *Language Isn't Needed: Nonverbal Assessments and Gifted Learners*.

Liebert, R. M., & Morris, L. W. (1967). Cognitive and emotional components of test anxiety: A distinction and some initial data. *Psychological reports*, 20(3), 975-978.

Linacre, J. M. (2000). Computer-adaptive testing: A methodology whose time has come. Chae, S.-Kang, U.–Jeon, E.–Linacre, JM (eds.): *Development of Computerised Middle School Achievement Tests*, MESA Research Memorandum, (69).

Lumsden, J., Edwards, E. A., Lawrence, N. S., Coyle, D., & Munafò, M. R. (2016). Gamification of cognitive assessment and cognitive training: a systematic review of applications and efficacy. *JMIR serious games*, 4(2).

- Martin, A. J., & Lazendic, G. (2017). Computer-adaptive testing: Implications for students' achievement, motivation, engagement, and subjective test experience.
- McCarthy, J. M., & Goffin, R. D. (2005). Selection test anxiety: Exploring tension and fear of failure across the sexes in simulated selection scenarios. *International journal of Selection and Assessment*, 13(4), 282-295.
- Morris, L. W., & Liebert, R. M. (1969). Effects of anxiety on timed and untimed intelligence tests: Another look. *Journal of Consulting and Clinical Psychology*, 33(2), 240.
- Moscoso, S., Salgado, J. F., & Anderson, N. (2017). How do I get a job, what are they looking for? *Personnel Selection and Assessment. An introduction to work and organizational psychology: An international perspective*, 25-47.
- Petersen, N. S., & Novick, M. R. (1976). An evaluation of some models for culture-fair selection. *Journal of Educational Measurement*, 13(1), 3-29.
- Rudner, L. M. (1998). An On-line, Interactive, Computer Adaptive Testing Mini-Tutorial, 11/98.
- Sarason, I. G. (1961). Test anxiety and the intellectual performance of college students. *Journal of Educational Psychology*, 52(4), 201.
- Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological bulletin*, 124(2), 262.
- Schmidt, F. L., Oh, I. S., & Shaffer, J. A. (2016). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 100 years of research findings.
- Schmit, M. J., & Ryan, A. M. (1992). Test-taking dispositions: A missing link?. *Journal of Applied Psychology*, 77(5), 629.
- Shute, V. J., Hansen, E. G., & Almond, R. G. (2008). You can't fatten A hog by weighing It—Or can you? evaluating an assessment for learning system called ACED. *International Journal of Artificial Intelligence in Education*, 18(4), 289-316.
- Shute, V. J., Ventura, M., Bauer, M., & Zapata-Rivera, D. (2009). Melding the power of serious games and embedded assessment to monitor and foster learning. *Serious games: Mechanisms and effects*, 2, 295-321.
- Shute, V. J., & Rahimi, S. (2017). Review of computer-based assessment for learning in elementary and secondary education. *Journal of Computer Assisted Learning*, 33(1), 1-19.
- Shute, V. J., & Ventura, M. (2013). *Stealth assessment: Measuring and supporting learning in video games*. MIT Press.
- Spielberger, C. D., & Vagg, P. R. (1995). Test anxiety: A transactional process model. In C. D. Spielberger & P. R. Vagg (Eds.), *Series in clinical and community psychology. Test anxiety: Theory, assessment, and treatment* (pp. 3-14). Philadelphia, PA, US: Taylor & Francis.

Spronck, P., Balemans, I., & Van Lankveld, G. (2012, October). Player Profiling with Fallout 3. In AIIDE.

Tekofsky, S., Van Den Herik, J., Spronck, P., & Plaat, A. (2013). Psyops: Personality assessment through gaming behavior. In Proceedings of the International Conference on the Foundations of Digital Games.

Tonidandel, S., Quiñones, M. A., & Adams, A. A. (2002). Computer-adaptive testing: The impact of test characteristics on perceived performance and test takers' reactions. *Journal of Applied Psychology*, 87(2), 320.

Turan, Z., & Meral, E. (2018). Game-Based Versus to Non-Game-Based: The Impact of Student Response Systems on Students' Achievements, Engagements and Test Anxieties. *Informatics in Education*, 17(1).

Volkow, N. D., Wang, G. J., Newcorn, J. H., Kollins, S. H., Wigal, T. L., Telang, F., ... & Wong, C. (2011). Motivation deficit in ADHD is associated with dysfunction of the dopamine reward pathway. *Molecular psychiatry*, 16(11), 1147.

Vos, L. (2015). Simulation games in business and marketing education: How educators assess student learning from simulations. *The International Journal of Management Education*, 13(1), 57-74.

Weiss, D. J. (1982). Improving measurement quality and efficiency with adaptive testing. *Applied psychological measurement*, 6(4), 473-492. Wise, S. L. (2015). Effort analysis: Individual score validation of achievement test data.

*Applied Measurement in Education*, 28(3), 237-252.

Weiss, D. J. (2004). Computerized adaptive testing for effective and efficient measurement in counseling and education. *Measurement and Evaluation in Counseling and Development*, 37(2), 70-84.

Wise, S. L. (2007). Examinee effort and test score validity. In annual meeting of the Northeastern Educational Research Association, Rocky Hill, Connecticut.

Wise, S. L., & DeMars, C. E. (2005). Low examinee effort in low-stakes assessment: Problems and potential solutions. *Educational assessment*, 10(1), 1-17.

Zeidner, M. (1998). *Test anxiety: The state of the art*. Springer Science & Business Media.





## About MindmetriQ

MindmetriQ is a series of gamified assessments which have been professionally validated. The MindmetriQ series is published by Test Partnership Ltd.

### **Test Partnership Ltd**

Mercury House,

London

SE1 8UL

+44 (0)20 7960 2897

**Web:** <https://www.testpartnership.com>

**Email:** [info@testpartnership.com](mailto:info@testpartnership.com)

