



Good gamers, good managers? A proof-of-concept study with Sid Meier's *Civilization*

Alexander Simons¹ · Isabell Wohlgenannt¹ · Markus Weinmann² · Stefan Fleischer³

Received: 14 August 2018 / Accepted: 9 January 2020
© The Author(s) 2020

Abstract

Human resource professionals increasingly enhance their assessment tools with game elements—a process typically referred to as “gamification”—to make them more interesting and engaging for candidates, and they design and use “serious games” that can support skill assessment and development. However, commercial, off-the-shelf video games are not or are only rarely used to screen or test candidates, even though there is increasing evidence that they are indicative of various skills that are professionally valuable. Using the strategy game *Civilization*, this proof-of-concept study explores if strategy video games are indicative of managerial skills and, if so, of what managerial skills. Under controlled laboratory conditions, we asked forty business students to play the *Civilization* game and to participate in a series of assessment exercises. We find that students who had high scores in the game had better skills related to problem-solving and organizing and planning than the students who had low scores. In addition, a preliminary analysis of in-game data, including players' interactions and chat messages, suggests that strategy games such as *Civilization* may be used for more precise and holistic “stealth assessments,” including personality assessments.

Keywords Assessment · Gamification · Recruitment · Human resources · Serious games · Video games

JEL Classification J24 · M51

✉ Alexander Simons
alexander.simons@uni.li

Extended author information available on the last page of the article

1 Introduction

“I’ve been playing Civilization since middle school. It’s my favorite strategy game and one of the reasons I got into engineering.”

Mark Zuckerberg on *Facebook*, 21 October 2016

Information technology (IT) has changed human resource (HR) management, particularly its assessment procedures. HR professionals are increasingly using IT-enhanced versions of traditional selection methods such as digital interviews, social-media analytics, and reviews of user profiles on professional social-networking sites instead of traditional selection interviews, personality tests, and reference checks (Chamorro-Premuzic et al. 2016). While business games have a long history in personnel assessment and development, the use of digital games and game elements is also increasing (see, e.g., Ferrell et al. 2016). For example, computerized personality surveys and assessment exercises have been “gamified” with elements such as narratives, progress bars, and animations (Armstrong et al. 2016) to create a more engaging experience for applicants, and “serious” games—that is, digital games that serve purposes other than entertainment (Michael and Chen 2006)—have been designed for assessment, education, and training (see, e.g., Bellotti et al. 2013).

The potential of commercial, off-the-shelf video games has long been ignored by HR research, but interest in them has recently surfaced. Several video games have been found to be able to be indicative of various skills that are professionally valuable, including persistence, problem-solving, and leadership (Lisk et al. 2012; Shute et al. 2009, 2015), which are often referred to as twenty-first-century skills (see, e.g., Chu et al. 2017). Therefore, Petter et al. (2018) recently proposed that employers could use video games to screen or test applicants and that applicants should indicate their gaming experiences and achievements on their résumés. In fact, being adept at video games can significantly boost one’s career. For example, Jann Mardenborough, a professional racing driver, is said to have started his career by participating in *Gran Turismo* competitions (Richards 2012), and Matt Neil’s performance in the video game *Football Manager* allegedly paved the way for his career as a football analyst (Stanger 2016).

The use of video games for assessment purposes is often referred to as a “stealth assessment” (e.g., Shute et al. 2009; Wang et al. 2015). During stealth assessments, candidates are less aware that they are being evaluated (Fetzer 2015) because they can fully immerse themselves in the game, so that test anxiety and response bias can be reduced (Kato and de Klerk 2017; Shute et al. 2016). However, different video games and game genres can indicate very different types of skills (Petter et al. 2018), so the challenge faced by research is to determine which games can be used to assess which types of skills. Against this backdrop, we explore if and to what extent strategy video games can be used to assess managerial skills using the video game *Civilization* (www.civilization.com). We focus on managerial skills because they are closely related to several of the twenty-first-century skills that previous research has assessed using video games, and

we use *Civilization* because it is an unusually broad and open video game that confronts players with a high level of complexity: Dealing with multifaceted and deeply connected game mechanisms requires players to plan their actions carefully, to develop sophisticated strategies, and, in the multiplayer mode, to interact and trade with other players. In fact, there is increasing anecdotal evidence that *Civilization* requires skills such as critical thinking and strategic planning—skills that are known to be important in managerial jobs. To determine which managerial skills influence game success, this exploratory study focuses on the following research question: *Can strategy video games such as Civilization be used to assess managerial skills and, if so, what skills are they indicative of?* To answer our research question, we asked business students to participate in a controlled, correlational laboratory study that involved a series of multiplayer games and assessment exercises. Accordingly, the students' managerial skills were measured using the assessment-center method, and to answer our research question, we compared the participants' game scores with their assessment results.

The article proceeds as follows. Section 2 provides the background on personnel assessments and reviews research on game-based assessment methods. Section 3 describes the basic game principles of *Civilization* and provides a rationale for why the game could be used to assess managerial skills. Section 4 outlines the procedures for data collection and analysis and explains how we organized the multiplayer games with the participants and how we designed the assessments. Section 5 presents our findings, which are discussed in Sect. 6. Section 7 acknowledges the limitations and Sect. 8 concludes the paper.

2 Game-based assessment

The history of personnel-selection research stretches back to the first decade of the twentieth century (Ghiselli 1973). Since then, researchers have studied various methods for assessing candidates, including general mental-ability tests, reference checks, work-sample tests, interviews, job-knowledge tests, peer ratings, grades, and assessment centers (e.g., Reilly and Chao 1982; Schmidt and Hunter 1998; Schmitt et al. 1984). Since the late 1950s, increasing numbers of organizations from the private and public sectors have used assessment centers to evaluate applicants (Spychalski et al. 1997) and to develop and promote personnel (Ballantyne and Povah 2004). Assessment centers' greatest advantage over other predictors is that they combine traditional assessments such as interviews, simulation exercises, and personality tests to provide an overall evaluation of an applicant's knowledge and abilities (see Thornton and Gibbons 2009). Therefore, assessment centers allow employers to collect detailed information about candidates' skills and abilities such as their communication skills, problem-solving skills, or their ability to influence or be aware of others (Arthur et al. 2003).

During the past few years, IT has disrupted traditional forms of personnel selection by producing new, technology-enhanced assessment methods (Chamorro-Premuzic et al. 2016). For example, reference checks are increasingly conducted online using business-oriented websites such as *LinkedIn*, which inform potential

employers about applicants' professional networks, work experience, and recommendations (Zide et al. 2014). While job interviews via videoconferencing services such as *Skype* (Straus et al. 2001), unlike face-to-face meetings, may even be used for voice mining (Chamorro-Premuzic et al. 2016), social-media platforms such as *Facebook* and *Twitter* provide information about the applicants' personal relationships, private hobbies, and interests—information that has long been unavailable to recruiters (Stoughton et al. 2015). Therefore, though they can save time and costs (Mead and Drasgow 1993), such assessments also raise several legal and ethical issues (Slovensky and Ross 2012) as well as privacy concerns (Stoughton et al. 2015), and they may even influence construct measurement (Morelli et al. 2017). For example, researchers have found that the results of computerized versions of cognitive-ability tests, personality tests, and situational-judgment tests can differ from those of written tests (see, e.g., Stone et al. 2015) because candidates may tend to answer more quickly but less accurately in IT-based assessments (Van de Vijver and Harsveld 1994). Against this background, researchers have been challenged to compare traditional assessments with the IT-based methods that are increasingly used in HR practice (Anderson 2003).

In addition to these technology-enhanced assessment methods, a recent trend in assessment is the “gameful” design of personnel-selection methods (see, e.g., Chamorro-Premuzic et al. 2016). Gamification, which refers to the use of game elements, and serious games, which refers to the design and use of purposeful games, have received special attention from researchers. Gamification generally describes the idea of using game elements in non-game contexts (Deterding et al. 2011) to increase user engagement (Huotari and Hamari 2012). Examples of such game elements are leader boards, progress bars, feedback mechanisms, badges, and awards (Hamari et al. 2014), which have been used in contexts as diverse as marketing, health, and education (e.g., Huotari and Hamari 2012; Kapp 2012; McCallum 2012). Among others, researchers have studied the gamification of personality surveys and assessment exercises using game elements such as narratives and progress bars (Armstrong et al. 2016; Ferrell et al. 2016). Today, the rapidly growing gamification market offers various applications that can support personnel selection. For example, *Nitro*, a cloud-based enterprise gamification platform by Bunchball, can be used to implement game elements such as challenges, badges, and leaderboards on websites, apps, and social networks to assess employee performance (www.bunchball.com); *HR Avatar*, a company that administers online employment tests, uses animations to create immersive simulations for various types of jobs (www.hravatar.com); and, *Visual DNA*, a web-based profiling technology, queries website visitors using images instead of text to learn about their personalities (www.visualdna.com).

Serious games are those that have been developed for purposes other than entertainment (Michael and Chen 2006). Serious games are especially common in education, where they have long been used to engage learners and help them to acquire new knowledge and abilities through play (see, e.g., Van Eck 2006). However, serious games are also becoming increasingly popular in other domains, including in the marketing, social-change, and health fields (e.g., McCallum 2012; Peng et al. 2010; Susi et al. 2007). HR has long used business games (i.e., serious games that have been developed for business training), which were once paper-and-pencil-based

but are mostly digital today (Bellotti et al. 2013). Examples for serious games that can be used for personnel assessment are *America's Army*, an online shooting game developed by the US Army to recruit soldiers (www.americasarmy.com); *Theme Park Hero*, a game-based cognitive ability assessment for recruitment by Revelian (www.revelian.com); and *Knack*, a mobile puzzle game that assesses players' skills in various dimensions and awards players with skill-related badges based on their game results (www.knackapp.com). In fact, such gaming apps may lead to a shift in the relationship between assessor and assessed, from business-to-business to business-to-consumer, and from reactive test-taking to proactive test-taking—such that “the testing market will increasingly transition from the current *push* model—where firms require people to complete a set of assessments in order to quantify their talent—to a *pull* model where firms will search various talent badges to identify the people they seek to hire” (Chamorro-Premuzic et al. 2016, p. 632; emphasis in original).

While gamification and serious games have received some attention from researchers, the market for recruitment games and gamified assessment applications has grown much more quickly than academic interest has, which “leaves academics playing catch up and human resources (HR) practitioners with many unanswered questions,” especially regarding these approaches' validity (Chamorro-Premuzic et al. 2016, p. 622). Commercial, off-the-shelf video games have received even less scientific attention, although researchers have recently shown increasing interest in video games. In fact, during the past few years, several video games have been found to be indicative of various skills other than gaming skills, including professional and digital skills, so Petter et al. (2018) encouraged applicants to share their gaming experiences on their résumés and during job interviews, and employers to use video games to screen or test candidates. As Barber et al. (2017, p. 3) put it, “similar to how an individual's background in competitive sports communicates information to a hiring manager, an individual's history in online gaming can be a signal to a hiring manager of attributes possessed by the potential job candidate.”

Various video games may qualify for skill assessment, including tactical games such as *Use Your Brainz* (a modified version of *Plants vs. Zombies 2*) and role-playing games such as *The Elder Scrolls: Oblivion*, which have been used to assess problem-solving skills (Shute et al. 2009, 2016); massively multiplayer online games such as *EVE Online* and *Chevaliers' Romance III*, which may indicate leadership skills and behavior (Lisk et al. 2012; Lu et al. 2014); and first-person shooters such as *Counter Strike*, which may be used to learn about players' creativity (Wright et al. 2002). In addition, video games may reflect intellectual abilities, for example, multiplayer online battle arenas such as *League of Legends* and *DOTA 2*, adventure games such as *Professor Layton and the Curious Village*, and puzzle games such as Nintendo's *Big Brain Academy* (Kokkinakis et al. 2017; Quiroga et al. 2009, 2016). Video games may even be used to train and develop these and related skills, for example, sandbox games such as *Minecraft*, which have been used to teach planning, language, and project-management skills (see Nebel et al. 2016); multiplayer games such as *Halo 4* and *Rock Band*, which have been found to improve team cohesion and performance (Keith et al. 2018); and puzzle games such as *Portal 2*, which have been found to improve players' spatial, problem-solving, and persistence skills

(Shute et al. 2015). A broader experimental study with various video games such as *Borderlands 2*, *Minecraft*, *Portal 2*, *Warcraft III*, and *Team Fortress 2* suggested that video games may generally be used to train individuals in communication skills, adaptability, and resourcefulness (Barr 2017).

Accordingly, in studying the relationship between gaming and skill assessment and development, researchers have mostly focused on twenty-first-century or digital skills. However, as Granic et al. (2014) explained, different game genres offer different benefits to gamers, thus it is still a challenge for research to determine what game genres can be used to assess and train which types of skills. In particular, strategy video games deserve the researchers' attention because they are both complex and social (Granic et al. 2014). Due to strategy games' complexity, players must carefully plan and balance their decisions, develop alternative game strategies, and deal with high levels of uncertainty; furthermore, since modern strategy games are typically played online with other players, they are also interactive and social, so that communication and negotiation skills are important. Therefore, strategy games could arguably be useful for skill assessment; however, they have not yet received much attention from researchers. Basak et al. (2008) used *Rise of Nations*, a real-time strategy video game, to train executive functions in older adults; Glass et al. (2013) found that *StarCraft*, another real-time strategy video game, can improve cognitive flexibility; and Adachi and Willoughby (2013) discovered a relationship between gaming and self-reported problem-solving skills for strategy games as opposed to fast-paced games. Still, most of the research has been dedicated to game genres other than that of strategy and has tended to neglect several skills that may be assessed using strategy games. Against this background, this study explores if strategy games such as *Civilization* are indicative of managerial skills, so they could be used for assessment purposes.

3 Sid Meier's Civilization

Civilization is a long-standing series of strategy games in which players move in turns, giving them time to think, which is why the game has been compared to chess (Squire and Steinkuehler 2005). Sidney K. "Sid" Meier and Bruce Shelley created the first *Civilization* game for *MicroProse* in 1991. Since then, five sequels and several expansion packs and add-ons have been released. With millions of copies sold and multiple awards won—the opening theme of *Civilization IV* was even awarded a Grammy—, *Civilization* is considered one of the best and most widely played turn-based video games to date (see Owens 2011). The current version of the series is *Civilization VI*, which was not available at the time when we collected our data, so we used *Civilization V*. However, most of the information we provide applies to the whole game series.

The idea of the *Civilization* game is to build a civilization from scratch from the ancient era to the modern age, which requires players to expand and protect their borders, build new cities, develop their infrastructures, discover novel technologies, maintain economies, promote their cultures, and pursue diplomacy. Including all downloadable content and the two expansion packs *Gods & Kings* and *Brave New*

World, forty-three civilizations are currently available in *Civilization V*, and each offers unique gameplay advantages. The world differs in each game, with differing geography, terrain, and resources. During the game, players must explore their world to uncover the randomly generated map, find new resources, identify suitable locations for founding cities, and outline the other civilizations' territories. The game can be played alone in single-player mode (i.e. against the computer) or together with other players in multiplayer mode (i.e. against each other). There are four main types of victory in the game—domination, science, culture, and diplomacy—, so it offers numerous avenues through which to pursue success:

- First, if all but one player has lost their original capital cities through conquest, the last player who still possesses his or her own capital city wins the domination victory. To achieve the domination victory, players can recruit more than 120 military units, ranging from archers and warriors to nuclear missiles and giant death robots. While all these units have their general advantages and disadvantages, their strength and speed further depend on a number of factors such as the opponents and the terrain. In addition, several buildings can be constructed to increase the strength of the military units (e.g., barracks, armories, and military academies) or to improve the defense of cities (e.g., walls, castles, and arsenals).
- Second, the first player whose technological development is advanced enough to build and launch a spaceship wins the science victory, for which technological progress is most important. Science progresses with every turn, and once players have researched enough, they can discover novel technology that yields new units, new buildings, or certain game advantages. More than eighty technologies (e.g., mining, biology, and nuclear fusion) in several eras (e.g., the ancient, medieval, and atomic eras) can be researched. Choosing a technology to explore is not easy because scientific discovery follows predefined and complex paths in the so-called tech tree. Various buildings can accelerate scientific progress (e.g., libraries, universities, and public schools).
- Third, the player whose cultural influence dominates all other civilizations wins the cultural victory. Players develop their civilizations' culture with every turn, which expands their borders and allows them to introduce social policies that yield certain gameplay bonuses. *Civilization* offers forty-five social policies (e.g., humanism, philanthropy, and reformation) and three ideologies (freedom, order, and autocracy) with sixteen tenets each. In addition, great works of artists, writers, and musicians as well as ancient artifacts that can be found in archeological digs together produce tourism, which helps civilizations spread their culture around the world. Several buildings (e.g., monuments, opera houses, and museums) support a cultural victory.
- Fourth, the player who wins a world-leader resolution in the World Congress achieves the diplomatic victory. All civilization leaders are represented by a certain number of delegates in the World Congress (which later in the game becomes the United Nations), where they can propose, enact, reject, or repeal resolutions that—for good or for bad—affect all of them (e.g., embargos, funding, and taxes). The number of delegates a civilization has is especially important for proposals to pass in the World Congress, and this mainly depends on the

number of that civilization's city-state allies. Players can seek allies from among sixty-four city-states (e.g., Zurich, Prague, and Hanoi) of differing types (e.g., religious, mercantile, and maritime), and diplomats help them find out how other civilizations think about their proposed resolutions and make diplomatic agreements.

If no player has achieved one of the four types of victory, the game ends in the year 2050, and the player with the highest score wins the time victory. It is not entirely clear how the game calculates the scores, but there are many websites, wikis, and forums that offer quite sensible estimates, suggesting that scores are calculated as a function of several factors with different weightings that reflect economic, scientific, cultural, and military progress. Among them are the number (and size) of cities owned, technologies researched, wonders built, and the amount of land controlled. As players can pursue different types of victory, there is no simple or ideal strategy for winning the game. Instead, they must develop balanced strategies, as weakness in any area can weaken other areas:

[T]he strategies in winning, whichever conditions the player might choose, are intricate and manifold. If a player attempts a military victory, he/she still needs to keep up scientific research, or the units will become obsolete. A strong economy must be maintained or the player won't be able to support all of the military units. A variety of cities are necessary to build units, but cities not only require maintenance, they also need to be defended from enemies. Regardless of what path the player chooses, an appropriate balance must be struck. Within this framework, there are many options for the player to explore (Camargo 2006, n.p.).

In sum, *Civilization* has a great variety of ways in which to play and win, making it an unusually broad and open game. While even the central game elements—terrain features, resource types, buildings, religion, happiness, espionage, trading, archeology, wonders, promotions, specialists, great people, barbarians, and many more—cannot be explained concisely, our broad overview should provide some sense of the game mechanics. (A more detailed description of the game can be found at <http://civilization.wikia.com>.) As explained, strategy games are both complex and social, which is especially the case with *Civilization*, so the game may indicate several skills other than gaming skills that are important when on the job: *Civilization* requires players to deal with multifaceted and deeply connected game mechanisms such as economics, science, culture, and religion—along with various units, buildings, and resources—, which demand careful planning and strategy development. In the multiplayer mode, players must also interact with each other, either cooperatively through diplomacy, trading, and research, or competitively through war, espionage, and embargos, so they must communicate and negotiate. Against this background, strategy video games such as *Civilization* may be indicative of analytical skills such as organizing, planning, and decision-making, and interpersonal skills such as communication and negotiation—skills that largely correspond to those that have been deemed important for managerial positions (see, e.g., Arthur et al. 2003).

According to Common Sense Media, a nonprofit provider of entertainment and technology recommendations to families and schools, *Civilization* provides an educational tool for classrooms and helps to develop players' creativity and thinking-and-reasoning ability (Sapieha n.d.). In fact, the game has also been used as an educational tool in, for example, history lessons (Squire 2004; also see Shreve 2005), so it is not surprising that it was planned to develop an educational version of the game for use in North American high schools (Carpenter 2016). Early on, Squire and Barab (2004, pp. 505 and 512) found that *Civilization* can not only help students learn about history, but also about the interplay between geography, politics, and economics, and that "powerful systemic-level understandings" can emerge through gameplay. Against this background, our study explores if strategy games such as *Civilization* can be used to assess managerial skills and what skills they can assess—"to ascertain exactly what it is that players are taking away from games such as [...] *Civilization*" (Shute et al. 2009, p. 298).

4 Method

4.1 Participants

We promoted the research project in lectures and via e-mail and offered participants a copy of *Civilization V* plus add-ons and the chance to win one of six prizes in a lottery—three tablet computers, a notebook, an e-book reader, and a *Civilization* board game—as an incentive to participate. Fifty business students, all native German speakers from a small European university, volunteered to participate. Shortly after a student had responded, we explained to him or her the conditions of participation via e-mail and provided copies of *Civilization V*, including the add-ons, *Brave New World* and *Gods & Kings*. The participants had one month to learn how to play *Civilization*, which was a challenge for some, as becoming competent in the game requires players to invest considerable time and effort. Therefore, ten students who applied for the study withdrew, citing time constraints. Table 1 provides descriptive statistics for the forty remaining participants.

Participants' average age was 24.10 years, and thirty of the forty participants were male. Twenty-three of the participants were undergraduate business-administration students, while the remaining seventeen were in business-oriented master's programs at the graduate level. Thirty-three percent had participated in an assessment center before. Their previous *Civilization V* playtimes—which we could measure because all participants became our "friends" on *Steam*, a software distribution platform—ranged from 3.80 to 260.30 h, with a standard deviation of 39.25 h. Still, as only a few of the volunteers had played the game before, their *Civilization* playtimes, with a few exceptions, were relatively equally distributed among them, with a mean of 33.40 h and a median of 26.95 h. The participants' self-estimated experience with other *Civilization* titles (e.g., *Civilization I–IV*, *Beyond Earth*) ranged from 0 to 200 h, with a mean of 23.90 h. They reported spending an average of around 4 h/week on video games of any kind (often action, sports, and strategy games).

Table 1 Participants' descriptive statistics

Variable	Unit	Obs	Mean	SD	Min	Max
Age	(years)	40	24.10	4.70	19.00	46.00
Gender	(female = 1)	40	.25	.44	.00	1.00
Study level	(Master's = 1)	40	.43	.50	.00	1.00
Gaming habits	(h/week)	40	4.08	5.54	.00	25.00
<i>Civilization V</i> playtime	(h)	40	33.40	39.25	3.80	260.30
Experience with other <i>Civilization</i> titles	(h)	40	23.90	54.09	.00	200.00
Experience with assessment centers	(yes = 1)	40	.33	.47	.00	1.00

4.2 Procedures

Multiplayer games We organized ten four-hour multiplayer games, each with four participants. The games were run as permanently supervised LAN games in a computer lab, where we had installed *Steam* and *Civilization V*. To ensure that participants could not identify each other during the game and team up with their friends, they were randomly assigned to groups, and they used anonymized *Steam* accounts and usernames. In addition, their workstations were surrounded by whiteboards so they could not see each other's screens, they were not permitted to speak aloud to each other, and they wore headphones so they could not hear each other when they were typing in the game's chat window. To ensure that the participants would try to play as skillfully as possible, the winner of one of the most expensive lottery prizes was drawn from among the ten participants who had earned the highest scores in the multiplayer games. Figure 1 illustrates the physical layout of the multiplayer games.

We informed the participants about the game setup via e-mail before the gaming sessions started. All of them played the "Washington" civilization to ensure that they had equal benefits. To rule out potential artificial intelligence (AI) biases, there were no computer players. The "Pangaea" map type was used so all players shared a single, huge landmass (as opposed to maps with several islands or continents). The difficulty level was set at medium-high ("emperor") to make the game challenging, the game pace was set to "quick" to shorten the time required for a game, the

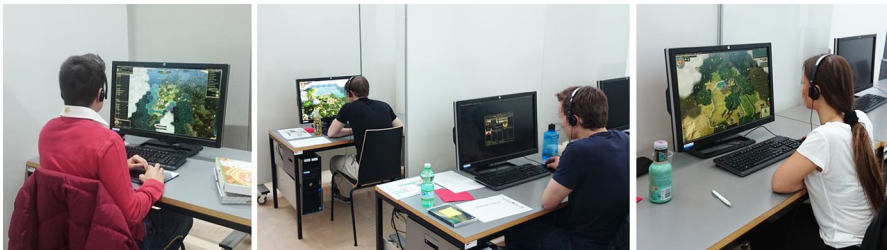


Fig. 1 Physical layout of the multiplayer games. This figure is not included in the article's Creative Commons licence

resource distribution was “balanced” so the geography was as fair as possible, and the turn timer was enabled to prevent players from delaying the game. In addition, the map size was “tiny,” the four main types of victory were enabled, movement and combat were set to “quick,” and downloadable content other than the approved add-ons, *Gods & Kings* and *Brave New World*, was disabled. All the other settings (e.g., game era, world age, number of city-states) were standard. With increasing playtime, *Civilization* tends to slow down, especially in the multiplayer mode, so we tested this setup in three one-day LAN games, each with at least four unique players, to ensure it would perform adequately.

Assessment centers We designed our assessments according to established guidelines and procedures from the academic and professional literature on personnel selection (e.g., Ballantyne and Povah 2004; Caldwell et al. 2003). For example, our design incorporated the ten recommendations established by the *International Task Force on Assessment Center Guidelines*, which address issues ranging from behavioral classification and simulation to recording and data integration (Joiner 2000) (Appendix 1). We assigned participants to groups based on the groups in which they played the games and we conducted ten assessments with four participants each. Each of the ten assessments took approximately 5 h.

To provide an incentive for the participants to perform as well as possible in the assessments, we drew one of the lottery prizes from among the ten participants who performed best. In addition, we offered all participants the chance to receive feedback on how they performed during the assessments. After a short introduction that provided an overview of the time schedule and exercises, participants signed a declaration of consent that stated that they had participated voluntarily, that they could quit at any time for any reason, and that they would keep the contents of the assessments confidential until the study was completed so their fellow participants could not prepare in advance. The assessments concluded with a short personality test and a debriefing in which the participants were presented with their preliminary results and could ask questions about the study.

Our assessments featured the probably most common types of assessment-center exercises: *presentations*, *in-basket exercises*, *case studies*, *role plays*, and *group discussions* (see Spychalski et al. 1997) (Appendix 2). All exercises, which were conducted in German to ensure sufficient comprehension, came from the academic and professional literature on personnel evaluation and selection. We supervised the participants’ work on all exercises, including the breaks, and videotaped all exercises except for the written case study and in-basket exercises to facilitate detailed data analysis. We selected only those exercises that did not require more than basic managerial knowledge and we adapted them slightly to match our objectives. Figure 2 illustrates the setting of the assessments based on screenshots we took from the videos.

Our exercises required participants to show the dimensions of managerial skill that are most commonly evaluated in assessment centers (Arthur et al. 2003): *consideration/awareness of others* (“awareness of others” hereafter), which reflects the extent to which individuals care about others’ feelings and needs; *communication*, which reflects how individuals deliver information in oral or written form; *drive*, which reflects individuals’ activity level and how persistently they pursue



Fig. 2 Assessment-center setup (role play, group discussion, presentation). This figure is not included in the article's Creative Commons licence

achievement; *influencing others*, which reflects how successfully an individual can steer others either to adopt a certain point of view or to do or not do something; *organizing and planning*, which reflects individuals' ability to organize their work and resources systematically to accomplish tasks; and *problem-solving*, which reflects how individuals gather, understand, and analyze information to generate realizable options, ideas, and solutions (Arthur et al. 2003).

The six skill dimensions categorize several skills that we could directly observe and measure with our exercises, so they represent the categories we used to classify behaviors displayed by participants (Joiner 2000), and we developed a hierarchical competency system (Chen and Naquin 2006) that defined which dimensions were assessed in which exercise. Each dimension was assessed in more than one exercise, and—even though the videos we took allowed for repeated and focused evaluations—we only assessed between two and five dimensions per exercise (see Woehr and Arthur 2003). We used twenty-five, more measurable and specific skills that we borrowed from the academic literature on personnel recruitment to evaluate the participants' performance in the six dimensions (Appendix 3).

4.3 Measures

Game success We measured participants' game success based on their final *Civilization* scores because it is nearly impossible to achieve any type of victory in *Civilization V* other than the domination victory in a 4-h game. As explained, these scores are automatically calculated by the game and are a function of several factors, each with its own weighting, that reflect economic, scientific, cultural, and military progress. Although all games were of equal length, participants' game scores varied with the number of turns a group took, and the number of turns varied with the game pace (e.g., war slowed the game down in some groups). To allow for group comparisons, we calculated a participant's *Mean points per turn* as the quotient of his or her total points in the game and the number of turns that his or her group took.

Managerial skills Two assessors, one of whom was not part of the project team, used a 7-point Likert scale (where 7 is high) to independently evaluate the participants' performance during the assessments. One of the main reasons that assessment

centers often fail is insufficient assessor training (Caldwell et al. 2003), so our assessors used detailed instruction and evaluation material that we created based on the literature and on notes that one of the researchers took while observing the participants' work. As is typically recommended, the assessors used sample solutions, criteria catalogs, and behavior checklists that described desirable and undesirable behavior (see, e.g., Reilly et al. 1990). The assessors independently reviewed the participants' written solutions for the case study and in-basket exercise and watched the videos of the other exercises at least twice. They took detailed notes to justify their ratings.

Accordingly, our assessors independently rated the skills that the participants demonstrated during their work on the exercises, and we averaged their individual ratings to get final skill ratings for each exercise. As the rating scale was ordinal, we measured the assessors' level of agreement using Kendall's coefficient of concordance. All coefficients of concordance were significant, so inter-rater reliability was generally high (Appendix 4). Next, we averaged the assessors' skill ratings across exercises to get composite skill-dimension ratings. For example, for measuring the skill dimension *Organizing and planning*, we used data collected from the case-study, in-basket, and presentation exercises and averaged the following skill ratings: *Coaching*, *Delegation*, *Strategic thinking*, *Planning and scheduling*, *Structuring and organizing*, and *Time sensitivity*; for measuring the skill dimension *Problem-solving*, we used data collected from the case-study and in-basket exercises and from the group discussion and averaged the following skill ratings: *Solution finding*, *Decisiveness*, *Problem analysis*, and *Fact finding*. (Appendix 3 provides additional information as to what skill ratings were used to measure what skill dimension.)

4.4 Model specification

We specified a linear mixed-effects regression model to estimate the relationship between participants' performance in the game (measured as mean points per turn) and their managerial skills (measured as skill-dimension ratings). Because the participants played *Civilization* in groups we used a mixed-effects model with varying intercepts to consider group effects, as observations within the same group might be correlated (Gelman and Hill 2007); as Barr et al. (2013) suggested, we specified a linear mixed-effects model with maximum random effects.

We also had to assume that the effects were not constant across groups, as group-specific game dynamics (e.g., war and alliances between players) may have had an influence, so the model also allowed for the coefficients (i.e., the slopes) to vary across groups. According to Snijders and Bosker (2012), random-coefficient models are especially useful for relatively small groups like the four-participant groups in our study. Therefore, we specified the following varying-intercept, varying-slopes model (see StataCorp 2019, p. 14):

$$SDR_{ijk} = (\beta_{00} + u_{0j}) + (\beta_{10} + u_{1j}) \cdot MPT_{ij} + \gamma' \cdot Controls_{ij} + \varepsilon_{ij},$$

where SDR_{ijk} is the skill-dimension rating k for a participant i in a group j ; β_{00} represents the overall mean intercept; β_{10} is the overall mean effect (slope) of *Mean points per turn* (MPT_{ij}); $Controls_{ij}$ are the control variables *Age*, *Gender*, *Civilization V playtime*, *Experience with other Civilization titles*, *Gaming habits*, *Study level*, and *Experience with assessment centers*; and ε_{ij} indicates level-one residuals (i.e., on the individual level), which are assumed to be normally distributed with mean 0. As observations from the four participants in a group might be correlated, u_{0j} is a level-two random effect (i.e., a group-specific random intercept) that describes the between-group variability of the outcome variable SDR_{ijk} and captures the non-independence between observations of SDR_{ijk} for participants i in a group j , so it allows the intercept β_{00} to vary across groups. Similarly, u_{1j} is a level-two random effect (i.e., a group-specific random slope) of MPT_{ij} that accounts for in-game group dynamics and allows the coefficient β_{10} to vary across groups. Both random effects, u_{0j} and u_{1j} , are assumed to be normally distributed with mean 0.¹

5 Results

5.1 Descriptive results

Table 2 shows the participants' game results and assessment results. The participants' *Total points* in the game (i.e., their final scores) ranged from 213 to 1291, with a mean of close to 700 points and a standard deviation of around 246 points. The number of *Turns* the groups took ranged from 131 to 205, with a mean of around 165 turns. The participants' *Mean points per turn* averaged 4.20, had a standard deviation of 1.30, and ranged from 1.28 to 6.62.

The participants' performance in each of the six skill dimensions ranged from 2.00 (*Drive*, *Influencing others*) to 6.67 (*Influencing others*). The mean and standard deviation for *Awareness of others* were 4.10 and .94, respectively, while they were 4.41 and .75 for *Communication*, 4.04 and .89 for *Drive*, 4.29 and 1.21 for *Influencing others*, 4.00 and .79 for *Organizing and planning*, and 4.04 and .81 for *Problem-solving*.

Next, we test whether the participants' game results correlated with their assessment results.

5.2 Regression results

Based on our model specification, we conducted a series of regression analyses to test whether the participants' game results correlated with their assessment results. That is, we ran separate regressions on the six skill dimensions using the same model

¹ We modelled the binary control variables *Gender* and *Experiences with assessment centers* as fixed factors because they contain all population levels in our study (Snijders and Bosker 2012). We modelled all other control variables in the same way as *Mean points per turn* (i.e., coefficients with fixed and random components; maximal random-effects structure; see Barr et al. 2013).

Table 2 Descriptive results

Variable	Unit	Obs	Mean	SD	Min	Max
<i>Video games</i>						
Turns	(abs. number)	40	165.20	23.00	131.00	205.00
Total points	(abs. number)	40	698.80	246.28	213.00	1291.00
Mean points per turn	(see text)	40	4.20	1.30	1.28	6.62
<i>Assessments</i>						
Awareness of others	(see text)	40	4.10	.94	2.17	6.08
Communication	(see text)	40	4.41	.75	3.00	6.00
Drive	(see text)	40	4.04	.89	2.00	6.33
Influencing others	(see text)	40	4.29	1.21	2.00	6.67
Organizing and planning	(see text)	40	4.00	.79	2.71	5.57
Problem-solving	(see text)	40	4.04	.81	2.33	5.50

specification, while participants' skill-dimension ratings provided the outcome variables (i.e., *Awareness of others*, *Communication*, *Drive*, *Influencing others*, *Organizing and planning*, and *Problem-solving*). While we found no significant relationships between *Mean points per turn* and *Awareness of others*, *Communication*, *Drive*, and *Influencing others*, we found *Mean points per turn* to significantly correlate with both *Organizing and planning* and *Problem-solving*. For each of these two skill dimensions, we estimated two models, one model without control variables and one model with control variables. Table 3 presents the regression results for *Organizing and planning* and Table 4 presents the regression results for *Problem-solving*. We used *Stata 13.1* to estimate the mixed-effects models ("mixed command"). By default, *Stata* uses the maximum-likelihood estimation (StataCorp 2019).²

For *Organizing and planning*, Model 1a (without controls) indicates a significantly positive coefficient for *Mean points per turn* ($\beta = .25, p < .00$), which remains robust when adding the control variables (Model 1b: $\beta = .18, p < .05$). Accordingly, both models suggest that game success is correlated with higher skill levels in *Organizing and planning*.

For *Problem-solving*, Model 2a (without controls) indicates a significantly positive coefficient for *Mean points per turn* ($\beta = .19, p < .04$), which remains robust when adding the control variables (Model 2b: $\beta = .19, p < .04$). Accordingly, both models suggest that game success is correlated with higher skill levels regarding *Problem-solving*.

In summary, the mixed-effects linear regression analysis suggests that participants who had high *Civilization* scores had significantly better problem-solving skills and organizing-and-planning skills on average than did participants who performed less

² The assumptions in linear mixed-effects models are weaker than they are in normal linear regression models (Gelman and Hill 2007, pp. 45–47). We tested for multicollinearity, which presented no problems, as all variance inflation factor (VIF) values were smaller than 2. We also tested for normality of errors using *qqplots*, which also presented no problems.

Table 3 Regression results for *Organizing and planning*

	Dependent variable: organizing and planning			
	Model 1a: without controls		Model 1b: with controls	
Mean points per turn	.25 (.08)	.00**	.18 (.09)	.05*
<i>Civilization V</i> playtime			.00 (.00)	.93
Experience with other <i>Civilization</i> titles			-.00 (.00)	.50
Gaming habits			.03 (.02)	.26
Age			-.02 (.03)	.57
Gender			-.52 (.32)	.11
Study level			.20 (.24)	.41
Experience with assessment centers			-.16 (.26)	.53
Intercept	2.95 (.38)	.00***	3.62 (.92)	.00***
Group-specific effects	Yes		Yes	
Log likelihood	-42.53		-39.73	

* $p < .05$; ** $p < .01$; *** $p < .001$. Standard errors are in parentheses. $N = 40$; number of groups = 10

Table 4 Regression results for *Problem-solving*

	Dependent variable: problem-solving			
	Model 2a: without controls		Model 2b: with controls	
Mean points per turn	.19 (.09)	.04*	.19 (.09)	.04*
<i>Civilization V</i> playtime			.00 (.00)	.79
Experience with other <i>Civilization</i> titles			-.00 (.00)	.37
Gaming habits			-.02 (.03)	.49
Age			-.04 (.03)	.16
Gender			-.56 (.30)	.06
Study level			.09 (.25)	.72
Experience with assessment centers			.06 (.25)	.80
Intercept	3.25 (.41)	.00***	4.31 (.91)	.00***
Group-specific effects	Yes		Yes	
Log likelihood	-45.87		-40.80	

* $p < .05$; ** $p < .01$; *** $p < .001$. Standard errors are in parentheses. $N = 40$; number of groups = 10

well in the game. This result suggests that game success is positively related to these two skill dimensions.

6 Discussion

Gamification, the use of game elements in non-game contexts (Deterding et al. 2011), has received considerable attention from researchers (see, e.g., Hamari et al. 2014), as has the design and use of serious games that have been developed

for purposes other than entertainment (Michael and Chen 2006). Researchers have long studied the negative effects of conventional video games and have only recently turned to their potentially positive effects (e.g., Liu et al. 2013). Vichitvanichphong et al. (2016, p. 10) examined video games' potential for indicating elderly persons' driving skills and concluded that "good old gamers are good drivers." Similarly, using the example of the strategy game *Civilization*, we explored video games' potential for indicating managerial skills and asked whether good gamers would be good managers. *Civilization* has already received attention from researchers in various disciplines (e.g., Hinrichs and Forbus 2007; Owens 2011; Squire and Barab 2004; Squire and Steinkuehler 2005; Testa 2014), but application scenarios in business contexts have not yet been explored. Against this backdrop, we explored the following research question: *Can strategy video games such as Civilization be used to assess managerial skills and, if so, what skills are they indicative of?*

Our results should be useful to researchers from various fields who are becoming increasingly aware of video games' potential to indicate several skills other than gaming skills. Our study revealed significant and positive relationships between the participants' game success and how they performed during our assessments. As explained, assessment centers can provide a comprehensive picture of an applicant's knowledge and abilities, thus they are increasingly used to predict future job performance. Therefore, we also used the data collected from the assessments to calculate an overall assessment rating, a commonly used job-performance predictor (e.g., Russell and Domm 1995). In creating an overall assessment rating, there are different approaches to data aggregation (Thornton and Rupp 2006, p. 161), and we tested two purely quantitative approaches: First, we aggregated the skill-dimension ratings into overall assessment ratings, with weightings based on the relevance of the skill dimensions to the exercises; second, we used the skill ratings to calculate exercise ratings, which we then aggregated into overall assessment ratings, with weightings based on the length of the exercises. For both aggregation approaches, we explored how the overall assessment results correlated with participants' game results, using the same model specification as before, and found that the students' overall assessment ratings were significantly related to their game scores. Accordingly, video games may not only be used to assess specific skills but could also be useful to predict performance at a more general level. In fact, assessment centers are one of the most commonly used tools to predict the future job performance of university graduates (see, e.g., Ballantyne and Povah 2004) who apply for managerial positions but typically lack work experience.

As there are several predictors other than assessment centers that can be used for evaluating and selecting personnel, including general mental-ability tests, reference checks, work-sample tests, peer ratings, and grades (e.g., Reilly and Chao 1982; Schmidt and Hunter 1998; Schmitt et al. 1984), we also compared the students' game results with their academic performance. While the results of this comparison have been presented elsewhere as research-in-progress (Simons et al. 2015), they confirmed that participants who had high scores in the game performed significantly better in their studies than did the participants who had low game scores. Clearly, even though grades are a common tool in hiring, some researchers have questioned their predictive power regarding job performance and adult achievement

(e.g., Bretz 1989; Cohen 1984). Still, several studies have suggested that grades and future job performance are related (e.g., Dye and Reck 1989; Roth et al. 1996), so our pre-test provided additional evidence for the usefulness of video games in personnel selection.

Accordingly, our results support the notion that gaming experiences and achievements may meaningfully inform personnel recruitment and assessment (Petter et al. 2018). As Efron (2016, n.p.) put it: “The more children play games to learn and navigate life, the more they will expect them as they enter the adult world. Employers who get ahead of this curve will have an advantage in the war for talent. The best of the best will be snared through games.” While games are unlikely to replace traditional assessment methods, they may provide a useful, innovative, and engaging supplement to other recruitment tests. In addition, if an off-the-shelf game such as *Civilization* can be an indicator for managerial skills, even if only to some extent, certainly strategy games developed specifically for that purpose offer potential for personnel recruitment. Having said that, this is a proof-of-concept study, so we do not recommend the use of *Civilization* for assessments in professional contexts, as using a standard video game such as *Civilization* for assessment purposes carries the risk that applicants who have played the game before will receive higher ratings than applicants who have not. (The participants’ previous *Civilization* playtimes were relatively equally distributed and only a few of them had played the game before, so gaming experience was not an issue in our study; instead, our measure of game success rather reflects how fast participants learned the game in the study-preparation phase.) In fact, it is a well-known challenge of game-based assessment that gamers may have an unfair advantage over non-gamers (Kim and Shute 2015). Accordingly, our results also suggest that “serious” strategy games that are designed for skill assessment offer companies an opportunity to save time and money, as recruitment procedures such as the use of assessment centers are time-consuming and expensive.

The design and use of video games for recruitment purposes requires understanding what skills and skill dimensions the games assess and what game mechanisms allow for skill assessment. Therefore, our study was exploratory and identified the dimensions of managerial skill that correlate with success in the *Civilization* game. We found significant positive correlations between the participants’ game results and their problem-solving skills and organizing-and-planning skills but no statistical evidence for other skills such as communication or the ability to influence others. However, this result does not necessarily mean that no strategy game can indicate the presence of other skill dimensions, because our study only focused on a specific game (i.e., *Civilization*) and used a highly aggregated measure of game success (i.e., the participants’ *Civilization* scores). In fact, video games offer much more data than what we analyzed in this study. For test purposes, we developed a *Civilization* mod (“modding” refers to changing a video game using development tools) (see Owens 2011) and ran it during the multiplayer games to collect various performance measures per player and per turn, including the players’ in-game chats, which provided a near-complete picture of each participant’s performance in the game (e.g., what was researched and in what order). A systematic exploration of the log files is outside the scope of this article, but a preliminary analysis suggests that in-game data analytics offers the potential to draw a more sophisticated picture of managerial talent. For

example, we extracted data on a participant's number of allies and opponents from the log files, both of which may reflect interpersonal skills. In fact, the number of opponents (allies) was negatively (positively) correlated with the participants' ability to influence others, while the average number of chat messages was positively correlated with the participants' communication skills. As modern video games produce tremendous amounts of data, they may thus inform employers about more than just the broad skills we measured.

Accordingly, our future research will explore the extent to which strategy games such as *Civilization* can be used for “stealth assessments,” which refers to “the real-time capture and analysis of gameplay performance data” such as game logs (Ke and Shute 2015, p. 301), and is “woven directly and invisibly into the [...] gaming environment” (Shute 2015, p. 62). As video games are immersive, stealth assessments can reduce test anxiety and the urge to respond in certain ways (Kato and de Klerk 2017), especially when it comes to non-cognitive skills such as conscientiousness that are usually assessed through self-reported means (Moore and Shute 2017). Theoretically grounded in evidence-centered design (see Mislevy et al. 2016), stealth assessments require the development of a competency model, which defines claims about candidates' competencies, an evidence model, which defines the evidence of a claim and how to measure that evidence, and a task model, which determines the tasks or situations that trigger such evidence (Van Eck et al. 2017; also see, e.g., Shute and Moore 2017). Accordingly, our future research will focus on developing such models and on exploring what skills and skill dimensions can be assessed with in-game data. For example, strategy games may also offer potential to measure social and interpersonal skills and personality traits, as people may behave differently in a gaming environment than they would in a job-application procedure—in fact, faking is a known limitation of personality tests (Morgeson et al. 2007). The qualitative analysis of players' in-game behavior during assessments, for example based on chats and performance data, may shed light on individuals' negotiation strategies, including opportunistic behavior, emotional intelligence, and persistence.

Finally, our study is correlational, so the causality is unclear—that is, our results do not suggest that *Civilization* can be used to develop managerial skills nor train individuals in these skills. Still, deliberately designed strategy games may not only measure performance but may also improve certain skills such as those at the analytical level. Therefore, our results might also stimulate research on the design of game-based personnel-development tools that companies might use for employee development and that job applicants might use to test and train their abilities before they participate in assessments.

7 Limitations

Our research has some limitations. First, as participation in our study was voluntary and time-consuming—participants spent an average of more than 25 h learning how to play the game, they all participated in a 4-h multiplayer game, and the assessment-center exercises took 5 h—our sample size was small, so the robustness of the observed effects could be questionable. Therefore, we also estimated the models

(without controls) using Bayesian data analysis, which can handle small sample sizes better than frequentist methods can (Hinneburg et al. 2007). According to the Bayesian estimation,³ the effect of game success on organization and planning was .26*** and that for problem-solving was .20**. Therefore, all effects are comparable to the effects estimated using the frequentist approach and different from zero, so they further support our results.

In addition, even though the participants were assigned randomly to groups, the groups' composition may still have affected individual performance. To account for the groups' differing playing times, we measured game success as mean points per turn, but other factors at the group level, especially the dynamics inherent in the game, may have biased the results. For example, if an unskilled player leaves a city (in the game) undefended, the player who conquers that city has a significant advantage for the rest of the game, which would affect the group's overall performance. We constructed linear mixed-effects models that were not only useful for our small group sizes but also allowed for the coefficients and the intercepts of the regression functions to vary across groups. Still, while we included several control variables, future research should use more holistic models. For example, general mental ability is a heavily used predictor of managerial performance (Schmidt and Hunter 2004), but we did not measure our participants' general mental ability, even though playing video games such as *Civilization* is cognitively demanding (see Granic et al. 2014).

The validity of our measures, especially at the skill-dimension level, presents another limitation. To assess their validity, we used confirmatory factor analysis where the latent variables were the exercises and the skill dimensions, and the observed variables were the skills (see, e.g., Gorsuch 1983). While most skills had significant factor loadings with their corresponding exercises, indicating high validity, many skills did not load on their corresponding skill dimension or were even insignificant. However, this does not necessarily indicate a measurement error, as assessment centers have repeatedly been found to lack construct validity across exercises (see, e.g., Bycio et al. 1987; Jansen and Stoop 2001; Sackett and Dreher 1982). For example, Archambeau (1979) found that skill-dimension ratings measured in the same exercise correlated strongly and positively, while the same skill-dimension ratings measured across exercises correlated far more modestly, and Neidig et al. (1979) presented similar results (both cited in Gibbons and Rupp 2009). These findings have led to a long and ongoing debate among HR researchers on the so-called construct-related validity paradox (see, e.g., Arthur et al. 2000). We used a structured literature review to identify a consistent and valid set of skills, but these skills were still diverse. For example, the skill dimension of *Communication* was measured with skills such as writing, spelling, and grammar (i.e., written communication), as well as clarity of speech and verbal ability (i.e., oral communication). However, a good speaker is not necessarily a good writer, which may explain the results of our

³ In a Bayesian analysis, the significance level of parameter estimates is based on highest-density intervals (HDIs). An HDI indicates which points of the posterior distribution are most credible (Kruschke 2014). Therefore, we consider values inside the HDI to be more credible than those that are outside the HDI and use the following significance levels: ***99%, **95%, *90% when the HDI does not contain 0.

validity tests. In addition, for some of the skill dimensions, we could only measure very few skills (Appendix 3), so it is still a challenge for future research to collect additional evidence on the relationship between gaming and managerial skills.

While our results are consistent with related work on inconsistency in assessment-center ratings, the low construct validity may also result from poor assessment-center design and implementation (Woehr and Arthur 2003). However, even though the design of assessment centers is generally not straightforward (see, e.g., Bender 1973), we believe that our assessments were demonstrably thorough. Caldwell et al. (2003) identified ten common assessment-center errors ranging from inadequate job analysis to sloppy behavior documentation. To avoid these errors, our assessment-center design followed established guidelines from the academic and professional literature on personnel recruitment (e.g., Ballantyne and Povah 2004). In particular, ten principles established by the *International Task Force on Assessment Center Guidelines* provided a framework for our assessments (Joiner 2000) (Appendix 1). Against this background, we are confident that our research takes an important step toward clarifying the potential of strategy games such as *Civilization* in assessment.

8 Conclusions

Our study suggests that video games such as *Civilization* can be used to assess problem-solving skills and organizing-and-planning skills—skills that are highly relevant for managerial professions. We thus conclude that collecting and analyzing data from strategy video games can offer useful insights for profilers and recruiters in the search for talent. A preliminary analysis of in-game data collected during the multiplayer games further suggests that strategy games offer the opportunity to assess other dimensions of managerial skill, including interpersonal skills. Our future research will thus explore if and to what extent strategy games such as *Civilization* can be used for stealth assessments, which collect and analyze gameplay performance data in real time to draw conclusions about individuals' management capabilities.

Acknowledgements We thank all of the students who participated for their time and effort. We also owe thanks to many of our colleagues at the University of Liechtenstein, especially Matthias Tietz for a brilliant performance in the role-play exercise and for his support in evaluating the assessment-center results; Roope Jaakonmäki for a terrific poster design and other help; Nicole Thöny and Sandra Beyer for organizing rooms, catering, and the like; and Bernd Schenk and Jan vom Brocke for general project support. This article is an extension and revision of a research-in-progress paper presented at the *36th International Conference on Information Systems (ICIS 2015)* in Fort Worth, Texas, and we thank the anonymous *ICIS* reviewers for their constructive comments on our research (Simons et al. 2015). Finally, we thank Martin Hibbeln from the University of Duisburg-Essen, Stephan Kramer from the Rotterdam School of Management, Oliver Müller from Paderborn University, Jan Recker from the University of Cologne, and Christoph Schneider from the IESE Business School for sharing their thoughts and ideas with us. Any remaining errors are our own.

Funding The study was funded by the Liechtenstein National Research Fund (Grant No. wi-1-16) and also received financial support from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement (Grant No. 645751).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix 1: Assessment-center design

We designed our assessment centers following established guidelines and procedures from the professional and academic literature. The following ten principles, established by the *International Task Force on Assessment Center Guidelines*, provided a framework for our assessments (Joiner 2000):

1. *Job analysis* held minor importance in our case, as we did not intend to evaluate participants' suitability for a specific job but focused on the assessment of general managerial skills. Therefore, we used the dimensions of managerial skill that are most commonly used in assessment centers (Arthur et al. 2003). In addition, either the exercises we selected did not require specific subject knowledge or we adapted them to match our objectives.
2. *Behavioral classification* refers to determining which behavior is representative of which type of managerial skill. For that purpose, Chen and Naquin (2006) recommended developing a hierarchical competency system, which can be used to categorize the skill dimensions, skills, and specific behaviors displayed by the participants during the exercises (Joiner 2000). Our evaluation schemes were designed accordingly, and the sample solutions, checklists, and criteria we used described desirable and undesirable behavior at a detailed level.
3. The *assessment techniques*, that is, the exercises used for the assessment, must allow the researcher to evaluate the defined skill dimensions (Joiner 2000). Our assessments featured the most common types of exercises (see Spsychalski et al. 1997), which we borrowed from established academic and professional textbooks (Appendix 2). We also established a link between skill dimensions and exercises by creating an exercise/competency matrix (Joiner 2000) (Tables 5 and 6).
4. *Multiple assessments* refer to the use of several exercises to elicit a variety of behaviors (Joiner 2000). We conducted a pre-test of our assessment techniques (Caldwell et al. 2003) to ensure that they allowed us to collect objective and reliable behavioral information in the defined skill dimensions (Joiner 2000). We also used a broad spectrum of short exercises instead of only a few, similar

exercises that require a long time to complete (Thornton and Byham 1982), and each skill dimension was assessed in several exercises (see Thornton and Gibbons 2009).

5. *Simulations* like group exercises, in-basket exercises, interaction simulations, presentations, and fact-finding exercises are an important element of assessment centers, as they can be used to observe individuals' behavioral responses to job-related situational stimuli (Joiner 2000). Accordingly, all our exercises provided simulations instead of simple evaluations of subject knowledge or multiple-choice tests.
6. *Multiple assessors* must observe the applicants' behavior and evaluate the performance based on the defined skill dimensions (see Thornton and Gibbons 2009). Among other factors, the number of assessors required depends on the types of exercises, the skill dimensions assessed, and the assessors' experience and training, but a typical ratio of subjects to assessors is two subjects to one assessor (Joiner 2000). Accordingly, we used two assessors for each group of four participants.
7. *Assessor training* includes behavioral-observation training and performance-dimension training, the former of which helps to sensitize assessors and supports note-taking, and the latter of which reduces the risk that performance is assessed based on overall impressions instead of actual skills (Ballantyne and Povah 2004; Jackson et al. 2005). Our assessors received detailed instructions and, even though they did not have psychology backgrounds, they were experienced in evaluating students' performance in terms of grading.
8. As to *recording behavior*, assessors should follow a systematic procedure and record their impressions accurately at the time of observation based on, for example, notes or checklists (Joiner 2000). Our assessors evaluated the participants' performance in a systematic, replicable, and reliable manner. In addition, they did not conduct their assessments during the participants' work on the exercises but did so afterward based on videos, which allowed for repeated and more focused evaluations.
9. Assessors should also create *reports* of their observations before the aggregation discussion or statistical aggregation (Joiner 2000). Our integration approach did not involve discussions between assessors but followed a purely quantitative model, and inter-rater reliability was high (Appendix 4). Still, our assessors took detailed notes to justify their assessments for each exercise.
10. There are various approaches to *data integration*. Thornton and Rupp (2006) distinguished five methods of integrating assessment-center observations and ratings, from the purely judgmental to the purely statistical. As we wanted to increase replicability and objectivity, we applied a purely statistical aggregation approach that was based on equal weightings for calculating the skill-dimension ratings.

Appendix 2: Exercises

Our assessments featured the most common types of assessment-center exercises: *presentations*, *in-basket exercises*, *case studies*, *role plays*, and *group discussions* (see Spsychalski et al. 1997). The case-study exercise was originally designed by Stärk (2011), the presentation exercise by Eck et al. (2007), the role-play exercise by Siewert (2004), the in-basket exercise by Obermann (2013), and the group discussion by Kleinmann (2013).

The *case-study* exercise dealt with location planning in the cinema industry. Participants stepped into the role of the head of the marketing department for a movie-house chain that, given the wide diffusion of home-theater systems and streaming services, has experienced a sharp fall in revenues. Participants were provided with detailed information about the chain's theatres, including organization charts, revenues, debts, visitors, showroom sizes, food/drink offerings, and technical equipment. The participants' assignment was to provide a written statement on the chain's location policy on behalf of the top management and to justify their recommendations on whether to close some of the chain's theatres and to develop strategies on how to run the other theatres in the future. Participants had 45 min to prepare this statement. Some of the information they were provided, which included a detailed glossary, was not necessarily required to complete the assignment successfully, so the search for useful information was part of the challenge in this exercise.

In the *presentation* exercise, participants were put into a fictitious job-application situation. A short description explained their future tasks in the company (e.g., market research, business analyses, customer support) and the job requirements (e.g., flexibility, commitment, social competence). They had to apply for the job using a 5-min presentation to the assessors, who represented the company's board of directors. The assessors did not ask questions during or after the presentations.

The *role-play* exercise put participants in the situation of a middle manager who is working for a company in the satellite-reception industry and observes employees celebrating with a glass of sparkling wine during work time. The participant was told that one of the employees, who was described as committed, loyal, and popular among colleagues, worked in the participant's department. The company had a strict anti-alcohol policy that established alcohol consumption as a reason for termination, so the employee was ordered to attend a meeting with the participant/manager, which set the stage for the role play. The purpose and content of this meeting, which took 10 min, were not fixed, so participants could either decide to fire the employee or to risk conflict with top management. A Ph.D. student from our department took the role of the employee and was provided with a script on how to react based on the participant's arguments. For example, he argued either that they had only celebrated the successful completion of an advanced training course, which he took for the good of the company, or that there was alcohol at the company's last Christmas party and other events. If the participant decided to fire him, he acted shocked and said he had heard that the other employees with whom he celebrated had not been fired by the other department heads. If the participant decided to not fire him,

he acted relieved and said that the employees from the other departments had been fired.

The *in-basket exercise* put participants in the situation of a middle manager of a cleaning company who had just returned from a holiday. Participants were provided with a short description of the company, including an organization chart, and with their assignment, which was to read and process sixteen e-mails, several of which were related and required immediate action. Participants had 30 min to read the e-mails and prepare and take notes, and another 60 min to answer as many of the e-mails as possible. To facilitate preparation and note-taking, we provided writing materials and print versions of the e-mails, but the participants answered the e-mails on laptops on which we had pre-installed standardized pdf templates that mimicked an e-mail software. Participants had to justify their decisions in the broader context of the company and, as time was an issue in this exercise, to explain how they prioritized the e-mails, for which the templates provided additional space.

Finally, in the *group discussion*, participants were put into a board meeting of a Swiss bank that was recruiting a manager for a new business unit that would be responsible for asset and securities management. They were provided with background information on the bank and on the job requirements, including the required practical experiences, language skills, and academic records. The participants received short CVs from eight short-listed applicants and were asked to review the materials and take notes independently before deciding collectively on one of the applicants. This assignment was a challenge, as the background information on the job requirements with which they were provided differed among them. During the group discussion, which took 30 min, they had to rule out the candidates one by one. Their assignment was all the more challenging because they were not allowed to review the background information during the discussion but only the notes they took during the preparation period, and because only one of the applicants fulfilled all of the requirements.

Appendix 3: Exercises, skill dimensions, and skills

Table 5 shows the exercise/competency matrix we created to establish a link between the skill dimensions, the more measurable and specific skills, and the exercises. Table 6 provides descriptions and references for each skill we assessed.

Table 5 Exercise/competency matrix

Skill dimensions	Skills	Exercises				
		Presentation	Role play	Case study	In-basket exercise	Group discussion
Awareness of others	Open communication		x			
	Interpersonal sensitivity		x			
	Team work					x
	Organizational orientation				x	
	Perception of social cues				x	
	Team building					x
Communication	Active listening					x
	Clarity of speech	x				
	Spelling and grammar			x		
	Verbal ability	x				
	Writing style			x		
Drive	Results orientation				x	x
	Initiative					x
Influencing others	Leadership				x	
	Persuasiveness		x			x
Organizing and planning	Coaching				x	
	Delegation				x	
	Strategic thinking			x		
	Planning and scheduling			x	x	
	Structuring and organizing			x		
	Time sensitivity	x				
Problem-solving	Solution finding				x	
	Decisiveness				x	
	Problem analysis			x	x	
	Fact finding			x		x

Table 6 Skill dimensions and skills

Skill dimensions	Skills	Description	Exemplary references
Awareness of others	Open communication	The ability to foster open communication and encourage others to share their ideas and concerns	Arthur et al. (2003)
	Interpersonal sensitivity	The ability to show consideration for the situations of subordinates	Lievens (1999)
	Team work	The ability to cooperate and work well with others in pursuit of common goals	Lievens et al. (2003)
	Organizational orientation	The ability to recognize the impact of decisions on other components of an organization	Thornton and Byham (1982)
Communication	Perception of social cues	The ability to recognize indirect hints in the environment as required on the job	Russell and Domm (1995)
	Team building	The ability to develop supportive relationships with team members and to create a team spirit	Lievens et al. (2003)
	Active listening	The ability to use facial expressions and body language to show that one is listening	Lievens (1999)
	Clarity of speech	The ability to speak grammatically correctly and to use appropriate verbal language	Jackson et al. (2005)
	Spelling and grammar	The ability to write grammatically correctly and to use appropriate written language	Thornton and Byham (1982)
	Verbal ability	The ability to speak clearly, fluently, and at an appropriate volume and pace	Lievens et al. (2003)
	Writing style	The ability to convey and interpret information and express ideas through written means	Love and DeArmond (2007)
Drive	Results orientation	The ability to show commitment to achieving high quality in work performance and reaching goals	Goffin et al. (1996)
Influencing others	Initiative	The ability to influence events actively rather than passively and to take action to achieve goals	Thornton and Byham (1982)
	Leadership	The ability to guide others towards task accomplishment and/or achievement of goals	Love and DeArmond (2007)
	Persuasiveness	The ability to obtain agreement or acceptance and to persuade or influence others	Rupp et al. (2003), as cited in Thornton and Rupp (2006)

Table 6 (continued)

Skill dimensions	Skills	Description	Exemplary references
Organizing and planning	Coaching	The ability to guide and help subordinates to develop, and facilitate their professional growth	Goffin et al. (1996)
	Delegation	The ability to allocate tasks, decisions, and other responsibilities to appropriate subordinates	Thornton and Byham (1982)
	Strategic thinking	The ability to think strategically, holistically, and long term about decisions	Ulrich (1987)
	Planning and scheduling	The ability to set goals and priorities and to identify and initiate goal-relevant actions	Goffin et al. (1996)
Problem-solving	Structuring and organizing	The ability to arrange ideas and organize information in a systematic manner	Love and DeArmond (2007)
	Time sensitivity	The ability to recognize time limitations of situations and to accomplish tasks within these limits	Ross and Wolter (1998)
	Solution finding	The ability to identify a problem situation and formulate alternative ways to solve a problem	Love and DeArmond (2007)
	Decisiveness	The ability and readiness to make decisions, render judgments, and take action	Thornton and Byham (1982)
	Problem analysis	The ability to find the underlying causes of problems	Lievens (1999)
	Fact finding	The ability to look for information and to identify the advantages and disadvantages of available options	Lievens (1999)

Appendix 4: Inter-rater reliability

As the rating scales were ordinal, we measured the assessors' level of agreement using Kendall's coefficient of concordance. As Table 7 shows, all coefficients of concordance were significant, so inter-rater reliability was generally high.

Table 7 Kendall's coefficient of concordance

Exercise ^a	Skills	W	<i>p</i>
Case study	Fact finding	.944	.001
	Planning and scheduling	.930	.001
	Problem analysis	.919	.001
	Spelling and grammar	.954	.001
	Strategic thinking	.930	.001
	Structuring and organizing	.930	.001
	Writing style	.936	.001
Group discussion	Active listening	.909	.001
	Initiative	.955	.001
	Fact finding	.962	.000
	Persuasiveness	.952	.001
	Results orientation	.912	.001
	Team building	.906	.001
	Team work	.886	.002
Presentation	Clarity of speech	.936	.001
	Time sensitivity	.956	.001
	Verbal ability	.919	.001
Role play	Interpersonal sensitivity	.958	.000
	Open communication	.963	.000
	Persuasiveness	.951	.001

^aFor the in-basket exercise, we used a detailed and comprehensive checklist of yes-or-no questions for each e-mail. Since the rating was straightforward and did not require subjective judgment, we did not calculate measures of inter-rater reliability for this exercise

References

- Adachi PJC, Willoughby T (2013) More than just fun and games: the longitudinal relationships between strategic video games, self-reported problem solving skills, and academic grades. *J Youth Adolesc* 42:1041–1052
- Anderson N (2003) Applicant and recruiter reactions to new technology in selection: a critical review and agenda for future research. *Int J Sel Assess* 11:121–136
- Archambeau DJ (1979) Relationships among skill ratings assigned in an assessment center. *J Assess Center Technol* 2:7–20
- Armstrong MB, Ferrell JZ, Collmus AB, Landers RN (2016) Correcting misconceptions about gamification of assessment: more than SJTs and badges. *Ind Organ Psychol* 9:671–677
- Arthur W Jr, Woehr DJ, Maldegen R (2000) Convergent and discriminant validity of assessment center dimensions: a conceptual and empirical reexamination of the assessment center construct-related validity paradox. *J Manag* 26:813–835
- Arthur W Jr, Day EA, McNelly TL, Edens PS (2003) A meta-analysis of the criterion-related validity of assessment center dimensions. *Pers Psychol* 56:125–154
- Ballantyne I, Povah N (2004) *Assessment and development centres*, 2nd edn. Routledge, London
- Barber CS, Petter SC, Barber D (2017) It's all fun and games until someone gets a real job!: from online gaming to valuable employees. In: *Proceedings of the 38th international conference on information systems*, Seoul
- Barr M (2017) Video games can develop graduate skills in higher education students: a randomised trial. *Comput Educ* 113:86–97
- Barr DJ, Levy R, Scheepers C, Tily HJ (2013) Random effects structure for confirmatory hypothesis testing: keep it maximal. *J Mem Lang* 68:255–278
- Basak C, Boot WR, Voss MW, Kramer AF (2008) Can training in a real-time strategy videogame attenuate cognitive decline in older adults? *Psychol Aging* 23:765–777
- Bellotti F, Kapralos B, Lee K, Moreno-Ger P, Berta R (2013) Assessment in and of serious games: an overview. *Adv Hum Comput Interact* 2013:1–11
- Bender JM (1973) What is “typical” of assessment centers? *Personnel* 50:50–57
- Bretz RD Jr (1989) College grade point average as a predictor of adult success: a meta-analytic review and some additional evidence. *Public Pers Manag* 18:11–22
- Bycio P, Alvares KM, Hahn J (1987) Situational specificity in assessment center ratings: a confirmatory factor analysis. *J Appl Psychol* 72:463–474
- Caldwell C, Thornton GC III, Gruys ML (2003) Ten classic assessment center errors: challenges to selection validity. *Public Pers Manag* 32:73–88
- Camargo C (2006) Interesting complexity: Sid Meier and the secrets of game design. *Crossroads* 13(2):4. <https://doi.org/10.1145/1217728.1217732>
- Carpenter N (2016) Education-focused Civilization game heading to schools in 2017. IGN. www.ign.com/articles/2016/06/24/education-focused-civilization-game-heading-to-schools-in-2017. Accessed 20 Dec 2019
- Chamorro-Premuzic T, Winsborough D, Sherman RA, Hogan R (2016) New talent signals: shiny new objects or a brave new world? *Ind Organ Psychol* 9:621–640
- Chen H-C, Naquin SS (2006) An integrative model of competency development, training design, assessment center, and multi-rater assessment. *Adv Dev Hum Resour* 8:265–282
- Chu SKW, Reynolds RB, Tavares NJ, Notari M, Lee CWY (2017) 21st century skills development through inquiry-based learning: from theory to practice. Springer, Singapore
- Cohen PA (1984) College grades and adult achievement: a research synthesis. *Res High Educ* 20:281–293
- Deterding S, Dixon D, Khaled R, Nacke L (2011) From game design elements to gamefulness: defining “gamification”. In: *Proceedings of the 15th international academic MindTrek conference*, Tampere
- Dye DA, Reck M (1989) College grade point average as a predictor of adult success: a reply. *Public Pers Manag* 18:235–241
- Eck CD, Jöri H, Vogt M (2007) *Assessment-Center*. Springer, Heidelberg
- Efron L (2016) How gaming is helping organizations accelerate recruitment. *Forbes*. www.forbes.com/sites/louisefron/2016/06/12/how-gaming-is-helping-organizations-accelerate-recruitment/#450c23e53d5. Accessed 20 Dec 2019

- Ferrell JZ, Carpenter JE, Vaughn ED, Dudley NM, Goodman SA (2016) Gamification of human resource processes. In: Gangadharbatla H, Davis DZ (eds) *Emerging research and trends in gamification*. IGI Global, New York, pp 108–139
- Fetzer M (2015) Serious games for talent selection and development. *Ind Organ Psychol* 52:117–125
- Gelman A, Hill J (2007) *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press, New York
- Ghiselli EE (1973) The validity of aptitude tests in personnel selection. *Pers Psychol* 26:461–477
- Gibbons AM, Rupp DE (2009) Dimension consistency as an individual difference: a new (old) perspective on the assessment center construct validity debate. *J Manag* 35:1154–1180
- Glass BD, Maddox WT, Love BC (2013) Real-time strategy game training: emergence of a cognitive flexibility trait. *PLOS ONE* 8:1–7
- Goffin RD, Rothstein MG, Johnston NG (1996) Personality testing and the assessment center: incremental validity for managerial selection. *J Appl Psychol* 81:746–756
- Gorsuch RL (1983) *Factor analysis*, 2nd edn. Lawrence Erlbaum Associates, Hillsdale
- Granic I, Lobel A, Engels RCME (2014) The benefits of playing video games. *Am Psychol* 69:66–78
- Hamari J, Koivisto J, Sarsa H (2014) Does gamification work? A literature review of empirical studies on gamification. In: *Proceedings of the 47th Hawaii International conference on system sciences*, Waikōloa Village, pp 3025–3034
- Hinneburg A, Mannila H, Kaislaniemi S, Nevalainen T, Raumolin-Brunberg H (2007) How to handle small samples: Bootstrap and Bayesian methods in the analysis of linguistic change. *Lit Linguist Comput* 22:137–150
- Hinrichs TR, Forbus KD (2007) Analogical learning in a turn-based strategy game. In: *Proceedings of the 20th international joint conference on artificial intelligence*, Hyderabad, pp 853–858
- Huotari K, Hamari J (2012) Defining gamification: a service marketing perspective. In: *Proceedings of the 16th international academic MindTrek conference*, Tampere, pp 17–22
- Jackson DJR, Stillman JA, Atkins SG (2005) Rating tasks versus dimensions in assessment centers: a psychometric comparison. *Hum Perform* 18:213–241
- Jansen PG, Stoop BA (2001) The dynamics of assessment center validity: results of a 7-year study. *J Appl Psychol* 86:741–753
- Joiner DA (2000) Guidelines and ethical considerations for assessment center operations: International Task Force on Assessment Center Guidelines. *Public Pers Manag* 29:315–332
- Kapp KM (2012) *The gamification of learning and instruction: game-based methods and strategies for training and education*. Pfeiffer, San Francisco
- Kato PM, de Klerk S (2017) Serious games for assessment: welcome to the jungle. *J Appl Testing Technol* 18:1–6
- Ke F, Shute V (2015) Design of game-based stealth assessment and learning support. In: Loh CS, Sheng Y, Ifenthaler D (eds) *Serious games analytics: methodologies for performance measurement, assessment, and improvement*. Springer, Cham, pp 301–318
- Keith MJ, Anderson G, Gaskin JE, Dean DL (2018) Team video gaming for team building: effects on team performance. *AIS Trans Hum Comput Interact* 10:205–231
- Kim YJ, Shute VJ (2015) The interplay of game elements with psychometric qualities, learning, and enjoyment in game-based assessment. *Comput Educ* 87:340–356
- Kleinmann M (2013) *Assessment-Center*, 2nd edn. Hogrefe, Göttingen
- Kokkinakis AV, Cowling PI, Drachen A, Wade AR (2017) Exploring the relationship between video game expertise and fluid intelligence. *PLOS ONE* 12:1–15
- Kruschke JK (2014) *Doing Bayesian data analysis: a tutorial with R, JAGS, and Stan*, 2nd edn. Academic Press, London
- Lievens F (1999) Development of a simulated assessment center. *Eur J Psychol Assess* 15:117–126
- Lievens F, Harris MM, Van Keer E, Bisqueret C (2003) Predicting cross-cultural training performance: the validity of personality, cognitive ability, and dimensions measured by an assessment center and a behavior description interview. *J Appl Psychol* 88:476–489
- Lisk TC, Kaplancali UT, Riggio RE (2012) Leadership in multiplayer online gaming environments. *Simul Gaming* 43:133–149
- Liu D, Li X, Santhanam R (2013) Digital games and beyond: what happens when players compete? *MIS Q* 37:111–124
- Love KG, DeArmond S (2007) The validity of assessment center ratings and 16PF personality trait scores in police sergeant promotions: a case of incremental validity. *Public Pers Manag* 36:21–32

- Lu L, Shen C, Williams D (2014) Friending your way up the ladder: connecting massive multiplayer online game behaviors with offline leadership. *Comput Hum Behav* 35:54–60
- McCallum S (2012) Gamification and serious games for personalized health. In: Blobel B, Pharow P, Sousa F (eds) Ebook: pHealth 2012. Studies in health technology and informatics, vol 177, pp 85–96. <http://ebooks.iospress.nl/volume/phealth-2012>. Accessed 20 Dec 2019
- Mead AD, Drasgow F (1993) Equivalence of computerized and paper-and-pencil cognitive ability tests: a meta-analysis. *Psychol Bull* 114:449–458
- Michael D, Chen S (2006) Serious games: games that educate, train, and inform. Course Technology, Mason
- Mislevy RJ, Corrigan S, Oranje A, DiCerbo K, Bauer MI, von Davier AA, John M (2016) Psychometrics and game-based assessment. In: Drasgow F (ed) Technology and testing: improving educational and psychological measurement. Routledge, New York, pp 23–48
- Moore GR, Shute VJ (2017) Improving learning through stealth assessment of conscientiousness. In: Marcus-Quinn A, Hourigan T (eds) Handbook on digital learning for K-12 schools. Springer, Cham, pp 355–368
- Morelli N, Potosky D, Arthur W Jr, Tippins N (2017) A call for conceptual models of technology in I–O psychology: an example from technology-based talent assessment. *Ind Organ Psychol* 10:634–653
- Morgeson FP, Campion MA, Dipboye RL, Hollenbeck JR, Murphy K, Schmitt N (2007) Reconsidering the use of personality tests in personnel selection contexts. *Pers Psychol* 60:683–729
- Nebel S, Schneider S, Rey GD (2016) Mining learning and crafting scientific experiments: a literature review on the use of Minecraft in education and research. *Educ Technol Soc* 19:355–366
- Neidig RD, Martin JC, Yates RE (1979) The contribution of exercise skill ratings to final assessment center evaluations. *J Assess Cent Technol* 2:21–23
- Obermann C (2013) Assessment Center: Entwicklung, Durchführung, Trends, 5th edn. Springer, Wiesbaden
- Owens T (2011) Modding the history of science: values at play in modder discussions of Sid Meier's Civilization. *Simul Gaming* 42:481–495
- Peng W, Lee M, Heeter C (2010) The effects of a serious game on role-taking and willingness to help. *J Commun* 60:723–742
- Petter S, Barber D, Barber CS, Berkley RA (2018) Using online gaming experience to expand the digital workforce talent pool. *MIS Q Exec* 17:315–332
- Quiroga MA, Herranz M, Gómez-Abad M, Kebir M, Ruiz J, Colom R (2009) Video-games: do they require general intelligence? *Comput Educ* 53:414–418
- Quiroga MA, Román FJ, De La Fuente J, Privado J, Colom R (2016) The measurement of intelligence in the XXI century using video games. *Spanish J Psychol* 19:1–13
- Reilly RR, Chao GT (1982) Validity and fairness of some alternative employee selection procedures. *Pers Psychol* 35:1–62
- Reilly RR, Henry S, Smither JW (1990) An examination of the effects of using behavior checklists on the construct validity of assessment center dimensions. *Pers Psychol* 43:71–84
- Richards G (2012) From gamer to racing driver. *The Guardian*. www.theguardian.com/sport/2012/apr/29/jann-ardenborough-racing-car-games. Accessed 20 Dec 2019
- Ross CM, Wolter SA (1998) Hiring the right person: using the assessment center as an alternative to the traditional interview process. *Recreational Sports Journal* 22:38–44
- Roth PL, BeVier CA, Switzer FS III, Schippmann JS (1996) Meta-analyzing the relationship between grades and job performance. *J Appl Psychol* 81:548–556
- Rupp DE, Gibbons AM, Runnels T, Anderson L, Thornton GC III (2003) What should developmental assessment centers be assessing? Paper presented at the 63rd annual meeting of the Academy of Management, Seattle
- Russell CJ, Domm DR (1995) Two field tests of an explanation of assessment centre validity. *J Occup Organ Psychol* 68:25–47
- Sackett PR, Dreher GF (1982) Constructs and assessment center dimensions: some troubling empirical findings. *J Appl Psychol* 67:401–410
- Sapieha C (n.d.) Sid Meier's Civilization V: game review. Common Sense Media. www.commonsensemedia.org/game-reviews/sid-meiers-civilization-v. Accessed 20 Dec 2019
- Schmidt FL, Hunter JE (1998) The validity and utility of selection methods in personnel psychology: practical and theoretical implications of 85 years of research findings. *Psychol Bull* 124:262–274
- Schmidt FL, Hunter J (2004) General mental ability in the world of work: occupational attainment and job performance. *J Pers Soc Psychol* 86:162–173

- Schmitt N, Gooding RZ, Noe RA, Kirsch M (1984) Metaanalyses of validity studies published between 1964 and 1982 and the investigation of study characteristics. *Pers Psychol* 37:407–422
- Shreve J (2005) Let the games begin. *Edutopia* 1:29–31
- Shute V (2015) Stealth assessment in video games. Paper presented at the 2015 Australian Council for Educational Research (ACER) research conference, Melbourne, pp 61–64
- Shute VJ, Moore GR (2017) Consistency and validity in game-based stealth assessment. In: Jiao H, Lissitz RW (eds) *Technology enhanced innovative assessment: development, modeling, and scoring from an interdisciplinary perspective*. Information Age Publishing, Charlotte, pp 31–51
- Shute VJ, Ventura M, Bauer M, Zapata-Rivera D (2009) Melding the power of serious games and embedded assessment to monitor and foster learning: flow and grow. In: Ritterfeld U, Cody M, Vorderer P (eds) *Serious games: mechanisms and effects*. Routledge, New York, pp 295–321
- Shute VJ, Ventura M, Ke F (2015) The power of play: the effects of Portal 2 and Lumosity on cognitive and noncognitive skills. *Comput Educ* 80:58–67
- Shute VJ, Wang L, Greiff S, Zhao W, Moore G (2016) Measuring problem solving skills via stealth assessment in an engaging video game. *Comput Hum Behav* 63:106–117
- Siewert HH (2004) Spitzenkandidat im Assessment-Center: Die optimale Vorbereitung auf Eignungstests, Stressinterviews und Personalauswahlverfahren. Redline, Frankfurt am Main
- Simons A, Weinmann M, Fleischer S, Wohlgenannt I (2015) Do good gamers make good students? Sid Meier's Civilization and performance prediction. In: *Proceedings of the 36th international conference on information systems*, Fort Worth
- Slovensky R, Ross WH (2012) Should human resource managers use social media to screen job applicants? Managerial and legal issues in the USA. *Info* 14:55–69
- Snijders TAB, Bosker RJ (2012) *Multilevel analysis: an introduction to basic and advanced multilevel modeling*, 2nd edn. Sage, London
- Spychalski AC, Quiñones MA, Gaugler BB, Pohley K (1997) A survey of assessment center practices in organizations in the United States. *Pers Psychol* 50:71–90
- Squire KD (2004) *Replaying history: learning world history through playing "Civilization III"*. Ph.D. thesis, Indiana University. <https://www.learntechlib.org/p/125618/>. Accessed 20 Dec 2019
- Squire K, Barab S (2004) Replaying history: engaging urban underserved students in learning world history through computer simulation games. In: *Proceedings of the 6th international conference on learning sciences*, Santa Monica, pp 505–512
- Squire K, Steinkuehler C (2005) Meet the gamers. *Libr J* 103:38–41
- Stanger M (2016) How an obsession with Football Manager could earn you a career in the game. *The Guardian*. www.theguardian.com/football/the-set-pieces-blog/2016/sep/29/football-manager-2017-sport-plymouth-argyle. Accessed 20 Dec 2019
- Stärk J (2011) *Assessment-Center erfolgreich bestehen: Das Standardwerk für anspruchsvolle Führungs- und Fach-Assessments*, 13th edn. Gabal, Offenbach
- StataCorp (2019) *Stata multilevel mixed-effects reference manual: release 16*. StataCorp LLC, College Station
- Stone DL, Deadrick DL, Lukaszewski KM, Johnson R (2015) The influence of technology on the future of human resource management. *Hum Resour Manag Rev* 25:216–231
- Stoughton JW, Foster Thompson L, Meade AW (2015) Examining applicant reactions to the use of social networking websites in pre-employment screening. *J Bus Psychol* 30:73–88
- Straus SG, Miles JA, Levesque LL (2001) The effects of videoconference, telephone, and face-to-face media on interviewer and applicant judgments in employment interviews. *J Manag* 27:363–381
- Susi T, Johannesson M, Backlund P (2007) *Serious games: an overview*. Technical Report HS-IKI-TR-07-001. School of Humanities and Informatics, University of Skövde. <https://www.diva-porta.lorg/smash/get/diva2:2416/FULLTEXT01.pdf>. Accessed 20 Dec 2019
- Testa A (2014) Religion(s) in videogames: historical and anthropological observations. *Heidelb J Relig Internet* 5:249–278
- Thornton GC III, Byham WC (1982) *Assessment centers and managerial performance*. Academic Press, New York
- Thornton GC III, Gibbons AM (2009) Validity of assessment centers for personnel selection. *Hum Resour Manag Rev* 19:169–187
- Thornton GC III, Rupp DE (2006) *Assessment centers in human resource management: strategies for prediction, diagnosis, and development*. Lawrence Erlbaum Associates, Mahwah
- Ulrich D (1987) Organizational capability as a competitive advantage: human resource professionals as strategic partners. *Hum Resour Plan* 10:169–184

- Van de Vijver FJR, Harsveld M (1994) The incomplete equivalence of the paper-and-pencil and computerized versions of the General Aptitude Test Battery. *J Appl Psychol* 79:852–859
- Van Eck R (2006) Digital game-based learning: it's not just the digital natives who are restless. *EDUCAUSE Rev* 41:16–30
- Van Eck RN, Shute VJ, Rieber LP (2017) Leveling up: game design research and practice for instructional designers. In: Reiser RA, Dempsey JV (eds) *Trends and issues in instructional design and technology*, 4th edn. Pearson, Upper Saddle River, pp 227–285
- Vichitvanichphong S, Talaie-Khoei A, Kerr D, Ghapanchi AH, Scott-Parker B (2016) Good old gamers, good drivers: results from a correlational experiment among older drivers. *Austr J Inf Syst* 20:1–21
- Wang L, Shute V, Moore GR (2015) Lessons learned and best practices of stealth assessment. *Int J Gaming Comput Mediated Simul* 7:66–87
- Woehr DJ, Arthur W Jr (2003) The construct-related validity of assessment center ratings: a review and meta-analysis of the role of methodological factors. *J Manag* 29:231–258
- Wright T, Boria E, Breidenbach P (2002) Creative player actions in FPS online video games: playing Counter-Strike. *Game Stud* 2
- Zide J, Elman B, Shahani-Denning C (2014) LinkedIn and recruitment: how profiles differ across occupations. *Empl Relat* 36:583–604

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Alexander Simons¹ · Isabell Wohlgenannt¹ · Markus Weinmann² · Stefan Fleischer³

Isabell Wohlgenannt
isabell.wohlgenannt@uni.li

Markus Weinmann
weinmann@rsm.nl

Stefan Fleischer
stefan.fleischer@ercis.uni-muenster.de

- ¹ Department of Information Systems, University of Liechtenstein, Fuerst-Franz-Josef-Strasse, 9490 Vaduz, Principality of Liechtenstein
- ² Rotterdam School of Management, Erasmus University, Burgemeester Oudlaan 50, 3062 PA Rotterdam, The Netherlands
- ³ European Research Center for Information Systems (ERCIS), University of Münster, Leonardo-Campus 3, 48149 Münster, Germany