

Hommel, Björn E.; Ruppel, Regina; Zacher, Hannes

Article — Published Version

Assessment of cognitive flexibility in personnel selection: Validity and acceptance of a gamified version of the Wisconsin Card Sorting Test

International Journal of Selection and Assessment

Provided in Cooperation with:

John Wiley & Sons

Suggested Citation: Hommel, Björn E.; Ruppel, Regina; Zacher, Hannes (2021) : Assessment of cognitive flexibility in personnel selection: Validity and acceptance of a gamified version of the Wisconsin Card Sorting Test, International Journal of Selection and Assessment, ISSN 1468-2389, Wiley Periodicals, Inc., Hoboken, USA, Vol. 30, Iss. 1, pp. 126-144, <https://doi.org/10.1111/ijsa.12362>

This Version is available at:

<https://hdl.handle.net/10419/264569>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<http://creativecommons.org/licenses/by-nc/4.0/>

Assessment of cognitive flexibility in personnel selection: Validity and acceptance of a gamified version of the Wisconsin Card Sorting Test

Björn E. Hommel  | Regina Ruppel  | Hannes Zacher 

Leipzig University, Institute of Psychology—
Wilhelm Wundt, Leipzig University, Leipzig,
Germany

Correspondence

Björn Hommel, Institute of Psychology—
Wilhelm Wundt, Leipzig University,
Neumarkt 9-19, Leipzig 04109, Germany.
Email: bjorn.hommel@uni-leipzig.de

Abstract

The present study contributes to the emerging field of gamification in personnel selection by examining validity and acceptance of the Gamified Set-Shifting Task (GSST), which is based on a well-established neuropsychological test of cognitive flexibility, the Wisconsin Card Sorting Test (WCST). Results based on a sample of 180 participants in an online study provided preliminary support for construct and criterion-related validity. The GSST was better accepted among test-takers than both the WCST and a cognitive ability test. Overall, the findings suggest that the GSST may be an attractive and valid method to assist organizations in selecting employees who are able to adapt to changing environments.

KEYWORDS

cognitive ability, cognitive flexibility, gamification, personnel selection

Practitioner points

- Cognitive flexibility—adapting to new, changing, and unexpected circumstances—may become increasingly important in modern work environments.
- We redesigned a neuropsychological test into a gamified instrument for means of measuring applicants' cognitive flexibility.
- Game scores are positively associated with academic performance and self-reported adaptability.
- Study participants appeared to prefer the gamified tool over traditional means of personnel selection.

1 | INTRODUCTION

Recent changes in the nature of work require that employers reassess the modus operandi of their personnel selection procedures. In a fast-changing knowledge economy, employees need to adapt quickly to

novel demands, make decisions in the face of uncertainty, and cope with unexpected challenges (Pulakos et al., 2000). At the same time, employees are increasingly expected to switch seamlessly between different job roles, tasks, organizations, and even occupations (Eby et al., 2003). To ensure sustained firm performance, organizations

Björn Hommel and Regina Ruppel contributed equally to this study.

Paper for the Special Issue on "Game-Based and Gamified Assessments: Advances at the Frontier of Psychometrics" of the International Journal of Selection and Assessment.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2021 The Authors. *International Journal of Selection and Assessment* published by John Wiley & Sons Ltd.

need a workforce with the necessary capacities to efficiently deal with ongoing transformation (Pulakos et al., 2006). For this reason, adaptability and flexibility have been widely acknowledged as key transversal skills that play a vital role in the long-term success of employees and, in turn, organizations (Griffin & Hesketh, 2003). To keep up with the demands of today's dynamic and diverse workplaces, personnel selection researchers and practitioners need to reconsider what and how to assess in the 21st century (Ployhart, 2006).

Addressing the "what" question, research has tried to reveal the constructs that underlie adaptable and flexible behaviors (e.g., Ployhart & Bliese, 2006; Pulakos et al., 2006). Thereby, cognitive flexibility has been identified as one central capacity that may be particularly predictive in this regard (Laureiro-Martínez et al., 2009; Pulakos et al., 2006). *Cognitive flexibility* is the ability to adjust cognitive processing strategies in response to new, changing, and unexpected circumstances, conditions, and situations (Cañas et al., 2003). It is a central part of the executive functions that enable people to adapt successfully to environmental changes through mental set shifting and through overcoming automatic responses (Diamond, 2013). Cognitive flexibility has been recognized as one of the most important job skills as it helps individuals to efficiently deal with the challenges of the modern workplace (World Economic Forum, 2016).

Cognitive flexibility can be assessed with a variety of neuropsychological tests, the most prominent being the Wisconsin Card Sorting Test (WCST; Heaton et al., 1993). In this task, participants have to sort a series of cards according to different rules and alter their strategy when the rules change unexpectedly. Typically, individuals who are less cognitively flexible struggle to adjust to changing rules, while those with higher aptitude can quickly switch their mode of thinking between an efficiency-driven adherence to a given rule and the exploration of new approaches (Anderson et al., 1991). As this task was originally designed for clinical use to detect executive dysfunction, its suitability for assessing performance in a personnel selection context has yet to be investigated.

To evaluate the utility of a potential selection method, two aspects need to be considered: First, an assessment method must be a valid tool for predicting future work performance and, second, a selection test must be perceived favorably by job candidates (Moscoso et al., 2017). However, traditional selection tests often fail to meet both requirements of high validity and positive applicant reactions. Psychological tests, and cognitive ability tests in particular, are known to be the best predictors of future job performance (Schmidt & Hunter, 2004), but are often not well received by job candidates due to their abstract design that is not context-derived (Hausknecht et al., 2004; Kersting, 2008). How applicants perceive and react to the selection process significantly impacts critical outcomes, such as employer attractiveness and candidates' intentions to accept a job offer (for an overview, see Ryan & Ployhart, 2000). In general, organizations with a more positive reputation have a competitive advantage in attracting and selecting talented applicants (Smither et al., 1993). This is particularly relevant today as tight labor

markets in knowledge-based, service, and technical occupations give job seekers considerable choice between employers (Ployhart, 2006). Therefore, organizations not only need to reconsider what skills to assess within personnel selection, but also which selection methods to use.

To stand out and to offer candidates a more enjoyable selection experience, organizations have recently begun to implement novel technologies and innovative tools in their selection process. One of the biggest trends in this context is the use of gamified assessments (GAs) and serious games (Woods et al., 2020). In general, gamification refers to the integration of game design elements into a nongame context (Deterding, Sicart, et al., 2011). With regard to personnel selection, this entails enhancing traditional psychometric assessments by adding game design elements or developing stand-alone games that can measure job-relevant skills and abilities (Armstrong et al., 2016). These GAs are expected to be superior to traditional selection methods as they may increase the quality and accuracy of measurement, while improving the candidate experience (Landers, 2015).

The present study was designed to investigate a new GA of cognitive flexibility based on the WCST. The *Gamified Set-Shifting Task* (GSST) builds upon the fundamental principles of the WCST and includes typical game design elements like collecting points, a performance graph, and storytelling (Sailer et al., 2017). Moreover, the abstract content of the original task has been replaced with a work-specific simulation. Instead of sorting cards according to stimulus attributes (i.e., color, shape, object count), the GSST requires participants to match products to target groups according to their priorities (e.g., channel of communication, pricing, product design). Similar to the WCST, the gamified version aims to measure cognitive flexibility as it requires test-takers to switch between adherence to a sorting strategy that works and exploring another approach as soon as the target groups' priorities change unexpectedly. The objective of this study is to explore the suitability of this new GA as a potential selection method by examining construct and criterion-related validity. Additionally, we explore its acceptance amongst test-takers as compared to the traditional WCST and a cognitive ability test.

This study contributes to the literature by taking the first steps to examine a new GA that was designed to assist organizations in selecting the most talented employees for an era of change, while improving applicant perceptions at the same time. More generally, this study will provide insights into whether combining a well-established psychological test with modern game technology has the potential to create selection methods with incremental value above and beyond traditional assessment formats. Moreover, this study makes an important contribution to the literature on gamification in personnel selection. Despite their growing popularity and prevalence in practice, research on this topic is still scarce (Woods et al., 2020). Therefore, several scholars have emphasized the need for empirical investigation of gamified selection methods, in particular regarding their validity and applicant reactions (e.g., Armstrong et al., 2016; Fetzer et al., 2017). By examining the GSST with regard to both of

these aspects, this study complements the small body of evidence on the application of gamification in personnel selection.

2 | COGNITIVE FLEXIBILITY

Cognitive flexibility has been defined as the ability to adjust cognitive processing strategies in response to new, changing, and unexpected circumstances, conditions and situations (Cañas et al., 2003). It enables people to switch from one activity to another, to consider multiple perspectives, to find new solutions to a problem, and to face novel conditions in the environment (Dajani & Uddin, 2015; Ionescu, 2012). In contrast, individuals who are cognitively inflexible, struggle to adapt their strategies when situations change and, therefore, tend to get stuck in habitual patterns (Morris & Mansell, 2018). The ability to shift cognitive sets is a key property of efficient executive functioning and has been found to be different from cognitive abilities (Friedman et al., 2006). Research in the field of neuroscience has revealed that cognitive flexibility encompasses several cognitive processes, such as attention shifting, conflict monitoring, and knowledge representation (Cañas et al., 2006; Ionescu, 2012). The ability to flexibly update mental sets further builds upon inhibitory control and working memory and, accordingly, constitutes the basis for higher-level executive functions, such as reasoning, problem-solving, and planning (Diamond, 2013).

As working conditions become increasingly dynamic and complex in the information age, the ability to adapt thoughts and behaviors according to changing context requirements becomes particularly relevant for work success (Pulakos et al., 2006). In this regard, the World Economic Forum (2016) recognized cognitive flexibility as one of the most important contemporary job skills. In an era of fast change and uncertainty, individuals who are flexible in their mental processing can maintain high levels of performance, whereas a lack of this skill likely leads to stagnation and failure. Indeed, empirical studies found that performance in complex and dynamic tasks is not dependent on a person's repertoire of strategies or the ability to execute a strategy, but on the capacity to flexibly switch strategies in response to changing conditions (Reder & Schunn, 1999; Schunn & Reder, 2001). Cognitive flexibility has also been identified as a crucial managerial capability since it enables managers to deal with complex problems by switching between fast, automatic and more deliberate, conscious cognitive processing (Laureiro-Martínez & Brusoni, 2018; Laureiro-Martínez et al., 2009). In this context, cognitive flexibility has been referred to as "the ability of understanding when to rely on habits versus when to explore new courses of action" (Laureiro-Martínez & Brusoni, 2018, p. 1031). Moreover, empirical evidence demonstrated a link between higher levels of this capacity and improved strategic decision-making performance (Laureiro-Martínez & Brusoni, 2018).

The evidence described above highlights the significance of cognitive flexibility for individual job performance as well as its potential role regarding long-term success at an organizational level. To secure a sustainable competitive advantage, employers should

actively recruit and hire employees who not only possess professional expertise, but who are cognitively flexible at the same time. Although research on cognitive flexibility and related constructs, such as adaptability, also exist in the context of human resource management and the fields of organizational and vocational behavior (e.g., Ployhart & Bliese, 2006; Pulakos et al., 2000; Savickas & Porfeli, 2012), no gamified aptitude measurements have yet been established.

An easy and cost-efficient option for assessing cognitive flexibility in the context of employee selection would be the use of self-report-methods. While frequently used in behavioral research and practice, these formats suffer from various drawbacks, such as acquiescence (Bentler et al., 1971), social desirability bias (Edwards, 1953), or deceptive response styles. Donovan et al. (2003) found the latter to be highly prevalent in a study among job applicants, where 50% of the respondents admittedly exaggerated desirable traits, while 30% indicated that they had supplied completely false or fabricated responses. In contrast, performance-based measures may be the preferable option, as they do not rely on applicants' introspective abilities and are not prone to intentional distortions by "faking-good." There are several standardized neuropsychological tests to measure cognitive flexibility that are frequently used in clinical psychology. Typically, these tasks involve some sort of switching between tasks or principles according to changing rules. The oldest and most well-researched paradigm for this purpose is the WCST (Heaton et al., 1993), which is commonly used as a test of prefrontal cortex function. In this test, cards need to be sorted by color, shape, or object count. Through trial and error, test-takers have to determine the correct sorting rule based on feedback they receive (i.e., *correct/wrong*) and must flexibly switch strategy whenever the sorting rule changes (Stemme et al., 2007). Beyond cognitive flexibility, the WCST also taps into conceptual problem-solving skills, decision-making, and the ability to learn from feedback, as well as to modify wrong strategies and to inhibit incorrect action impulses (Heaton et al., 1993). Since the WCST was originally developed for clinical populations, it is unclear whether it would be appropriate for measuring performance among healthy individuals in the context of personnel selection. Possibly, this test would be too easy for this purpose, which could result in reduced motivation amongst applicants and limited discriminatory power in determining interindividual differences, as ceiling effects become increasingly likely (Laureiro-Martínez & Brusoni, 2018). Another potential drawback could be that job candidates usually react poorly to psychological tests in personnel selection as has been shown repeatedly in previous work (e.g., Gilliland, 1993; Hausknecht et al., 2004). This, in turn, can result in the withdrawal of applicants (Ryan et al., 2000), thereby diminishing the utility of a selection procedure. It appears that traditional methods often face tradeoffs between quality gains in measurement and losses in applicant acceptance (Benit & Soellner, 2012). By employing a gamification-approach, the study aims to overcome these limitations yielding a performance-based measure of cognitive flexibility with appropriate difficulty for the context of personnel selection.

3 | GAMIFICATION IN PERSONNEL SELECTION

Gamification is used as an umbrella term comprising a variety of techniques inspired by research in game design and generally refers to the integration of game design elements into nongame contexts (Deterding, Dixon, et al., 2011; Seaborn & Fels, 2015). The primary idea is to take advantage of the motivational nature of games to enhance the effectiveness of existing methods. By tapping into people's natural desire for competition and achievement, gamification promises to encourage participation, to increase productivity and, thus, to improve the quality and quantity of outcomes in any domain. Over the past few years, gamification has been increasingly applied within a variety of areas, including work, education, training, marketing, healthcare, wellness, and sustainability (Koivisto & Hamari, 2019; Seaborn & Fels, 2015). More recently, researchers and practitioners within the field of human resource management and industrial and organizational psychology have recognized gamification as a promising tool to improve recruitment and personnel selection. The central goal of using gamification within this context is to make the selection procedure more enjoyable while increasing the quality of measurement at the same time (Armstrong et al., 2016).

Game principles may be implemented in the employee selection process in various ways. For one, traditional selection methods, such as personality questionnaires and situational judgment tests, can be enhanced by integrating game design elements and are then referred to as *gamified assessments* (e.g., Georgiou et al., 2019; Landers et al., 2020). In contrast, serious games are mainly used in educational and training settings and aim to modify target behaviors rather than to assess trait-level information (Connolly et al., 2012; Sanchez & Langer, 2020).

In this article, we use the term GA as a general term for different kinds of assessments that include game design elements with the aim of improving traditional selection formats.

3.1 | Current prevalence of GAs in research and practice within personnel selection

To draw on the proposed benefits of GA, a growing number of large companies have started to implement game-like assessments in their recruitment process (Armstrong et al., 2016; Buil et al., 2020). Moreover, several third-party recruitment and assessment companies have developed a series of short games that intend to measure in-game performance and predict job-relevant behaviors. However, despite the rising popularity and use of gamified recruitment and selection in practice, empirical evidence regarding the validity and appropriateness of these new methods is still very limited (Armstrong et al., 2016; Woods et al., 2020). Hence, scholars have emphasized the need for published research about GAs in personnel selection so that research does not fall behind practice (Hawkes et al., 2017). It has been pointed out that future research should focus on criterion-related validity studies, in particular those investigating incremental

validity compared to traditional selection methods (Fetzer et al., 2017) as well as on applicant reactions (Armstrong et al., 2016; McCarthy et al., 2017).

3.2 | Criterion-related validity of GAs

It has repeatedly been suggested that GAs and serious games can improve the prediction of future work performance above traditional selection methods (e.g. Armstrong et al., 2016; Fetzer et al., 2017). First, GAs may be able to obtain higher quality information about job candidates compared to self-report methods. By mimicking work-related tasks, they can elicit and directly measure job-relevant behaviors instead of relying on accurate self-reflection of job candidates (Landers, 2015). Thus, similarly to work-sample assessments, they can reduce inferential leaps and improve the prediction of job performance (Fetzer et al., 2017). Second, the quality of measurement may also be increased compared to cognitive tests due to motivational benefits of gamification. Game features are expected to induce levels of engagement and immersion that are similar to a state of flow (Csikszentmihályi, 1990) and, thus, to distract applicants from the fact of being assessed (Fetzer et al., 2017). In this way, GAs may evoke less applicant tension typically associated with selection tests (Collmus et al., 2016). With decreased anxiety, the accuracy of assessments should be improved as test results are less contaminated in high-stake situations (Hausknecht et al., 2004). In line with this, empirical evidence suggests that GAs are associated with higher motivation and lower anxiety compared to traditional assessments (Gödöllei Lappalainen, 2017).

3.3 | Applicant reactions to GAs

Personnel selection can only be effective and financially worthwhile with a sufficient quantity of applicants (Ployhart, 2006). Thus, another important reason why organizations are using GAs in their selection process is to create a more fun and appealing experience for job candidates. In this way, companies hope to attract a wider pool of job candidates, enhance their employer image and positively influence applicants' job pursuit behaviors (Armstrong et al., 2016; Chow & Chapman, 2013; Fetzer et al., 2017).

The far-reaching significance of applicant reactions to the selection process has been investigated extensively within personnel selection research in the past decades. The realization has been that not only companies choose their employees, but job seekers also place ever higher demands on their future employers (Rynes, 1993). Gamification is a design method that has already been successfully applied within various work domains to make existing tasks more enjoyable (Cardador et al., 2017). In the same way, it is assumed that the implementation of game principles to assessment methods will provide a more positive experience for candidates (Landers, 2015). For example, it has indeed been shown that gamified surveys are rated as more enjoyable by respondents than the non-gamified

versions (Guin et al., 2012; Mavletova, 2015). However, it is important to consider that, in a high-stake context such as personnel selection, job candidates are not primarily seeking a fun and entertaining experience, but rather value criteria such as the perceived fairness of an assessment method. For this reason, researchers have asked for empirical studies that examine applicants' perception of gamified personnel selection methods more carefully (Armstrong et al., 2016; McCarthy et al., 2017).

To yield benefits that exceed traditional methods, new technologies for personnel selection like GAs must be developed in alignment with these findings from decades of applicant reactions research. Therefore, they should not only be created with the intention of making selection tools simply more enjoyable, but also ensuring that applicants will perceive them as fair and appropriate methods for selection purposes.

4 | DEVELOPMENT OF THE GAMIFIED SORTING TEST

The WCST was developed to test neurologically impaired patients and is therefore not an appropriate tool for personnel selection. Due to the profound differences between both target groups, those of patients and those of applicants, we deemed it necessary to develop a new paradigm for measuring cognitive flexibility. The objective of the present study is to investigate a newly developed GA with regard to its validity as well as test-takers' acceptance of it, thus addressing the present research-practitioner gap. The GSST incorporates several game design elements as defined within the taxonomy of game design elements by Sailer et al. (2017). First, the test was embedded in a narrative context of a fictive marketing agency where the player is welcomed as the new marketing director and is requested to carry out their first work task. We employed *storytelling* in the GA to evoke immersion, meaning, and explicit calls for action. Meaningful stories can inspire and motivate test-takers, especially when the story is aligned with their personal interests (Nicholson, 2015). Furthermore, by displaying *points* collected by the player, participants are provided with immediate feedback and reward (Sailer et al., 2013). Adding points to an assessment has been found to positively affect engagement, motivation, and overall performance (Papastergiou, 2009; Seaborn & Fels, 2015; Koivisto & Hamari, 2019). For a similar purpose, we added a *performance graph* to the GA which graphically represents a player's progression during the game. According to motivational theory, this promotes a mastery orientation, which facilitates learning and improvement (Sailer et al., 2013). Altogether, point systems, performance graphs, and narrative stories address human needs for achievement, competence, curiosity, and overcoming challenges that increase extrinsic as well as intrinsic motivation (Seaborn & Fels, 2015).

The GSST follows the same fundamental principles as the short version of the WCST (WCST-64; Kongs et al., 2000). Both tests involve the sorting of 64 objects according to changing rules. Participants are not told how to match them, but have to infer the correct

rule based on the feedback (i.e., wrong/right) they receive after each trial. After a certain number of consecutive correct trials, the rule changes without notice and subjects need to be flexible to adjust their strategy. Although, instead of matching cards to reference cards based on stimulus attributes (i.e., shape, color, object count) the GSST requires the matching of products to target groups according to their preferences (e.g., channel of communication, product category, pricing). Also, the GSST includes a few modifications aimed at increasing the task difficulty since it is designed for a nonclinical population. For this purpose, it entails an increased number of possible sorting rules (5 instead of 3) as well as a higher frequency of rule changes (after 7 instead of 10 consecutive correct trials).

We conducted a pilot study to explore the clarity of instructions, test difficulty, as well as test-takers reactions. In total, nine participants took part in this study. Target groups and products were printed out as paper cards and the game was carried out in a paper-based manner. The GSST results of the pre-study showed a high variability between the participants, thus indicating an appropriate difficulty level.

4.1 | Development of hypotheses

4.1.1 | Construct validity

To investigate convergent validity, WCST scores as well as GSST scores were compared with a self-reported adaptability measure. Adaptability is defined as the ability to adjust one's thoughts and behaviors to effectively respond to uncertainty, as well as new circumstances, conditions and situations (Martin et al., 2012). Both adaptability and (cognitive) flexibility are widely acknowledged as key competencies for today's employees (Griffin & Hesketh, 2003). The two constructs are often used in conjunction or interchangeably because their definitions overlap considerably and both lack a clear conceptualization in psychological literature (Pulakos et al., 2000). However, building on research within the cognitive sciences, it has been suggested that cognitive flexibility may be a cognitive micro-foundation that underlies human adaptability (Laureiro-Martínez & Brusoni, 2018). Flexible cognition may enable adaptive behavior in that it helps individuals to adjust their mental models following changes in the environment (Deák, 2004). Indeed, empirical evidence indicates that higher levels of cognitive flexibility relate to improved adaptive decision-making in well-structured and ill-structured problems (Laureiro-Martínez & Brusoni, 2018). Similarly, the WCST in the original and the gamified version involves making decisions in the face of inherent ambiguity and adjusting to changing demands, which is at the core of both cognitive flexibility and adaptability (Deák, 2004). Thus, we assume that there will be a positive relationship between WCST and GSST scores and adaptability.

Hypothesis 1 – Both the (a) WCST score and (b) GSST score relate positively to adaptability.

4.1.2 | Criterion-related validity

Next, we hypothesize that the WCST as well as the newly developed GSST have criterion-related validity in regard to two academic performance measures. Academic performance relates to job performance (Roth et al., 1996) and is usually measured with student grades or grade point average (GPA). A large body of literature demonstrates that the ability to switch flexibly between modes of thought relates to skills that overlap with or affect those required to succeed in an academic or work context. It has been shown that cognitive flexibility is associated with learning (Deák, 2004), multitasking (Ionescu, 2012), creativity (Figueroa & Youmans, 2013), leadership (Reiter-Palmon, 2003), problem-solving, and decision-making (Laureiro-Martínez & Brusoni, 2018), thus suggesting broad implications for academic and work achievement. As a classic neuropsychological test, the WCST has been primarily validated for its ability to detect executive dysfunction and barely for its utility as an ecological instrument to predict performance. However, a number of studies indicate that the WCST is predictive of academic achievement (Kercood et al., 2017), work behavior (Ready et al., 2001), work-related skills, and occupational status (Kibby et al., 1998).

The WCST and the GSST require test-takers to recognize and follow abstract rules, learn from feedback, and flexibly adjust their strategy when the rules change unexpectedly. In this way, this task might simulate some of the fundamental characteristics of today's complex and dynamic environments and tap into the same abilities that are necessary to thrive in the face of changing demands. Students constantly need to shift between subjects, come up with different approaches to solve complex problems, and choose appropriate strategies for various assignments. For this reason, the assumption can be made that individuals who struggle to be flexible in this way in real-life situations will experience comparable difficulties when performing the WCST as well as in the GSST.

Hypothesis 2 – *Both the (a) WCST score and (b) GSST score relate positively to academic performance.*

4.1.3 | Acceptance

Besides validity, applicant reactions are the second important aspect that determines the utility of a selection method (Moscoso et al., 2017). Meta-analytic findings have shown that psychological tests, and cognitive ability tests in particular, are often categorized as unfavorable in the perspectives of applicants due to their abstract design (Gilliland, 1993; Hausknecht et al., 2004). It is expected that gamified selection methods will improve applicant reactions (e.g., Armstrong et al., 2016; Landers, 2015). Thus, we assume that test acceptance for the newly developed GA will be higher than for the WCST and a cognitive ability test.

First, the GSST includes points, a performance graph, and storytelling as game elements that relate to a player's desire for competence and achievement, which in turn increase extrinsic as well

as intrinsic motivation (Sailer et al., 2017; Seaborn & Fels, 2015). Empirical work investigating applicant reactions to GAs is still limited (Armstrong et al., 2016). However, a number of studies have demonstrated that the implementation of game elements to a survey leads to more positive test-taker evaluations as compared to the original version (Guin et al., 2012; Mavletova, 2015). Moreover, the GSST simulates a work scenario in which individuals are required to perform a job-related task. In this way, this task may not only be more engaging and fun, but participants may also perceive it to be more suitable and fairer for the purpose of selection. In contrast to the WCST, the GSST no longer involves sorting cards but the matching of products to target groups, integrated in a fictive scenario within a marketing agency. Thus, the GSST replicates a real-life work scenario and may appear more closely related to actual work requirements. Recent studies have shown that simply changing the abstract design of a GMA test to a work-related design resulted in higher face validity (Benit & Soellner, 2012; Krumm et al., 2011). In turn, higher levels of face validity and job-relatedness have proven to be the most important aspects that drive overall fairness perception and test acceptance (Hausknecht et al., 2004; Ployhart, 2006). Consequently, we assume that the new GA will be rated more positively than the original WCST and a cognitive abilities test.

Hypothesis 3 – *Test-takers' acceptance for the GSST will be higher than for the WCST and a cognitive abilities test.*

We further hypothesize that the congruency between the GSST and the advertised job position will increase test acceptance. Congruency is given when the story-context of the GSST matches the domain of the advertised position (i.e., marketing). Previous work has shown that selection methods that are clearly relating to a specific job position are rated most favorably (Hausknecht et al., 2004). As the assessment content of the newly developed GA is embedded in a fictive marketing context, we assume that its test acceptance will be highest when it is used as a selection test with regard to a position in the area of marketing.

Hypothesis 4 – *Congruency between the GSST and the advertised job position will positively relate to test acceptance.*

5 | METHOD

5.1 | Participants and procedure

Data for this study came from 181 participants in Germany. One participant was excluded because their self-reported grades were based on a different grading system and thus their academic performance measures could not be used for the analysis. The final sample included 123 (68.3%) female and 57 (31.7%) male participants. Mean age was 27.15 years (SD = 7.58) and ranged from 19 to 58 years. More specifically, 144 participants (80%) were 30 years or younger, and 36 participants (20%) were older than 30. In terms of

highest level of education achieved, one (0.6%) participant had a general education degree, one (0.6%) had a middle school degree, four (2.2%) had completed vocational training, 96 (53.3%) had a high school degree, 76 (42.2%) had a university degree, and two (1.1%) indicated to have another not further specified degree. Regarding their current state of employment, the largest part of the sample ($N = 145$) were students, the majority of whom were enrolled in undergraduate or postgraduate psychology ($N = 113$). The remaining sample included 20 (11.1%) employees, two (1.1%) civil servants, six (3.3%) self-employed, three (1.7%) unemployed participants and four (2.2%) who indicated "other" as their current state of employment.

Data was collected over a 12-week period via a web-based survey. Participants were recruited via online advertisement posted on social networking websites such as Facebook, LinkedIn, Xing, and Social Science Survey. Students enrolled in psychology courses received ungraded course credit for their participation.

5.2 | Measures

5.2.1 | The Wisconsin Card Sorting Test

The WCST used in this study was an adaptation, based on the WCST from PsyToolkit (Stoet, 2017; www.psytoolkit.org). Since the available version of the WCST on this platform slightly differs from the traditional principles of the WCST (Heaton et al., 1993) two adjustments were made, so that the administration of the online WCST was equivalent to the original test. The number of trials was increased to a number of 64 (the WCST version on PsyToolkit only has 60 trials) and the changing of selection criteria was adjusted so that the rule changes every time the test-taker has completed 10 consecutive correct trials (in the WCST on PsyToolkit the classification rule changes every 10 cards). Against a black background, the four reference cards (one red circle, two green triangles, three blue crosses, and four yellow stars) were displayed at the top of the screen (see Figure 1). At the bottom of the screen, the response cards were presented one at a time. Participants were instructed to match a series of response cards with one of the four reference cards without being told how to sort them. Through trial and error, subjects had to infer the correct sorting criterion (object count, color, or shape) based on the feedback (correct/incorrect) they received after each trial. After 10 consecutive correct selections the sorting criterion changed without warning and participants had to learn and apply the new rule to sort the cards. The final test score consisted of the total number of correct trials, which represents the ability to recognize and adhere to a sorting rule as well as the flexibility to switch to a new rule after the sorting criteria changes. Incorrect attempts were penalized by covertly subtracting a point from the participant's score, without displaying the current score to the participant. Thus, higher scores indicate higher levels of cognitive flexibility. In total, there are six different scores that can be derived from this task; however, research on the psychometric properties of this test found that there is

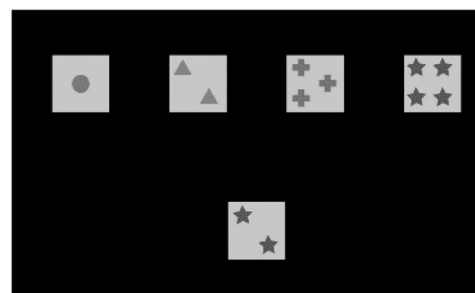


FIGURE 1 The Wisconsin Card Sorting Test

substantial redundancy between them and therefore their calculation is unnecessary (Bowden et al., 1998).

5.2.2 | The Gamified Set-Shifting Task

As described earlier, the GSST follows the same fundamental principles as the original WCST but incorporates game elements (i.e., points, performance graph, storytelling), a higher degree of difficulty as well as a work-related design. Instead of matching cards according to the attributes shape, color and object count, test-takers had to match products to target groups according to common characteristics.

The game starts with a fictive employee of a marketing agency welcoming the participant in the team and congratulating them to their new position as marketing manager. He instructs the subject to the goal of implementing a new marketing strategy to reduce costs and improve the efficiency of marketing campaigns for consumer products. After this brief introduction, the game interface is introduced (see Figure 2). The consumer products are presented in a market stand one at a time on the right side of the screen and five target groups are represented by five avatars. Participants start with a budget of €10,000 and are tasked with maximizing profits by correctly assigning products to one of the five target groups. As in the WCSTs, the correct matching rule is not revealed to the subject. Each target group is receptive to one attribute of the consumer product (e.g., pricing) until the target attribute changes. At first, it would be plausible to match a product to any of the five avatars. The participant has to infer which of the attributes the target groups currently are receptive to, based on the feedback they receive after each trial. Each time a consumer product is matched in accordance with the current underlying rule, visual and auditory reward cues are played (i.e., an exploding confetti cannon, accompanied by the sound of a fanfare) and the participant is notified that their balance increased by €500. Unsuccessful trials are penalized by subtracting €500 from the players' score. After seven correct consecutive matches, the underlying rule changes and the participant has to adapt by abandoning the previous strategy in favor of a new approach. The current game score (account balance) is displayed at the bottom right of the screen during the entire gameplay, as well as a performance graph that indicates the progress over previous trials. After completing all

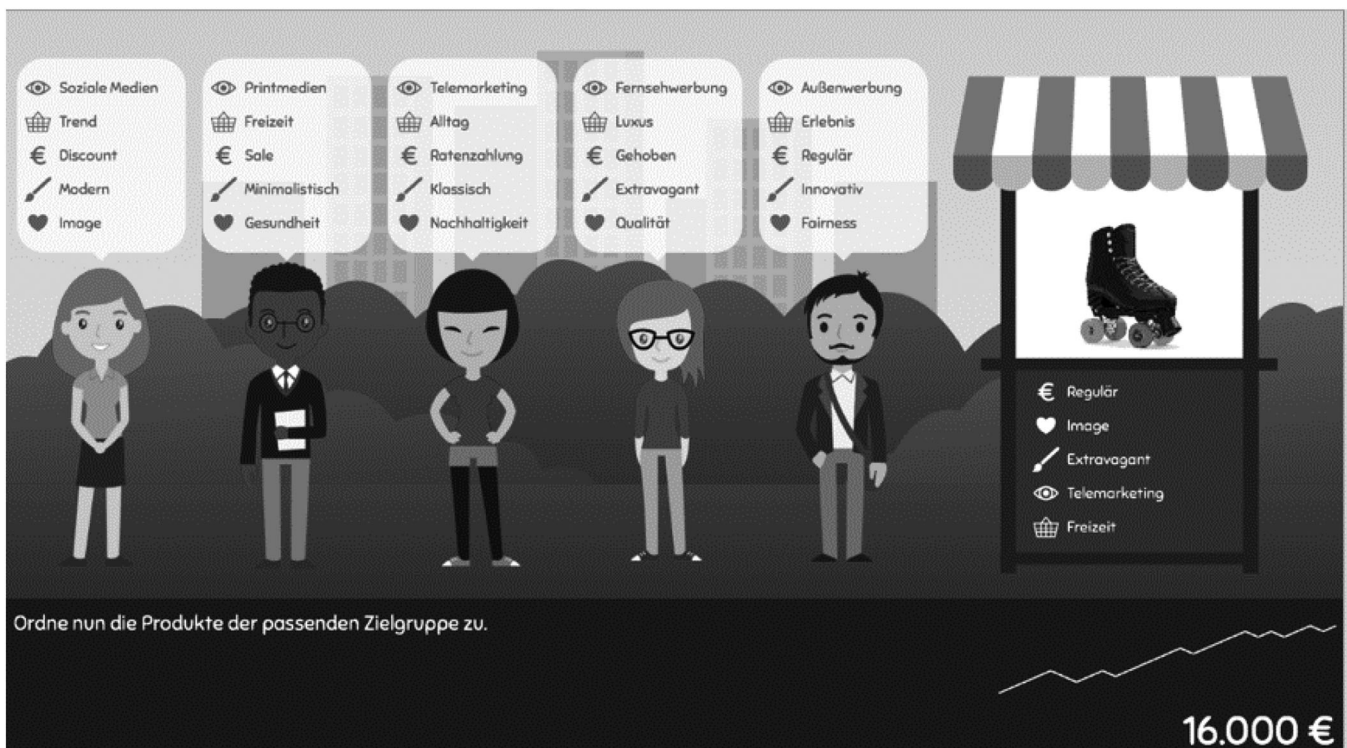


FIGURE 2 Game interface of the Gamified Set-Shifting Task. Participants were instructed to match 64 products one after another to one of the five target groups. Test-takers had to infer the correct sorting rule based on the feedback from previous trials. The current game score as well as a performance graph is displayed in the bottom right corner

64 trials, the final game score is displayed. Analogous to the WCST, the final game score is based on the total of all correct and incorrect trials, and thereby operationalizes the ability to detect and respond to rule changes by adopting new strategies. Note that the scoring mechanism for the GSST and the WCST is the same and, although instrument scores differ in scaling, both are the result of a linear function and thus translatable.

5.3 | Cognitive abilities

We included matrix reasoning items from the International Cognitive Skill Database (ICAR; <http://www.icar-project.com>; Condon & Revelle, 2014). These items contain similar stimuli as those used in Raven's Progressive Matrices, addressing abstract reasoning, and consist of 3×3 arrays of geometric shapes with one of the nine shapes missing. Test-takers are instructed to identify, which of the six possible response choices will correctly complete the matrix. For each of the items, participants were asked to choose the right answer from seven response choices, including the option "I don't know." We selected eight out of the 11 available items based on varying levels of item difficulty. Test-takers were instructed to complete this section within a time limit of 10 min. Due to low inter-item correlation and the insufficient reliability ($\alpha = .60$) in our sample, we refrained from using data from this instrument in further analyses. Nevertheless, we used data from test takers' acceptance ratings for this test.

5.3.1 | Acceptance

Participants' acceptance of the assessment methods was measured with 15 items adopted from a questionnaire developed by Smither et al. (1993) to examine applicant reactions to selection procedures. For this study, the original items were translated into the German language (and back-translated into English by a bilingual person to increase translation accuracy; see Brislin, 1970) and, wherever necessary, the word game was substituted for the word examination. The questionnaire includes items measuring five different aspects of applicants' reactions to selection methods. The five subscales as well as item examples are listed in Table 1. The original items used (Smither et al., 1993), as well as the German translation can be found in Appendix A. Participants were presented with the following instruction: "You have just undertaken a personnel selection test. Now we are interested in your opinion. Please rate how strongly you agree or disagree with the following statements." The items were rated on five-point Likert scales ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). Cronbach's α for the overall scale was .90.

5.3.2 | Adaptability

To examine construct validity, the Adaptability Scale (Martin et al., 2012) was translated into German and back-translated into

Subscale	Items	Example item
Face validity	5	The actual content of the examination was clearly related to the job
Perceived predictive validity	5	I am confident that the examination can predict how well an applicant will perform on the job
Affective reaction	2	I enjoyed the examination to a great degree
Procedural justice	2	Overall, I believe that the examination was fair
Recommendation	1	Based on my experience with the examination I would encourage others to apply for this job

TABLE 1 Acceptance questionnaire: subscales and example items

English by a bilingual person to increase translation accuracy (see Brislin, 1970); the German version was used in this study (see Appendix B). In support of structural validity, parallel analysis (Horn, 1965) suggested a single component structure. Principal component analysis with oblique rotation indicated good factor loadings ($N = 180$; $\lambda \geq 0.63$). This questionnaire consisted of nine items measuring cognitive (e.g., "I am able to think through a number of possible options to assist me in a new situation"), behavioral (e.g., "I am able to seek out new information, helpful people, or useful resources to effectively deal with new situations"), and emotional (e.g., "I am able to reduce negative emotions [e.g., fear] to help me deal with uncertain situations") adaptability. Participants were asked to indicate their agreement on Likert-scales from 1 (*strongly disagree*) to 7 (*strongly agree*) for each item. Previous work has indicated validity for the one-factor scale through confirmatory factor analysis, adequate levels of reliability, and invariance in measurement properties across key participant subgroups (e.g., gender, age; Martin et al., 2013). Cronbach's α for this scale was 0.87.

5.3.3 | Academic performance

For each participant, academic performance was assessed with two measures. First, all participants had to indicate the final grade with which they had graduated from school. Second, working participants were asked about the final grade of their professional or academic education, whereas students had to indicate their current GPA. They were instructed to give an honest and precise estimation of their GPA in case the actual GPA is unknown to them. The grades awarded in the German education system range from 1.0 (*very good*) to 6.0 (*insufficient*).

5.3.4 | WCST experience

WCST experience was measured with a single question asking whether the participant has already had experience with the Wisconsin Card Sorting Task. Participants had to select whether they have never heard about this task (1), have already heard about it (2) or have heard about and already performed this task (3).

5.3.5 | Computer experience

Computer experience was measured with a single question asking subjects to rate their level of experience with computers compared to people of the same age. Participants responded to this item on a scale from 1 (*very low*) to 7 (*very high*).

5.3.6 | Online game experience

Online game experience was measured with a single question asking subjects to rate their level of experience with online games compared to people of the same age. Participants responded to this item on a scale from 1 (*very low*) to 7 (*very high*).

5.4 | Study design and procedure

After accessing the study by clicking on the survey link participants were informed about the purpose and procedure of the study and were asked to confirm their voluntary participation. Next, they answered questions concerning demographic data and academic performance, and were then instructed to imagine that they were applying for a job and had to go through a selection process in which they would have to perform two different selection tests. All of the participants then completed the cognitive ability test as the first assessment method and were subsequently asked to indicate their acceptance of this test. The second selection method varied between the subjects. One half of the participants performed the WCST and the other half completed the GSST. In both cases, subjects were asked to evaluate the test afterwards. Finally, all the participants completed the Adaptability Scale and answered the questions regarding prior WCST, computer and gaming experience. Hypothesis 5 stated that the GSST will be judged more favorably by participants if its context appears relevant for a given purpose. To test this assumption, participants were told to apply for a position as a Marketing Manager or as a Customer Service Manager. Since the GSST simulates an environment in which the player is requested to make strategic marketing decisions, the former condition is *congruent* with the GSST. We therefore expected lower acceptance for the Customer Service Manager position, which is *incongruent* with the setting of the GSST.

Thus, the study followed a 2 (WCST/GSST) \times 2 (congruent/incongruent) \times 2 (cognitive ability test/cognitive flexibility test) design with the first and the second factor being between-subject factors and the latter a within-subject factor. The allocation to one of the four resulting experimental conditions occurred randomized after the participants had accessed the survey.

5.5 | Statistical analyses

To test construct validity, we conducted a correlational analysis between WCST score and adaptability as well as between GSST score and adaptability. Concerning the outcome variables, *current grade* was created as a composite academic performance measure, including self-reported current GPA from the $n = 145$ students and self-reported final grades of their professional or academic education from the $n = 35$ nonstudent participants. As a result, we obtained two academic performance measures (final school grade and current grade) for each participant and investigated correlations of these measures with WCST score and with GSST score to determine the criterion-related validity of both tests.

Regarding test acceptance, we conducted one-tailed t tests to compare average acceptance levels of the different assessment methods (cognitive ability test, WCST, GSST) within and between participants. To investigate the effect of gamification on acceptance we compared GSST acceptance and WCST acceptance between subjects as well as GSST acceptance with cognitive ability test acceptance within subjects. Finally, to examine the effect of congruency on acceptance, we computed a one-tailed t test, comparing GSST acceptance for the congruent study condition with GSST acceptance in the incongruent study condition. The data that support the findings of this study are openly available in the Open Science Framework at <https://osf.io/jcdvq/>.

6 | RESULTS

6.1 | Descriptive statistics and correlations

Means, standard deviations, intercorrelations, and reliability estimates for the surveyed variables are provided in Table 2. Regarding academic performance measures, subjects indicated on average very good final school grades ($M = 1.74$, $SD = 0.66$) and current grades ($M = 1.78$, $SD = 0.58$). Both academic performance measures were significantly correlated with each other ($r = .35$, $p < .001$).

We observed that GSST score related positively to GSST acceptance ($r = .15$, $p = 0.015$), indicating that participants who performed better also rated the test more favorably. Furthermore, cognitive ability test acceptance correlated positively with WCST acceptance ($r = .69$, $p < .001$) and with GSST acceptance ($r = .27$, $p < .001$). As the three control variables, WCST experience, computer experience, and gaming experience did not significantly relate to neither GSST performance nor GSST acceptance, they were not included in further analyses.

Although participants were randomly assigned to the four study conditions, we checked for differences between the four groups in the demographics, before testing the hypotheses. The initial analyses demonstrated that the four groups were statistically similar as there were no significant differences in age, $F(1, 178) = 0.23$, $p = .635$, gender, $F(1, 178) = 0.00$, $p = .985$, highest level of education, $F(1, 178) = 0.01$, $p = .908$, or current employment status, $F(1, 178) = 0.02$, $p = .889$, between the four groups. Therefore, demographic variables were not used as controls in subsequent analysis.

6.2 | Test of hypotheses

6.2.1 | Construct validity

Consistent with Hypothesis 1a, results of a one-tailed Spearman's rank correlation test revealed a significant positive association between GSST score and adaptability ($r_s = .28$, $p = .003$). Contrary to Hypotheses 1b, we found no significant correlation between WCST score and adaptability ($r_s = -.10$, $p = .820$).

6.2.2 | Criterion-related validity

Hypothesis 2 stated that both GSST score and WCST score would positively relate to academic performance as measured by self-reported final school grade and self-reported current grade. One-tailed Spearman correlation showed that GSST score was significantly related to final school grade ($r_s = -.22$, $p = .018$) and to current grade ($r_s = -.21$, $p = .025$), thus supporting criterion-related validity of the GSST. Note that a negative correlation between the GSST scores and the academic performance measures represents a positive relationship since lower grades reflect better academic achievement in the German grading system. Against our assumptions, we found that WCST score was neither significantly related to final school grade ($r_s = -.17$, $p = .060$) and nor to current grade ($r_s = -.12$, $p = .130$).

6.2.3 | Acceptance

Next, we explored differences across the different methods in terms of their acceptance amongst participants. Figure 3 shows the mean acceptance of the cognitive ability test, the WCST and the GSST. Consistent with Hypothesis 3, a one-way ANOVA revealed differences on average acceptance ratings between WCST, GSST and the cognitive ability test [$F(2, 359) = 32.48$, $p < .001$]. Rated acceptance was highest for the GSST ($M = 3.13$, $SD = 0.72$), lower for the WCST ($M = 2.49$, $SE = 0.08$) and lowest for the cognitive ability test ($M = 2.37$, $SD = 0.74$). Results indicated that acceptance of the GSST was significantly higher than for the cognitive ability test [$t(89) = 7.80$, $p < .001$, one-tailed], which represented a large effect ($d = 0.82$). In addition, results showed that the difference between acceptance of the GSST and the WCST became also significant

TABLE 2 Means (M), SD, and intercorrelations of study variables

Variables	M	SD	N	1	2	3	4	5	6	7	8	9	10	11	12	13
1. WCST score	37.64	18.67	180	-												
2. GSST score	7.066	13.154	90	NA	-											
3. Adaptability	5.01	0.85	90	-0.10	0.28**	(0.87)										
4. Final school grade	1.79	0.50	180	-0.17	-0.22*	0.07	-									
5. Current grade	1.77	0.50	180	-0.12	-0.21*	-0.01	0.35***	-								
6. Acceptance cognitive ability test	2.37	0.74	180	-0.04	-0.20	0.10	0.02	0.13	(0.90)							
7. Acceptance WCST	2.47	0.73	180	0.06	NA	0.09	-0.10	-0.04	0.67***	(0.88)						
8. Acceptance GSST	3.13	0.78	90	NA	0.22*	0.01	0.09	0.06	0.26*	NA	(0.88)					
9. Age	27.15	7.58	90	-0.30**	-0.19	0.18*	0.49***	0.15*	0.03	-0.25*	-0.09	-				
10. Gender	0.32	0.47	180	-0.17	-0.10	0.10	0.22**	0.05	0.15*	0.01	-0.04	0.24**	-			
11. WCST experience	1.39	0.58	180	0.20	-0.09	-0.03	-0.06	-0.11	-0.01	-0.05	-0.09	0.08	-0.07	-		
12. Computer experience	4.68	1.17	180	-0.06	-0.03	0.18*	0.13	0.05	0.14	0.08	0.20	0.15*	0.28***	-0.09	-	
13. Gaming experience	3.46	1.77	180	0.08	0.11	0.08	0.13	0.02	0.12	0.19	0.03	0.07	0.33***	0.04	0.57***	-

Note: Cronbach α s, where available, are shown in parentheses along the diagonal. Since test scores (cognitive ability, WCST, and GSST) as well as academic performance measures (final school grade and current grade) were nonparametric, Spearman correlation analyses were conducted and displayed in this table. GSST scores have been converted into WCST scores for a better comparability. For gender, 0 = female, 1 = male.

Abbreviations: GSST, Gamified Set-Shifting Task; NA, correlations cannot be computed because there are no pairs for these variables; WCST, Wisconsin Card Sorting Test.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

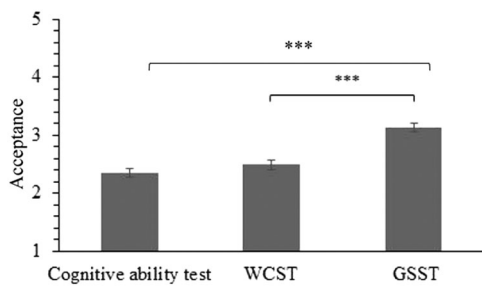


FIGURE 3 Acceptance for different assessment methods

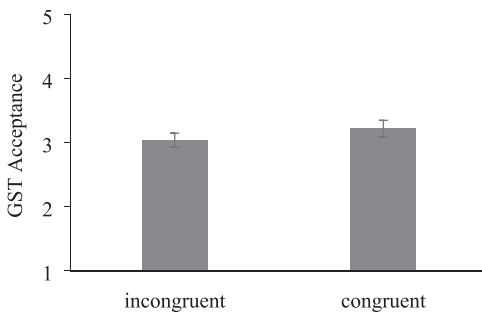


FIGURE 4 GSST acceptance for different study conditions. GSST, Gamified Set-Shifting Task

[$t(176.74) = 5.75, p < .001$, one-tailed], and constituted a large effect ($d = 0.86$). Thus, Hypothesis 3 was supported by the results.

Note. Mean acceptance for the cognitive ability test ($N = 180$), the WCST ($N = 90$) and the GSST ($N = 90$). Scale ranges from 1 to 5. Error bars represent standard errors.

According to Hypothesis 4, congruency would have a positive impact on GSST acceptance. As illustrated in Figure 4, mean GSST acceptance in the congruent condition was greater ($M = 3.22, SD = 0.13$) than in incongruent condition ($M = 3.04, SD = 0.11$). This divergence represented a small effect ($d = 0.22$), however the difference between the two conditions was not statistically significant [$t(85.34) = 1.06, p = .147$]. Therefore, our findings did not support Hypothesis 4.

Note. Mean GSST acceptance in the incongruent study condition ($N = 44$) compared to the congruent condition ($N = 46$). Scale ranges from 1 to 5. Error bars represent standard errors.

7 | DISCUSSION

7.1 | Summary and interpretation of results

The goal of this study was to explore the appropriateness of the Gamified Set-Shifting Task as a personnel selection method. Drawing on the fundamental principles of the WCST, the GSST requires test takers to continuously develop, apply and abandon successful problem-solving strategies in a job-related context. In that way, this GA is designed to measure cognitive flexibility.

The study findings demonstrated that GSST scores related to self-reported adaptability and two academic performance measures, thus providing first evidence for its construct and criterion-related validity. The results further suggested that the GSST was more accepted among test-takers than the WCST and the cognitive ability test. This study contributes to the literature by examining a GA for personnel selection with regard to both its validity as well as test takers' reactions, thus providing important evidence for research and practice in the field.

Consistent with expectations, GSST scores related positively to self-reported adaptability, thereby offering preliminary support for the instruments construct validity. Adaptability was included in this study to investigate convergent validity based on the premise that cognitive flexibility is closely related to or may be the predictor of the ability to adapt to changing environmental demands (Laureiro-Martínez et al., 2009). Contrary to our assumptions, the association between WCST scores and adaptability was not significant. A potential explanation for this may be that, because of its intended target domain, the WCST was not able to capture facets of more complex adaptive behaviors in healthy individuals that is shown in nonclinical settings. However, it is also possible that the adaptability measure was not adequate to determine construct validity in this context. Future research is needed to support construct validity of the GSST, in particular a cross-validation study comparing results from the WCST and this gamified adaptation.

In line with expectations, results showed that GSST scores were associated with both academic performance measures, final school grade and current grade. That is, individuals who achieved higher scores in the GA had on average better grades than those with lower scores. These findings support criterion-related validity and provide a first indication that the new GSST may be a valuable method to predict future performance outcomes. Again, contrary to predictions, WCST scores did not significantly relate to final school grade and current grade. Overall, our findings on the WCST show that ceiling effects were of greater magnitude than we had initially anticipated, which suggests that the WCST is not suitable for use in the context of personnel selection.

As already suspected in previous work (Laureiro-Martínez et al., 2009), the WCST might fail at discriminating among a population of healthy individuals since it had been developed as a test for clinical use and thus may be too easy for a nonclinical sample. In contrast, the GSST might be suitable for a selection context as it involves several modifications to increase the task difficulty. This was also reflected by study results showing that participants achieved on average lower scores in the GSST than in the WCST. On the other hand, earlier studies were able to demonstrate a relationship between WCST performance and performance in academic (Kercood et al., 2017) and work contexts (Kibby et al., 1998; Ready et al., 2001). It has to be noted that the present study sample comprised a large proportion of psychology students, who typically obtain very good grades in Germany. Thus, it may also be the case that the relationship was obscured by low variance in our measure of academic performance. Future research with more diverse samples and

additional performance measures may provide a better understanding on the predictive power of the original WCST and its gamified adaption in a selection context.

With regard to the second research goal, the study results showed that overall acceptance of the GSST was superior to the original WCST and the cognitive ability test. That is, participants rated the GSST better in terms of face validity, perceived predictive validity, justice perception, affect, and recommendation to others. Meta-analytical findings demonstrated that psychological ability tests are rather unpopular among applicants due to their tedious nature and abstract construction without any visible connection to future work requirements (Hausknecht et al., 2004; König et al., 2010). Gamification has already been applied successfully to different work areas to make tedious tasks more enjoyable and to improve task performance (Koivisto & Hamari, 2019; Seaborn & Fels, 2015). Similarly, GAs have recently been recognized as promising tools that may replace unpleasant traditional assessment formats and increase candidate experience within a personnel selection context (Armstrong et al., 2016; Woods et al., 2020). Particularly, presenting assessment content as a game that mimics a work-related task has been expected to bring several potential advantages such as improved applicant reactions (Landers, 2015). The present study contributes to the literature by empirically examining these assumptions, thus providing evidence that test-takers are perceiving a GA indeed more positively than traditional ability tests.

Finally, the data in this study did not conform to our expectation, that participants would rate the GSST more favorably when congruency between the selection method and a fictitious vacant position was communicated to the test takers. This finding disagrees with earlier work (e.g., Hausknecht et al., 2004) and might imply that the GSST generally evokes positive test-takers' reactions, independently of its field of application. From a practical point of view, this would entail an advantage as it might not be necessary to modify the GA for use in different occupational contexts. The fact that we found no effect of congruency on test acceptance may also be due to methodological weaknesses of the present research. In this study, participants were simply instructed to imagine that they were either applying for a job as a marketing manager or for a job as a service manager. It is possible that participants might have rated the GSST as significantly different if the manipulation had involved a more detailed and realistic description of the different job positions and requirements, for example like a typical job advertisement. Future research will need to examine whether a job-specific construction of the GA yields enough benefits such as acceptance gains to outweigh the disadvantages of a limited field of application.

An additional finding that is noteworthy is that prior computer and gaming experience in this study did not relate to neither GSST score nor GSST acceptance. In contrast, a recent study found that a gamified situational judgment test only had a positive effect on test-takers' perceptions for individuals with a high level of video gaming experience (Gkorezis et al., 2020). Further research into gamified tests found that computer game experience affected participants' performance, in that those with high prior gaming experience had an

advantage in achieving good test results (Burt et al., 2018; Kim & Shute, 2015). This posits a risk for GAs because in these cases, test scores are conflated and job candidates with little or no prior gaming experience may perceive the use of GAs in a high-stakes context as unfair (Armstrong et al., 2016). Therefore, recent studies have called to take the impact of technology usage into account when investigating GAs for personnel selection (e.g., Armstrong et al., 2016; Nikolaou et al., 2019). The present study addressed this concern and found that neither GSST performance nor acceptance were dependent on participants' familiarization and experience with such methods. Additional research will be necessary to gain a deeper understanding about when previous technological expertise may influence relevant outcomes to avoid possible adverse impacts or disadvantages for certain individuals when using GAs.

7.2 | Limitations and future research

This study has several limitations that need to be addressed in future research. First, the present research design did not allow for a direct comparison between the GSST and the established WCST, as participants completed merely one version. Although the relationship of GSST scores with self-reported adaptability provides a first indication of construct validity, the congruency with the original WCST has yet to be examined. To fully validate the GSST as a measure of cognitive flexibility, it will be necessary to investigate convergence with the WCST in a follow-up cross-validation study. Moreover, the translated acceptance and adaptability scales should undergo additional validation before they are used in future research.

A further limitation concerns the criterion variables in this study. Since we anticipated the majority of the sample to be students and not yet in employment, we used academic performance measures as criterion variables instead of job performance measures. Although academic achievement is highly related to future job success (Roth et al., 1996), the predictive power of the GSST with regard to relevant future work-related outcomes needs further investigation. Also, we relied on self-reported performance measures in this study due to the unavailability of participants' actual grades. To gain more robust findings about the predictive validity of the GSST, future research should use critical job performance outcomes, ideally including objective performance data or supervisor ratings. Taken together, this initial study requires support from further examinations of both construct and criterion-validity to ensure that the measurement target, cognitive flexibility, is rigorously assessed and indicative of work-related outcomes.

Next, some restrictions need to be considered when interpreting the results regarding the acceptance of the GSST. For one, there is considerable variability in game-like selection methods, ranging from traditional assessments that have been enhanced with single game design elements to full-fledged assessment games (Armstrong et al., 2016). Thus, the present study results may not apply to GAs in general. In addition, our sample consisted predominantly of undergraduate students who may lack the experience necessary to assess

the appropriateness of selection methods for means of performance prediction. This study could demonstrate that the GSST was more accepted by test-takers than the WCST and a cognitive ability test, but it cannot pinpoint what individual features contributed to this positive effect. As the GSST involves multiple modifications of the original WCST, it is unclear whether the higher test-acceptance can be attributed to its game elements, its work-related game content and graphic, or to a combination of these. To acquire a deeper understanding about the underlying mechanisms, future research could systematically vary each implementation and carefully examine which elements exhibit measurable effects on acceptance. Such studies may result in fruitful contributions to the development of stronger theoretical frameworks of gamification and GAs. Second, this study focused only on a limited set of applicant reactions which are considered to be amongst the most relevant aspects based on prior meta-analytical findings (e.g., Hausknecht et al., 2004). Future studies could extend the study results by examining other factors affecting GA acceptance like chance to perform or feedback (Bauer et al., 2001) and also secondary outcomes such as organizational attractiveness (Highhouse et al., 2003). Researchers within the field of industrial and organizational psychology have already critically commented on the inconsistent and fragmented operationalization of applicant reactions to selection methods (Truxillo & Bauer, 2011). There is a multitude of measurement instruments to assess different facets of applicant perceptions, which makes it challenging to compare study results (Ryan & Ployhart, 2000). Future research in the field of applicant reactions to GAs would profit from a systematic investigation of the effect of individual game features as well as the use of a consistent acceptance measurement methodology. While this paper has focused on test-takers' acceptance of the GSST, it remains unclear if the practitioners perceive cognitive flexibility as a relevant characteristic worth selecting for. In addition, acceptance of the GSST among human resource professionals should be a primary concern for further research and may be closely linked to face validity.

We also note that the use of game design elements may exert a confounding effect on the measurement of cognitive flexibility. While both the WCST and the GSST provide feedback to the test-taker, that is, whether a trial has been successful or not, the WCST additionally employs a performance graph. This may tap into test takers' individual differences in motivation which would violate the implicit underlying assumption of unidimensionality (Nunnally & Bernstein, 1994).

Finally, several methodological aspects may limit the generalizability of the present findings. One major constraint is that the sample consisted dominantly of university students, with a majority of psychology students. Psychology students are likely to be more familiarized with neuropsychological and cognitive ability tests, therefore test results and test perceptions may differ from the general population. To account for potential biases study participants were asked to indicate their prior WCST experience and the results revealed that this did not relate to neither performance nor acceptance of the WCST or the GSST. Although it seems justifiable to start

a line of research with a student sample for practical and economic reasons, subsequent studies should extend the findings with a sample that involves the intended target group of the GSST. Moreover, this study only involved a hypothetical selection process which most likely entails different motivational characteristics and outcomes (Landers & Behrend, 2015). It can be expected that actual applicants in a real selection situation are more involved or engaged, thus not only achieving better test results but also perceiving the selection process differently than participants in a selection simulation. In particular, whether someone receives a favorable outcome (hired or not), substantially impacts their perceptions of the selection procedure (Bauer et al., 1998). As this study investigated GA acceptance without real consequences or feedback about the selection decision, results might differ from acceptance ratings post-decision in high-stake situation (Ryan & Ployhart, 2000). To provide greater external validity, future research should endeavor to replicate the present findings in a field study involving an authentic selection situation.

8 | CONCLUSION

Within the scope of this study, we devised a gamified adaption of the WCST for the means of personnel selection, and provided preliminary support for its validity, as well for its acceptance among test takers. Cognitive flexibility, as measured by the Gamified Set-Shifting Task, may provide a valuable tool to assist organizations in selecting employees who can operate effectively in changing and dynamic environments while improving job candidates' perceptions of the selection procedure at the same time. In a more general way, the present findings indicate that transforming a traditional abstract psychological test to a gamified and work-related format, may be an effective method to ensure high-quality psychometric measurement while leveraging technological advances. This study adds to the emerging field of GAs in personnel selection by providing empirical evidence for their unique benefits over traditional selection methods and will hopefully encourage further research in this area.

ACKNOWLEDGEMENTS

The authors thank Sascha Klatt for his help with programming the gamified version of the Wisconsin Card Sorting Test. Björn Hommel's work on this paper was supported by a PhD scholarship from Leipzig University. Open Access funding enabled and organized by Projekt DEAL.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available at <https://osf.io/jcdvq/>

ORCID

Björn E. Hommel  <http://orcid.org/0000-0002-7375-006X>

Regina Ruppel  <http://orcid.org/0000-0002-8775-6868>

Hannes Zacher  <http://orcid.org/0000-0001-6336-2947>

REFERENCES

- Anderson, S. W., Damasio, H., Jones, R. D., & Tranel, D. (1991). Wisconsin Card Sorting Test performance as a measure of frontal lobe damage. *Journal of Clinical and Experimental Neuropsychology*, 13, 909–922. <https://doi.org/10.1080/01688639108405107>
- Armstrong, M. B., Landers, R. N., & Collmus, A. B. (2016). Gamifying recruitment, selection, training, and performance management: Game-thinking in human resource management. In H. Gangadharbatla, & D. Davis (Eds.), *Emerging research and trends in gamification* (pp. 140–165). IGI Global.
- Bauer, T. N., Maertz, C. P., Dolen, M. R., & Campion, M. A. (1998). Longitudinal assessment of applicant reactions to employment testing and test outcome feedback. *Journal of Applied Psychology*, 83, 892–903. <https://doi.org/10.1037/0021-9010.83.6.892>
- 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, 387–419. <https://doi.org/10.1111/j.1744-6570.2001.tb00097.x>
- Benit, N., & Soellner, R. (2012). Misst gut, ist gut? Vergleich eines abstrakten und eines berufsbezogenen Matrizentests. *Journal of Business and Media Psychology*, 3, 22–29.
- Bentler, P. M., Jackson, D. N., & Messick, S. (1971). Identification of content and style: A two-dimensional interpretation of acquiescence. *Psychological Bulletin*, 76, 186–204.
- Bowden, S. C., Fowler, K. S., Bell, R. C., Whelan, G., Clifford, C. C., Ritter, A. J., & Long, C. M. (1998). The reliability and internal validity of the Wisconsin Card Sorting Test. *Neuropsychological Rehabilitation*, 8, 243–254. <https://doi.org/10.1080/713755573>
- Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology*, 1, 185–216. <https://doi.org/10.1177/135910457000100301>
- Buil, I., Catalán, S., & Martínez, E. (2020). Understanding applicants' reactions to gamified recruitment. *Journal of Business Research*, 110, 41–50. <https://doi.org/10.1016/j.jbusres.2019.12.041>
- Burt, C. B. D., Crowe, L., & Thomas, K. (2018). Validation of a gamified measure of safety behavior: The SBT. In S. Haugen, A. Barros, C. van Gulijk, T. Kongsvik, & J. E. Vinnem (Eds.), *Safety and reliability—Safe societies in a changing world* (pp. 263–270). Taylor & Francis Group. <https://doi.org/10.1201/9781351174664-34>
- Cañas, J. J., Fajardo, I., & Salmeron, L. (2006). Cognitive flexibility. *International Encyclopedia of Ergonomics and Human Factors*, 1, 297–301. <https://doi.org/10.13140/2.1.4439.6326>
- Cañas, J. J., Quesada, J., Antolí, A., & Fajardo, I. (2003). Cognitive flexibility and adaptability to environmental changes in dynamic complex problem-solving tasks. *Ergonomics*, 46, 482–501. <https://doi.org/10.1080/0014013031000061640>
- Cardador, M. T., Northcraft, G. B., & Whicker, J. (2017). A theory of work gamification: Something old, something new, something borrowed, something cool? *Human Resource Management Review*, 27(2), 353–365. <https://doi.org/10.1016/j.hrmr.2016.09.014>
- Chow, S., & Chapman, D. (2013). *Gamifying the employee recruitment process*. *Proceedings of the First International Conference on Gameful Design, Research, and Applications*. pp. 91–94. <https://doi.org/10.1145/2583008.2583022>
- Collmus, A. B., Armstrong, M. B., & Landers, R. N. (2016). Game-thinking within social media to recruit and select job candidates. In R. N. Landers, & G. B. Schmidt (Eds.), *Social Media in Employee Selection and Recruitment* (pp. 103–124). Springer. https://doi.org/10.1007/978-3-319-29989-1_6
- Condon, D. M., & Revelle, W. (2014). The international cognitive ability resource: Development and initial validation of a public-domain measure. *Intelligence*, 43, 52–64. <https://doi.org/10.1016/j.intell.2014.01.004>
- Connolly, T. M., Boyle, E. A., MacArthur, E., Hainey, T., & Boyle, J. M. (2012). A systematic literature review of empirical evidence on computer games and serious games. *Computers & Education*, 59, 661–686. <https://doi.org/10.1016/j.compedu.2012.03.004>
- Csikszentmihályi, M. (1990). *Flow: The psychology of optimal experience*. Harper & Row.
- Dajani, D. R., & Uddin, L. Q. (2015). Demystifying cognitive flexibility: Implications for clinical and developmental neuroscience. *Trends in Neurosciences*, 38, 571–578. <https://doi.org/10.1016/j.tins.2015.07.003>
- Deák, G. O. (2004). The development of cognitive flexibility and language abilities. *Advances in Child Development and Behavior*, 31, 271–327. [https://doi.org/10.1016/S0065-2407\(03\)31007-9](https://doi.org/10.1016/S0065-2407(03)31007-9)
- Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). *From game design elements to gamefulness: Defining "gamification."* *Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments*, pp. 9–15.
- Deterding, S., Sicart, M., Nacke, L., O'Hara, K., & Dixon, D. (2011). Gamification. using game-design elements in non-gaming contexts. *Proceedings of the 2011 annual conference extended abstracts on human factors in computing systems*, 2425–2428. <https://doi.org/10.1145/1979742.1979575>
- Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, 64, 135–168. <https://doi.org/10.1146/annurev-psych-113011-143750>
- 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. https://doi.org/10.1207/S15327043HUP1601_4
- Eby, L. T., Butts, M., & Lockwood, A. (2003). Predictors of success in the era of the boundaryless career. *Journal of Organizational Behavior*, 24, 689–708. <https://doi.org/10.1002/job.214>
- Edwards, A. L. (1953). The relationship between the judged desirability of a trait and the probability that the trait will be endorsed. *Journal of Applied Psychology*, 37(2), 90–93. <https://doi.org/10.1037/h0058073>
- Fetzer, M., McNamara, J., & Geimer, J. L. (2017). Gamification, serious games and personnel selection. In H. W. Goldstein, E. D. Pulakos, J. Passmore, & C. Semedo (Eds.), *The Wiley Blackwell handbook of the psychology of recruitment, selection and employee retention* (1st ed., pp. 293–309). Wiley. <https://doi.org/10.1002/9781118972472.ch14>
- Figueroa, I. J., & Youmans, R. J. (2013). Individual differences in cognitive flexibility predict poetry originality. In D. Harris (Ed.), *Engineering psychology and cognitive ergonomics. understanding human cognition* (Vol. 8019, pp. 290–296). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-39360-0_32
- Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., DeFries, J. C., & Hewitt, J. K. (2006). Not all executive functions are related to intelligence. *Psychological Science*, 17, 172–179. <https://doi.org/10.1111/j.1467-9280.2006.01681.x>
- Georgiou, K., Gouras, A., & Nikolaou, I. (2019). Gamification in employee selection: The development of a gamified assessment. *International Journal of Selection and Assessment*, 27, 91–103. <https://doi.org/10.1111/ijsa.12240>
- Gilliland, S. W. (1993). The perceived fairness of selection systems: An organizational justice perspective. *Academy of Management Review*, 18(4), 694–734.
- Gkorezis, P., Georgiou, K., Nikolaou, I., & Kyriazati, A. (2020). Gamified or traditional situational judgement test? A moderated mediation model of recommendation intentions via organizational attractiveness. *European Journal of Work and Organizational Psychology*, 30, 1–11. <https://doi.org/10.1080/1359432X.2020.1746827>
- Gödöllei Lappalainen, A. F. (2017). *Game-based assessments of cognitive ability: Validity and effects on adverse impact through perceived stereotype threat, test-taking motivation and anxiety* [Unpublished master's thesis]. University of Calgary. <https://doi.org/10.11575/PRISM/26255>

- Griffin, B., & Hesketh, B. (2003). Adaptable behaviours for successful work and career adjustment. *Australian Journal of Psychology*, 55, 65–73. <https://doi.org/10.1080/00049530412331312914>
- Guin, T. D. -L., Baker, R., Mechling, J., & Ruyle, E. (2012). Myths and realities of respondent engagement in online surveys. *International Journal of Market Research*, 54, 613–633. <https://doi.org/10.2501/IJMR-54-5-613-633>
- Hausknecht, J. P., Day, D. V., & Thomas, S. C. (2004). Applicant reactions to selection procedures: An updated model and meta-analysis. *Personnel Psychology*, 57, 639–683. <https://doi.org/10.1111/j.1744-6570.2004.00003.x>
- Hawkes, B., Cek, I., & Handler, C. (2017). The gamification of employee selection tools: An exploration of viability, utility, and future directions. In J. C. Scott, D. Bartram, & D. H. Reynolds (Eds.), *Next generation technology-enhanced assessment* (1st ed., pp. 288–314). Cambridge University Press. <https://doi.org/10.1017/9781316407547.013>
- Heaton, R. K., Chelune, G. J., Talley, J. L., Kay, G. G., & Curtiss, G. (1993). *Wisconsin Card Sorting Test manual: Revised and expanded*. Psychological Assessment Resources.
- Highhouse, S., Lievens, F., & Sinar, E. F. (2003). Measuring attraction to organizations. *Educational and Psychological Measurement*, 63, 986–1001. <https://doi.org/10.1177/0013164403258403>
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30, 179–185. <https://doi.org/10.1007/BF02289447>
- Ionescu, T. (2012). Exploring the nature of cognitive flexibility. *New Ideas in Psychology*, 30, 190–200. <https://doi.org/10.1016/j.newideapsych.2011.11.001>
- Kercood, S., Lineweaver, T. T., Frank, C. C., & Fromm, E. D. (2017). Cognitive flexibility and its relationship to academic achievement and career choice of college students with and without Attention Deficit Hyperactivity Disorder. *Journal of Postsecondary Education and Disability*, 30, 329–344.
- Kersting, M. (2008). Zur Akzeptanz von Intelligenz- und Leistungstests. *Report Psychologie*, 33, 420–433.
- Kibby, M. Y., Schmitter-Edgecombe, M., & Long, C. J. (1998). Ecological validity of neuropsychological tests: Focus on the California Verbal Learning Test and the Wisconsin Card Sorting Test. *Archives of Clinical Neuropsychology*, 13, 523–534.
- Kim, Y. J., & Shute, V. J. (2015). The interplay of game elements with psychometric qualities, learning, and enjoyment in game-based assessment. *Computers & Education*, 87, 340–356. <https://doi.org/10.1016/j.compedu.2015.07.009>
- Koivisto, J., & Hamari, J. (2019). The rise of motivational information systems: A review of gamification research. *International Journal of Information Management*, 45, 191–210. <https://doi.org/10.1016/j.ijinfomgt.2018.10.013>
- Kongs, S. K., Thompson, L. L., Iverson, G. L., & Heaton, R. K. (2000). *Wisconsin Card Sorting Test-64 Card Version (WCST-64)*. Lutz, FL: PAR.
- König, C. J., Klehe, U. -C., Berchtold, M., & Kleinmann, M. (2010). Reasons for being selective when choosing personnel selection procedures. *International Journal of Selection and Assessment*, 18, 17–27. <https://doi.org/10.1111/j.1468-2389.2010.00485.x>
- Krumm, S., Hüffmeier, J., Dietz, F., Findeisen, A., & Dries, C. (2011). Towards positive test takers' reactions to cognitive ability assessments: Development and initial validation of the Reasoning Ability at Work Test. *Journal of Business and Media Psychology*, 2, 11–18.
- Landers, R. N. (2015). An introduction to game-based assessment: Frameworks for the measurement of knowledge, skills, abilities and other human characteristics using behaviors observed within videogames. *International Journal of Gaming and Computer-Mediation Simulations*, 7(4), iv–viii.
- Landers, R. N., Auer, E. M., & Abraham, J. D. (2020). Gamifying a situational judgment test with immersion and control game elements: Effects on applicant reactions and construct validity. *Journal of Managerial Psychology*, 35, 225–239. <https://doi.org/10.1108/JMP-10-2018-0446>
- Landers, R. N., & Behrend, T. S. (2015). An inconvenient truth: Arbitrary distinctions between organizational, Mechanical Turk, and other convenience samples. *Industrial and Organizational Psychology*, 8, 142–164. <https://doi.org/10.1017/iop.2015.13>
- Laureiro-Martínez, D., Brusoni, S., & Zollo, M. (2009). Cognitive flexibility in decision-making: A neurological model of learning and change. *CROMA-Center for Research in Organization and Management-Bocconi University*, 1, 1–43.
- Laureiro-Martínez, D., & Brusoni, S. (2018). Cognitive flexibility and adaptive decision-making: Evidence from a laboratory study of expert decision makers. *Strategic Management Journal*, 39, 1031–1058. <https://doi.org/10.1002/smj.2774>
- Martin, A. J., Nejad, H., Colmar, S., & Liem, G. A. D. (2012). Adaptability: Conceptual and empirical perspectives on responses to change, novelty and uncertainty. *Australian Journal of Guidance and Counselling*, 22, 58–81. <https://doi.org/10.1017/jgc.2012.8>
- Martin, A. J., Nejad, H. G., Colmar, S., & Liem, G. A. D. (2013). Adaptability: How students' responses to uncertainty and novelty predict their academic and non-academic outcomes. *Journal of Educational Psychology*, 105, 728–746. <https://doi.org/10.1037/a0032794>
- Mavletova, A. (2015). A gamification effect in longitudinal web surveys among children and adolescents. *International Journal of Market Research*, 57, 413–438. <https://doi.org/10.2501/IJMR-2015-035>
- McCarthy, J. M., Bauer, T. N., Truxillo, D. M., Anderson, N. R., Costa, A. C., & Ahmed, S. M. (2017). Applicant perspectives during selection: A review addressing “So what?,” “What's new?,” and “Where to next?” *Journal of Management*, 43, 1693–1725. <https://doi.org/10.1177/0149206316681846>
- Morris, L., & Mansell, W. (2018). A systematic review of the relationship between rigidity/flexibility and transdiagnostic cognitive and behavioral processes that maintain psychopathology. *Journal of Experimental Psychopathology*, 9, 204380871877943. <https://doi.org/10.1177/2043808718779431>
- Moscoco, S., Salgado, J. F., & Anderson, N. (2017). How do I get a job, what are they looking for? Personnel selection and assessment. In N. Chmiel (Ed.), *An introduction to work and organizational psychology* (pp. 25–47). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781119168058.ch2>
- Nicholson, S. (2015). A recipe for meaningful gamification. In T. Reiners, & L. C. Wood (Eds.), *Gamification in education and business* (pp. 1–20). Springer International Publishing. https://doi.org/10.1007/978-3-319-10208-5_1
- Nikolaou, I., Georgiou, K., & Kotsaralidou, V. (2019). Exploring the relationship of a gamified assessment with performance. *The Spanish Journal of Psychology*, 22, 1–10. <https://doi.org/10.1017/sjp.2019.5>
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed). McGraw-Hill.
- Papastergiou, M. (2009). Digital game-based learning in high school computer science education: Impact on educational effectiveness and student motivation. *Computers & Education*, 52, 1–12. <https://doi.org/10.1016/j.compedu.2008.06.004>
- Ployhart, R., & Bliese, P. (2006). Individual Adaptability (I-ADAPT) Theory: Conceptualizing the antecedents, consequences, and measurement of individual differences in adaptability. *Advances in human performance and cognitive engineering research* (Vol. 6, pp. 3–39). Elsevier. [https://doi.org/10.1016/S1479-3601\(05\)06001-7](https://doi.org/10.1016/S1479-3601(05)06001-7)
- Ployhart, R. E. (2006). Staffing in the 21st Century: New challenges and strategic opportunities. *Journal of Management*, 32, 868–897. <https://doi.org/10.1177/0149206306293625>
- Pulakos, E. D., Arad, S., Donovan, M. A., & Plamondon, K. E. (2000). Adaptability in the workplace: Development of a taxonomy of

- adaptive performance. *Journal of Applied Psychology*, 85, 612–624. <https://doi.org/10.1037/0021-9010.85.4.612>
- Pulakos, E. D., Dorsey, D., & White, S. (2006). Adaptability in the workplace: Selecting an adaptive workforce. In C. Shawn Burke, L. G. Pierce, & E. Salas (Eds.), *Advances in human performance and cognitive engineering research* (Vol. 6, pp. 41–71). Elsevier. [https://doi.org/10.1016/S1479-3601\(05\)06002-9](https://doi.org/10.1016/S1479-3601(05)06002-9)
- Ready, R. E., Stierman, L., & Paulsen, J. S. (2001). Ecological validity of neuropsychological and personality measures of executive functions. *The Clinical Neuropsychologist*, 15, 314–323. <https://doi.org/10.1076/clin.15.3.314.10269>
- Reder, L. M., & Schunn, C. D. (1999). Bringing together the psychometric and strategy worlds: Predicting adaptivity in a dynamic task. In D. Gopher, & A. Koriat (Eds.), *Attention and performance XVII: Cognitive regulation of performance: Interaction of theory and application* (pp. 315–342). MIT Press.
- Reiter-Palmon, R. (2003). Predicting leadership activities: The role of flexibility. *Individual Differences Research*, 1, 124–136.
- Roth, P. L., BeVier, C. A., Switzer, III, F. S., & Schippmann, J. S. (1996). Meta-analyzing the relationship between grades and job performance. *Journal of Applied Psychology*, 81, 548–556.
- Ryan, A. M., & Ployhart, R. E. (2000). Applicants' perceptions of selection procedures and decisions: A critical review and agenda for the future. *Journal of Management*, 26, 565–606. <https://doi.org/10.1177/014920630002600308>
- Ryan, A. M., Sacco, J. M., McFarland, L. A., & Kriska, S. D. (2000). Applicant self-selection: Correlates of withdrawal from a multiple hurdle process. *Journal of Applied Psychology*, 85, 163–179. <https://doi.org/10.1037/0021-9010.85.2.163>
- Rynes, S. L. (1993). When recruitment fails to attract: Individual expectations meet organizational realities in recruitment. In H. Schuler, J. L. Farr, & M. Smith (Eds.), *Individual and organizational perspectives* (pp. 27–40). Psychology Press.
- Sailer, M., Hense, J., Mandl, H., & Klevers, M. (2013). Psychological perspectives on motivation through gamification. *Interaction Design and Architecture Journal*, 19, 28–37.
- 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. <https://doi.org/10.1016/j.chb.2016.12.033>
- Sanchez, D. R., & Langer, M. (2020). Video game pursuit (VGPU) scale development: Designing and validating a scale with implications for game-based learning and assessment. *Simulation & Gaming*, 51, 55–86. <https://doi.org/10.1177/1046878119882710>
- Savickas, M. L., & Porfeli, E. J. (2012). Career Adapt-Abilities Scale: Construction, reliability, and measurement equivalence across 13 countries. *Journal of Vocational Behavior*, 80, 661–673. <https://doi.org/10.1016/j.jvb.2012.01.011>
- Schmidt, F. L., & Hunter, J. (2004). General mental ability in the world of work: Occupational attainment and job performance. *Journal of Personality and Social Psychology*, 86, 162–173. <https://doi.org/10.1037/0022-3514.86.1.162>
- Schunn, D. C., & Reder, L. M. (2001). Another source of individual differences: Strategy adaptivity to changing rates of success. *Journal of Experimental Psychology: General*, 130, 59–76. <https://doi.org/10.1037/0096-3445.130.1.59>
- Seaborn, K., & Fels, D. I. (2015). Gamification in theory and action: A survey. *International Journal of Human-Computer Studies*, 74, 14–31. <https://doi.org/10.1016/j.ijhcs.2014.09.006>
- Smither, J. W., Reilly, R. R., Millsap, R. E., Pearlman, K. P., & Stoffey, R. W. (1993). Applicant reactions to selection procedures. *Personnel Psychology*, 46, 49–76. <https://doi.org/10.1111/j.1744-6570.1993.tb00867.x>
- Stemme, A., Deco, G., & Busch, A. (2007). The neuronal dynamics underlying cognitive flexibility in set shifting tasks. *Journal of Computational Neuroscience*, 23, 313–331. <https://doi.org/10.1007/s10827-007-0034-x>
- Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology*, 44, 24–31. Retrieved from <https://psytoolkit.org/>
- Truxillo, D. M., & Bauer, T. N. (2011). Applicant reactions to organizations and selection systems. In S. Zedeck (Ed.), *APA handbook of industrial and organizational psychology: Selecting and developing members for the organization* (Vol. 2, pp. 379–397). American Psychological Association. <https://doi.org/10.1037/12170-012>
- Woods, S. A., Ahmed, S., Nikolaou, I., Costa, A. C., & Anderson, N. R. (2020). Personnel selection in the digital age: A review of validity and applicant reactions, and future research challenges. *European Journal of Work and Organizational Psychology*, 29, 64–77. <https://doi.org/10.1080/1359432X.2019.1681401>
- World Economic Forum. (2016). *The future of jobs: Employment, skills and workforce strategy for the fourth industrial revolution*. Global challenge insight report. World Economic Forum. http://www3.weforum.org/docs/WEF_Future_of_Jobs.pdf

How to cite this article: Hommel, B. E., Ruppel, R., & Zacher, H. (2022). Assessment of cognitive flexibility in personnel selection: Validity and acceptance of a gamified version of the Wisconsin Card Sorting Test. *International Journal of Selection and Assessment*, 30, 126–144. <https://doi.org/10.1111/ijsa.12362>

APPENDIX A

TABLE A1 Acceptance items

1	Der Inhalt des Tests hatte eindeutig etwas mit der beschriebenen Arbeitsstelle zu tun. (FV) <i>(The actual content of the examination was clearly related to the job.)</i>
2	Ich bin überzeugt, dass mit diesem Verfahren vorhergesagt werden kann, wie gut eine Person später ihre Arbeit ausüben wird. (PPV) <i>(I am confident that the examination can predict how well an applicant will perform on the job.)</i>
3	Ich konnte keinen Zusammenhang zwischen dem Auswahlverfahren und den Anforderungen der Jobposition erkennen. (FV) <i>(I could not see any relationship between the examination and what is required on the job.)</i>
4	Meine Leistung in dem Test liefert einen guten Hinweis dafür, dass ich für den Job geeignet bin. (PPV) <i>(My performance on the examination was a good indicator of my ability to do the job.)</i>
5	Es wäre für jeden offensichtlich, dass die Auswahlmethode mit der Arbeitsstelle zusammenhängt. (FV) <i>(It would be obvious to anyone that the examination is related to the job.)</i>
6	Bewerber, die bei dieser Art von Test gut abschneiden, werden mit größerer Wahrscheinlichkeit gute Leistungen bei der Arbeit erbringen als Bewerber, die dabei schlecht abschneiden. (PPV) <i>(Applicants who perform well on this type of examination are more likely to perform well on the job than applicants who perform poorly.)</i>
7	Ich habe nicht verstanden, was das Auswahlverfahren mit dem Job zu tun hatte. (FV) <i>(I did not understand what the examination had to do with the job.)</i>
8	Ein schlechtes Testergebnis weist eindeutig darauf hin, dass man nicht qualifiziert für den Beruf ist. (PPV) <i>(Failing to pass the examination clearly indicates that you can't do the job.)</i>
9	Es gab keine wirkliche Verbindung zwischen dem soeben bearbeiteten Auswahlverfahren und der Arbeitsstelle. (FV) <i>(There was no real connection between the examination that I went through and the job.)</i>
10	Ein Arbeitgeber kann aus den Ergebnissen dieses Testverfahrens wichtige Informationen über die Fähigkeiten der Bewerber herausfinden. (PPV) <i>(The employer can tell a lot about the applicant's ability to do the job from the results of the examination.)</i>
11	Die Bearbeitung des Tests hat mir Spaß gemacht. (AR) <i>(I enjoyed the examination to a great degree.)</i>
12	Ich würde mich freuen, in zukünftigen Bewerbungen ein ähnliches Testverfahren zu durchlaufen. (AR) <i>(I would look forward to going through the same type of examination again in the future.)</i>
13	Insgesamt bin ich der Meinung, dass diese Auswahlmethode fair ist. (PJ) <i>(Overall, I believe that the examination was fair.)</i>
14	Ich hatte ein gutes Gefühl bei der Art und Weise, wie das Test angeleitet und ausgeführt wurde. (PJ) <i>(I felt good about the way the examination was conducted and administered.)</i>
15	Aufgrund meiner Erfahrungen mit diesem Auswahlverfahren würde ich anderen eine Bewerbung auf diese Stelle weiterempfehlen. (R) <i>(Based on my experience with the examination I would encourage others to apply for this job)</i>

Note: Adapted from Smither et al. (1993). Answer format: 5-point scale from 1 = does not apply at all to 5 = applies completely.

Abbreviations: AR, affective reaction; FV, face validity; PJ, procedural justice; PPV, perceived predictive validity; R, recommendation.

APPENDIX B

TABLE B1 The adaptability scale

1	In einer unbekanntem Situation fällt es mir leicht, eine Reihe von möglichen Handlungsoptionen zu durchdenken. <i>(I am able to think through a number of possible options to assist me in a new situation.)</i>
2	Ich kann meine Denkweise verändern, um mit neuen Situationen zurechtzukommen. <i>(I am able to revise the way I think about a new situation to help me through it.)</i>
3	Um eine neue Lebenslage erfolgreich zu meistern, passe ich mein Denken oder meine Erwartungen gegebenenfalls an. <i>(I am able to adjust my thinking or expectations to assist me in a new situation if necessary.)</i>
4	Es fällt mir leicht, nützliche Informationen, Ressourcen oder Kontakte zu finden, die mir in neuen Situationen weiterhelfen. <i>(I am able to seek out new information, helpful people, or useful resources to effectively deal with new situations.)</i>
5	In einer unbekanntem Situation gelingt es mir gut, neue Strategien zu entwickeln (z.B. eine andere Art Fragen zu stellen oder Informationen zu finden). <i>(In uncertain situations, I am able to develop new ways of going about things [e.g. a different way of asking questions or finding information] to help me through.)</i>
6	Um mich in einer unvertrauten Situation zurechtzufinden, kann ich meine Vorgehensweise bei Bedarf verändern. <i>(To assist me in a new situation, I am able to change the way I do things if necessary.)</i>
7	Ich bin in der Lage, negative Emotionen (z.B. Angst) abzubauen, um auch unter ungewissen Bedingungen zurechtzukommen. <i>(I am able to reduce negative emotions [e.g. fear] to help me deal with uncertain situations.)</i>
8	Um mich in einer unvertrauten Situation zurechtzufinden, kann ich meine Vorgehensweise bei Bedarf verändern. <i>(When uncertainty arises, I am able to minimize frustration or irritation so I can deal with it best.)</i>
9	Es fällt mir leicht, auf positive Gefühle und Emotionen zurückgreifen (z.B. Freude, Zuversicht), um mit unbekanntem Situationen bestmöglich umzugehen. <i>(To help me through new situations, I am able to draw on positive feelings and emotions [e.g. enjoyment, satisfaction].)</i>

Note: Martin et al. (2012). Answer format: 7-point scale from 1 = does not apply at all to 5 = applies completely.