

Validation of the Team Trust Game: a Serious Game for the Assessment of Cooperation Skills

Name: Karch Kooijman

Student number: 432625

Supervisor: Dr. Alec Serlie

Erasmus University Rotterdam

Master Positive Organizational Psychology

Abstract

People in organizations must work together in some way or another in order to reach goals. Effective cooperation is beneficial for organizations as it can lead to increased productivity, motivation and job satisfaction. The current study addresses the need for a new measurement method for assessing cooperation skills, by developing a serious game: *The Team Trust Game*. Compared to traditional measures like self-rated questionnaires, serious games are found to be less prone to faking and socially desirable answers. Furthermore, in an assessment context, serious games can provide insight into individual differences between intention and behaviour. A theoretical framework for cooperation was made consisting of five cooperation components: coaching, collaboration, empathy, integrity and loyalty. Additionally, the study looked into the relationship between cooperation and trust and knowledge sharing and the personality dimensions extraversion and agreeableness. The quality of this new type of measurement was assessed by its construct validity, congruent validity, face validity, and usability. Seventy-nine participants filled in a self-rated questionnaire and played the serious game. Results showed that even though participants had positive perceptions of the serious game, it did not show significant correlations between cooperation as measured in the game and the self-rated cooperation. Future studies should further investigate the use of serious games for the measurement of competencies like cooperation within an assessment context.

Keywords: *Cooperation, Extraversion, Agreeableness Trust, Knowledge Sharing, Assessment,*

Introduction

People in organizations must work together in some way or another in order to reach goals. To our knowledge, there are very few, if any, jobs in which collaborating with others is not required. In recent years global and technological developments have brought new challenges and opportunities to the way people cooperate. For example, the pandemic outbreak of Covid-19 which reached the western world in 2020 suddenly forced many people to work from home. Working from home comes with a set of unique benefits and challenges (Feitosa & Salas, 2020). While on the one hand working in casual clothes or doing household chores between meetings can be seen as positive adjustments, on the other hand dealing with connection issues or family matters can become quite troublesome when the domestic environment becomes the workplace. Due to technological developments, people now have access to new communication devices like Microsoft Teams and Zoom that allow them to interact and cooperate virtually. Moreover, this novel, virtual aspect of cooperation unveils new factors that influence effective cooperation (Alsharo et al, 2017).

Because effective cooperation directly alters the way in which a team performs and the quality of the work they deliver, the impact of these factors is worth examining. In recent years the workplace has changed and become more complex in many ways. An example of this is that jobs tend to include more non-routine and contain interactive tasks (Neubert et al., 2015). This shift has led to new challenges for organizations and therefore to new job requirements. Researchers have identified a wide range of skill sets workers should possess that accompany these new job requirements. These skill sets are regularly labelled as 21st-century skills (Neubert et al., 2015).

A recent study (Rios et al., 2020), attempted to empirically rank-order these 21st-century skills that are necessary for workplace success. To do so, the scholars identified the most in-demand 21st-century skills present in job advertisements. Through job listings, employers are able to directly communicate their expectations of the work skills that potential employees should possess in order to perform well on the job (Rios et al., 2020). One of the main findings was that collaboration skills are in high demand by employers (Rios et al., 2020).

It is hardly a surprise that employers would want their employees to be able to work together as effective cooperation has positive effects on organizational outcomes (Klein et al., 2008). Research has shown that cooperation is an important work group characteristic that

boosts team performance in different ways. For instance, teams are more productive when individuals are able to cooperate effectively (Campion et al., 1993). Likewise, cooperation can have positive effects on an individual level, as employees tend to be more intrinsically motivated and more satisfied with their job when they are given the opportunity to work together with their colleagues (Campion et al., 1993; Tauer, & Harackiewicz, 2004).

The increasingly interactive nature of jobs requires individuals from various backgrounds to be able to work together when solving problems (Neubert et al., 2015). Yet, the individual ability to solve complex problems is not enough when employees are working in teams. If a group has to find a solution to a problem together, they must be able to share their understanding and perception of the effort required to reach the said solution. The pooling of knowledge, skills, and efforts to do so is essential when dealing with and solving a problem (Neubert et al., 2015). Significantly, the importance of these collaborative problem-solving skills has not exclusively been highlighted by research on 21st-century skills, but also by research on competencies. Collaborative problem solving skills show great resemblance with competencies like analytical thinking, communication and teamwork (Sliter, 2015).

Competencies can be described as a combination of knowledge, skills, abilities, and other individual differences that are required for a job (Campion et al., 2011). The traditional approach to job requirements focuses on discrete tasks and traits that must be met for a person to perform well at their job. The competency approach, however, describes these requirements more freely because it allows a combination of different skills and abilities (Campion et al., 2011). This makes the competency approach especially applicable to modern jobs that are constantly changing and becoming more complex (Neubert et al., 2015; Sliter, 2015). While the competency approach still tries to identify what is most important for job performance, its focus shifts towards the person and the combination of characteristics that are required to do well on that job.

Great Eight Competency framework by Bartram (2005)

In their research, Kurz and Bartram (2002) describe competencies as sets of behaviours that are instrumental in the delivery of desired results or outcomes. They see competencies as multi-dimensional since they can include a variety of capabilities, activities, and responses that enable a wide range of work demands (Kurz & Bartram, 2002). Here, the competency itself is defined in its significance for performance at work, and the included sets of behaviours are specified in actions (Kurz & Bartram, 2002; Bartram, 2005).

Organizations can benefit from using competencies for assessment because their goals, culture, and values can be expressed in terms of the behaviour that is expected of employees. Furthermore, competency-based assessment encourages person-specific discussions for development because it does not solely focus on antecedent person factors like personality and general cognitive ability as predictors for job-relevant behaviour (Bartram, 2005). Instead, individual differences are identified and measured in terms of specific work-related constructs and observable behaviours that are important for successful job performance (Kurz & Bartram, 2002; Bartram, 2005).

Kurz and Bartram (2002) developed the Great Eight Competency framework based on extensive content analyses of academic and practice-based competency models. This framework consists of three levels: a top tier, a middle tier, and a bottom tier. The top tier presents the eight broad competency factors that were derived from factor analysis. The middle tier consists of 20 competency dimensions that load on the eight competency factors. The bottom tier of the framework consists of a set of 110 component competencies. These components are defined as behavioural indicators and can be considered as the building blocks for a specific competency set. Table 1 shows the eight competency factors and the 20 competency dimensions of the competency framework by Kurz & Bartram (2002).

Table 1

The Great Eight Competencies and 20 Competency Dimensions (Kurz & Bartram, 2002)

<i>Factor</i>	<i>Competency domain title</i>	<i>Competency Dimensions</i>
1	Leading and Deciding	1.1 Deciding & Initiating Action 1.2 Leading and Supervising
2	Supporting and Cooperating	2.1 Working with People 2.2 Adhering to Principles and Values
3	Interacting and Presenting	3.1 Relating & Networking 3.2 Persuading and Influencing 3.3 Presenting and Communicating Information
4	Analyzing and Interpreting	4.1 Writing and reporting 4.2 Applying Expertise and Technology 4.3 Analyzing
5	Creating and Conceptualizing	5.1 Learning and Researching 5.2 Creating and Innovating 5.3 Formulating Strategies and Concepts
6	Organizing and Executing	6.1 Planning and Organizing 6.2 Delivering Results and Meeting Customer Expectations 6.3 Following Instructions and Procedures
7	Adopting and Coping	7.1 Adapting and Responding to Change 7.2 Coping with Pressure and Setbacks

This study focuses on one of the great eight competency factors: Supporting and Cooperating. The definition of this competency factor consists of three components: putting people first, working effectively with individuals and teams, and behaviour that is consistent with personal and organizational values (Bartram, 2005). Likewise, this competency factor has two competency dimensions: *Working with people* and *Adhering to Principles and Values* (see Table 1). Table 2 shows the competency components of the bottom tier. These are the behavioural indicators for the competency factor Supporting and Cooperating and its competency dimensions *Working with people* and *Adhering to Principles and Values*.

Table 2

Competency components of the Supporting & Cooperating competency factor (Bartram, 2005)

Supporting & Cooperating	
<i>Working with people</i>	<i>Adhering to Principles and Values</i>
- Understanding Others	- Upholding Ethics and Values
- Adapting to the Team	- Acting with Integrity
- Building Team Spirit	- Utilizing Diversity
- Recognizing and Rewarding Contributions	- Showing Social and Environmental Responsibility
- Listening	
- Consulting Others	
- Communicating Proactively	
- Showing Tolerance and Consideration	
- Showing Empathy	
- Supporting Others	
- Caring for Others	
- Developing and Communicating Self-Knowledge and Insight	

The following section will focus on different components of cooperation that are often discussed in cooperation literature. A theoretical conceptualization of cooperation as a competency will be developed based on these components, and subsequently used for the remainder of the study.

Components of cooperation

Personality. When it comes to predicting workplace behaviour, personality continues to be a major topic of interest. One of the most accepted models of personality is the Big Five model by Costa and McCrae (1997). This model distinguishes five personality traits: Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. These personality traits

have been linked to the competency factors, discussed in the previous section. Two of the former personality traits seem to contribute to an individual's ability to cooperate due to their interpersonal nature (Bell, 2007).

The first personality trait that is linked with cooperation is extraversion. Extraverted people are characterized by being sociable, energetic, assertive, and ambitious (Barrick et al., 2005). These individuals tend to be comfortable when making contact with others and easy to get along with. They are also likely to have an open personality and show attraction towards their team (Porter et al., 2003). Furthermore, extraversion has been shown to be related to other team processes like seeking help from others when needed (Bell, 2007). In this sense, extraversion appears to be a predictor of job performance for occupations where interacting with others plays a significant role (Barrick et al., 2001). Similarly, Bartram (2005) found that extraversion was related to competency factors that revolve around interpersonal interactions.

The second personality trait linked to cooperation is agreeableness. Agreeable people are characterized by being considerate, trusting, friendly, helpful, and tolerant (Barrick et al., 2005). They tend to pursue positive social interactions by seeking social harmony and avoiding competition or conflict within the team (Bell, 2007). Like extraversion, agreeableness has been found to be a predictor of job performance for occupations that involve interpersonal interactions, especially when such interaction involves helping, cooperating, and nurturing others (Mount et al., 1998; Barrick et al., 2005). Additionally, Bartram (2005) found a relation between agreeableness and the competency factor Cooperating and Supporting. Similar to extraversion, behaviours that fall under this competency factor involve interpersonal interaction (Bartram, 2005).

Both extraversion and agreeableness are expected to relate to behavioural tendencies across a wide range of interpersonal settings (Koole et al., 2001). By virtue of this tendency towards interpersonal interaction, extraverted and agreeable individuals are likely to have an easier time cooperating with others. As follows, these personality traits contribute to effective cooperation and can therefore be described as the antecedent factors of effective cooperation. Given that in this study cooperation is seen as a competency, the ability to cooperate is not only dependent on personality traits, but also on specific behavioural components (Bartram, 2005).

Cooperation components. On the behavioural level, the framework by Bartram (2005) allows us to distinguish the five components of cooperation described below. The first one is *collaboration*, which can be described as the aim to achieve a common goal by way of providing helpful communication and by disregarding personal interests (Kurz & Bartram, 2002). The second behavioural component is *coaching*, which includes behaviours such as

'helping others to identify opportunities for personal growth and development' and 'helping others to make use of their competencies' (Kurz & Bartram, 2002). Coaching is perceived as a contributing factor for cooperation because of its focus on helping coworkers. The third behavioural component is *empathy* (Kurz & Bartram, 2002). The latter involves exhibiting concern for and openness towards the feelings and attitudes of others. It is likely that showing empathy improves communication and consequently one's ability to cooperate with coworkers. The fourth component is *integrity* and can be described as the behaviours that maintain social and ethical standards at work (Kurz & Bartram, 2002). This becomes relevant because the disregard for such standards will hinder successful cooperation between individuals. The last component is *loyalty*, which implies behaving accordingly to the policies and interests of the organization or group (Kurz & Bartram, 2002). Loyalty contributes to cooperation as working with someone uninterested in the group's general standpoints becomes difficult (Kurz & Bartram, 2002).

Trust and Knowledge sharing. The next two components of cooperation, *trust* and *knowledge sharing*, do not directly result from the framework by Kurz & Bartram (2002). However, research has highlighted the importance of these components by investigating their relation with effective virtual teams. A virtual team is similar to a traditional team, as it consists of a group of people working together towards a shared goal or objective. A big difference, however, is that members of a virtual team often do not work at the same location but are instead geographically dispersed around the world (Malhotra et al., 2007; Pangil & Chan, 2014). In this sense, members of a virtual team depend on technology to be able to communicate with each other. With the help of new technological communication programs such as Zoom and Teams, virtual teams within organizations are no longer bound to bring team members to one specific location (Pangil & Chan, 2014). Another distinctive feature of virtual teams is that they work interdependently while still sharing responsibility for their outcomes (Malhotra et al., 2007).

Virtual teams face different challenges compared to traditional teams. For example, if team members are based in distant locations, time zone disparities may prompt meeting scheduling issues. The most obvious challenge is that virtual team members do not interact face-to-face. This can lead to less opportunities to pick up on non-verbal cues, have informal conversations or build friendships (Nunamaker et al., 2009). Despite having a lack of face-to-face interactions, virtual teams are still expected to perform. Effective communication and collaboration in a technological work environment will be essential in order for the team to

succeed (Alsharo et al., 2017). Research has focused on finding team processes that help virtual teams with overcoming challenges and make them effective.

The first component in effective virtual teams is *trust* among team members. Trust is an important factor in maintaining social relationships and effective teamwork (Alsharo et al., 2017). In a trusting environment, people are more likely to believe that their actions will have positive consequences because they are able to collaborate with other team members that are willing to help (Pangil & Chan, 2014). However, building this trusting environment is more difficult for virtual teams because team members are not able to meet face-to-face. Members of a virtual team that have never worked with each other have no past (work) performance or (work) relation to base their trust on. In such cases, individuals will have to develop the trust that others will complete their own tasks effectively and efficiently (Malhotra et al., 2007; Pangil & Chan, 2014). Since virtual team members are not able to directly monitor the work of other team members, their trust is based on the perceptions that members have of each other. These perceptions are built on the interactions that team members have with each other. Teams that interact frequently will therefore have an easier time building a trusting environment (Pangil & Chan, 2014). In virtual teams, these interactions are held with the help of technological communication programs which can lead to fewer interactions among team members. It is important for virtual teams to be able to develop trust through both formal and informal interactions. New communication programs can provide better face-to-face communication, making it easier to pick up on body language and facial expressions (Morrison-Smith & Ruiz, 2020).

Overall, research has repeatedly shown that trust is an important factor both in effective face-to-face and virtual teams (Alsharo et al., 2017; Pangil & Chan, 2014). Trust has been found to be a key variable that is crucial for cooperation and team effectiveness and can be seen as the glue that holds a team together (Breuer et al., 2016; De Jong et al., 2016). In a trusting environment, team members are more inclined to ask for help, share feedback, and discuss issues and conflicts (Breuer et al., 2016).

The second component that occurs in effective teams is *knowledge sharing*. In a team, all members possess individual knowledge or expertise that is necessary to complete their own tasks. In a virtual team, members working on a common task are openly asked to interact and share their individual knowledge to complete the objective (Pangil & Chan, 2014). In that sense, knowledge sharing is perhaps more evident in a virtual environment. By sharing knowledge, teams are able to create new knowledge that can help them with their tasks. Research has shown that the occurrence of knowledge sharing is significantly related to virtual

team effectiveness (Pangil & Chan, 2014; Alsharo et al., 2017). It is essential that team members are willing to share their knowledge with each other because this will be beneficial to the way they cooperate and perform (Alsharo et al., 2017). Vice versa, withholding relevant information has the opposite effect and can lead to missed objectives. Furthermore, knowledge sharing has been found to be connected with trust (Staples & Webster, 2008). When team members trust each other they will be more convinced that others will share their knowledge in order to cooperate effectively. This can make them feel obligated to participate in knowledge sharing (Staples & Webster, 2008). Furthermore, the development of trust can be positively influenced by knowledge sharing (Alsharo et al., 2017). In short, it can be concluded that both trust and knowledge sharing can be seen as contributors and therefore components of cooperation.

Based on the findings discussed above we suggest a main theoretical concept for cooperation through the behavioural components: *collaboration, coaching, empathy, integrity and loyalty* (Bartram, 2005). Furthermore, we suggest that the personality traits extraversion and agreeableness are antecedents components of cooperation (Bell, 2007). Lastly, we suggest that *trust* and *knowledge sharing* are determinant components for effective cooperation (Alsharo et al., 2017). The following section will discuss potential assessment methods for cooperation. Then, the main goal of the study will be presented along with the hypotheses.

Assessment of cooperation

Traditional methods. Effective cooperation can lead to many positive outcomes for organizations and employees, as described above. Therefore, organizations could benefit from knowing about their employees' competence to cooperate. This assessment of competencies can be done in various ways, for example by self-assessment. A problem with self-assessment, however, is that people are influenced by their self-perception and therefore have a tendency to overestimate or underestimate their own abilities (Harris & Schaubroeck, 1988; Kim et al., 2016). Moreover, self-assessments are prone to faking and social desirability (Dubbelt et al., 2014).

In combination with self-assessment, organizations can implement evaluations rated by others. Other people are found to be better at evaluating the capabilities of an individual because they are not or at least less influenced by this self-perception (Harris & Schaubroeck, 1988). In assessment, the addition of other-rated evaluations is based on the idea of 360-degree feedback. In this type of feedback the assessment of subordinates, peers, and/or supervisors is added to a self-reported assessment. The goal of 360-degree feedback is to provide a complete

review of someone's competencies based on the assessment of a variety of people that this person is working with. The inclusion of 360-degree feedback is found to be especially helpful if people overestimate their own capabilities since it can help them become more self-aware (Johnson & Ferstl, 1999; Brett & Atwater, 2001).

While traditional assessment methods are able to give insight into a broad range of constructs, they are often considered as boring and repetitive (Shute & Rahimi, 2017). In recent years, new methods for assessment have been developed that show more promise regarding the experience of the assessment.

Situational Judgment Tests. Another method to measure and assess constructs that are related to work performance is with situational judgment tests (SJTs). SJTs assess individual judgement by presenting a problem scenario and a list of response options (Whetzel et al., 2020). Through these scenarios, SJTs are able to address job-related problems and through the chosen responses SJT's can measure the application of knowledge or behavioural tendencies (Whetzel et al., 2020). Research has suggested that the implementation of SJTs can be helpful when measuring teamwork and other interpersonal skills to predict job performance (Christian et al., 2010). One specific advantage of using SJTs for assessment is that they show incremental validity over cognitive ability and personality (Whetzel et al., 2020). Furthermore, studies suggest that especially for SJT's that focus on behavioural tendency rather than the application of knowledge, group differences are found to be smaller in SJT's compared to cognitive ability tests (Lievens et al., 2008; Whetzel et al., 2020).

SJTs can be presented in several formats, for example, text-based and video-based. With video-based SJT's actors are being hired to play out the problem scenario. However, the development of video-based SJTs can be very costly with regard to money and time. Developing avatar-based SJT's can be a less costly method, while still holding advantages over text-based SJT's (Whetzel et al., 2020). Avatar-based SJTs are less costly because computer-generated avatars are used to interact with the participants instead of real actors. Moreover, they are easier to edit since it will not be necessary to reshoot an entire scene when changes need to be made (Whetzel et al., 2020).

Video- and avatar-based SJTs differ from written SJTs in terms of fidelity, which refers to the extent the format is able to represent the job environment (Oostrom et al., 2011). By implementing relevant pictures, audio, and video elements to an SJT it is possible to create a detailed and realistic job-related scenario. This increased fidelity of the test can lead to more favorable reactions by the applicants regarding interest and motivation in comparison with traditional paper and pencil tests (Woods et al., 2019; Dubbelt et al., 2014). Research has also

shown that high fidelity SJTs are able to show higher levels of criterion-related validity compared to low fidelity SJTs (Lievens & Sackett, 2006; Oostrom et al., 2011).

Gamification of SJTs

The implementation of multimedia elements into SJTs has been taken a step further through a process called gamification. Gamification is a process where game elements are incorporated into non-gaming activities in any context, such as the workplace (Georgiou et al., 2019). This process has been applied to assessment in order to make the experience more fun, engaging, and meaningful (Gkorezeis et al., 2020; Georgiou et al., 2019). Gamified assessments are assessments that have been enhanced with game-like elements such as the use of a cover story, avatars, movement systems, and scoring (Gkorezis et al., 2020). Previous research (Armstrong et al., 2016) has suggested that the use of gamified assessment methods is able to reduce faking and socially desirable responses. Through the game design and elements used in the assessment, applicants' attention can be shifted from the fact that their responses are being recorded. Consequently, they are more likely to show their true behaviours and reduced faking and socially desirable answers (Armstrong et al., 2016). This increases the quality of information about applicants and therefore the prediction of job performance. In their study, Georgiou et al. (2019) converted a traditional SJT into a gamified assessment and found support for the construct validity of this instrument. They showed that this new gamified assessment was suitable for the measurement of interpersonal competencies like resilience, adaptability, flexibility, and decision-making because it focuses on behaviour and not self-reported traits (Georgiou et al., 2019). Since the scenarios in the gamified assessment are able to assess work-related behaviours, they are likely to better predict future work behaviours compared with self-reported assessment methods (Armstrong et al., 2016; Georgiou et al., 2019).

Serious games

While gamified assessments incorporate game elements to the assessment, they are not considered as fully developed and stand-alone games. There has been a trend of using fully developed serious games in contexts like education, healthcare, the military, and more recently assessment (Kato & de Klerk, 2017). Serious games are digital games that are designed to achieve goals beyond entertaining, and are instead used for educational and training purposes (Kato & de Klerk, 2017; Naul & Liu, 2019). Moreover, serious games can be used for

assessment purposes by identifying personal traits, skills, and competencies (Barends et al., 2021).

Serious games have the potential to be even more fun, immersive and engaging compared to the earlier discussed assessment methods. Like normal games, serious games are able to implement game design approaches like reward systems, freedom of choice, and interactive environments. Furthermore, narratives in serious games can vary from fantasy worlds to realistic everyday contexts.

Serious games can be differentiated from gamified assessments with regard to the greater use of storyline elements and narrative in their game design (Naul & Liu, 2019). The use of narrative in serious games is important because it is able to promote immersion, engagement, and motivation (Naul & Liu, 2019). When players become immersed in the game through the narrative and game environment they can experience time loss and the feeling that they are being transported to the game world (Naul & Liu, 2019). This feeling of immersion can make players less aware of the original goal of the serious game, for example, the fact they are completing an assessment. Narratives in serious games can also positively influence engagement. This refers to the attention and interest of the players to the tasks in the game. Wouters et al. (2011) showed in their study that the implementation of foreshadowing and back story led to greater self-reported curiosity which contributed to prolonged engagement during the game. Lastly, the use of narrative in serious games is able to positively influence motivation. Through narrative elements like fantasy, challenges, freedom of choice, and meaningful interactions the game is likely to become more fun and interesting (Naul & Liu, 2019; Westera, 2019). What follows is that players will be more intrinsically motivated to play the game (Westera, 2019).

Previous research has suggested that serious games can be used in an assessment context to measure and assess a broad range of skills and constructs that are not easily measured with traditional assessment methods (Kato & de Klerk, 2017; Barends et al., 2021). Regarding the psychometric properties of serious games, studies (Dubbelt et al., 2014; Barends et al., 2021). have shown that serious games are able to show construct validity and predictive validity As with gamified assessments, faking becomes more difficult in serious games since the desirable answer is less obvious while playing the game (Armstrong et al., 2016). Moreover, serious games are likely to receive more favourable reactions from applicants compared to traditional pen and paper tests (Dubbelt et al., 2014; Georgiou et al., 2019).

Serious games have the potential to become powerful measurement and learning tools for behaviour-based skills and competencies. Through the implemented technology in serious

games, it is possible to record behaviours and evaluate them after playing. When used for assessment, serious games can give insight into a particular skill or competency that is not developed in the game itself (Barends et al., 2021). Combining the results of the serious game with self-reported data on the same skill or competency can provide information on potential differences between intention and behaviour, which can be useful for future development of this skill or competency (Georgiou et al., 2019).

To our knowledge, there is no one consistent method to assess cooperation. Moreover, previous methods like the cooperation and competition scale by Simmons et al. (1988) can be considered outdated for today's working environment. A new possible method to assess a competency like cooperation is with a serious game. This combines the need for an up-to-date assessment tool for cooperation and the several potential advantages that serious games have over traditional questionnaires (Dubbelt et al., 2014; Kato & de Klerk, 2017; Barends et al., 2021).

In a time where people make substantial use of digital methods as a means to communicate and cooperate at work, it could be helpful to first get an idea of how an applicant behaves in such a simulated environment of a serious game. Moreover, it will be possible to assess potential differences between intention and behaviour. The self-rated questionnaire measures the intention to cooperate and the game measures the actual behaviours (Georgiou et al., 2019). The scores on both the self-rated questionnaire and the serious game can give insight into which areas of cooperation the applicant can develop. For example, high cooperation scores on the self-rated questionnaire and low cooperation scores on the serious game can indicate that the applicant is not aware of his or her own behaviour and needs training. Vice versa, low cooperation scores on the self-rated questionnaire and high cooperation scores on the serious game can indicate that the applicant underestimates his or her own cooperation capabilities. With the help of the serious game, the applicant can be reminded of using these cooperation skills during the assessment process.

Present study

In this study, we aim to validate a serious game called *The Team Trust* game that has been designed to measure cooperation skills. The game puts the player in a scenario where they have to interact with different types of characters. The player is presented with scenes fitting a storyline along with responses in a multiple choice format and is asked to respond in a way that they think is most fitting. For the majority of the interactions in the game, the player is given

four different responses. These responses are set up so that sometimes they differ greatly and other times they differ slightly. Because of this, the player is either able to respond quickly when an answer is obvious as well as slowly when they want to consider the difference in details. The game makes sure to let the player know that their responses are being recorded by reacting to the responses in specific ways. The players can face different dialogue in the game depending on how they play it. Moreover, the player is given a certain amount of free choice which consequently influences the course and possibly the ending of the game. Despite there being different endings in the game, the player can't fail. This is because the game was designed for assessment purposes with the intention of being a positive experience.

Team Trust Game

The Team Trust Game follows a scenario where the player gets assigned as the new Project Manager for the *Vaccine Defense Consortium*. This consortium consists of a team that developed a working vaccine for an unknown virus. However, the team experiences difficulties with the production of the vaccine. This is because some team members have different priorities when it comes to the quality or quantity of the vaccine. What follows is that team members start to distrust each other, which negatively affects the way they cooperate. The player will follow the production process of the vaccine and interact with either the team as a whole or with members separately. Through these interactions, the player gets to know more about the personality, motivations, or even personal problems of each team member. Considering these factors, the player gets the opportunity to restore trust within the team so that they can find a solution to the production problem together. Appendix C provides a description of the story, characters, and game design including pictures of the Team Trust Game.

Validity analysis and hypotheses

To establish the serious game as a new and valid measurement tool for cooperation skills, it is necessary to analyze the game on criteria like reliability, validity, fairness and usability (Cook, 2016). A measurement method is considered reliable when the results of the method end in the same results whenever it is used (Cook, 2016). However, the Team Trust Game that is used in this study is newly developed and has no prior data. Moreover, the serious game measures cooperation behaviours through dynamic and interactive simulation gameplay. A consequence of this is that players can chose different paths while playing the game and therefore not all players will be shown the same scenes. Since there is, to our knowledge, no

appropriate way to measure the reliability of such an instrument, it will not be used as one main criterion for analysis.

Another criterion that is used to evaluate measurement tools is validity. Validity is the extent to which the method measures what it claims to measure (Cook, 2016). The first type of validity that will be analyzed is construct validity. Construct validity reveals what a method is actually measuring and can be assessed by comparing one method with another (Cook, 2016). For this study, the serious game as a new measurement method will be compared with a traditional measurement method in the form of self-rated questionnaires. Previous studies were able to find support for the construct validity of new digital measures in such a manner (Dubbelt et al., 2014; Georgiou et al., 2019; Barends et al., 2021). Therefore, we hypothesize:

H1: Cooperation scores of the serious game will be positively correlated with cooperation scores on the self-questionnaire

Earlier, we discussed two antecedents of cooperation in the form of the personality traits extraversion and agreeableness. Research suggests that people with higher levels of extraversion and agreeableness are likely to have an easier time cooperating (Barrick et al., 2005; Bartram, 2005; Bell, 2007). Therefore, we expect that the antecedents of cooperation measured with a self-rating questionnaire and cooperation measured with the serious game will be positively correlated. To establish the relationship between these components, we will compare the cooperation scores of the serious game with self-rated extraversion and agreeableness. Based on previous findings we hypothesize:

H1b: Cooperation scores of the serious game will be positively correlated with self-rated extraversion.

H1c: Cooperation scores of the serious game will be positively correlated with self-rated agreeableness

The second type of validity that will be analyzed is congruent validity. Congruent validity is the extent to which tests that measure the same concept correlate with each other. In this study, the congruent validity will be assessed by comparing the results of the serious game with results on both self- and other-rated questionnaires that measure cooperation through the same concepts. Self-questionnaires alone are not sufficient because they tend to be prone to biases like faking and social desirability (Dubbelt et al., 2014). Furthermore, the use of other-rated questionnaires can be meaningful as they are not influenced by the self-perceptions of the

participants (Harris & Schaubroeck, 1988). This study provides a 360-degree review of cooperation as it is measured with the serious game, a self-questionnaire and an other-rated questionnaire. We expect that the cooperation scores of these measures correlate with each other because they ought to measure the same concept. The hypotheses regarding the congruent validity of the serious game are formulated as followed:

H2a: Cooperation scores on the self-rated questionnaire will be positively correlated with cooperation scores on the other-rated questionnaire.

H2b: Cooperation scores of the serious game will be positively correlated with cooperation scores on the other-rated questionnaire

To further investigate the relationship between scores of the serious game and scores on the other-rated questionnaire we will look at the components of trust and knowledge sharing. Earlier we discussed trust and knowledge sharing as two components of effective cooperation. Research has shown that both trust and knowledge sharing can positively influence cooperation in effective teams (Alsharo et al., 2017; Pangil & Chan, 2014). Therefore, we expect that trust and knowledge sharing as components of cooperation can be effectively evaluated by colleagues. We hypothesize:

H2c: Trust scores measured by the serious game will be positively correlated with trust scores on the other-rated questionnaire.

H2d: Knowledge sharing scores measured by the serious game will be positively correlated with knowledge sharing scores on the other-rated questionnaire.

The next type of validity that will be assessed in this study is face validity. This is the extent to which the serious game appears to measure cooperation as a job-related construct according to the participants. Face validity is strongly connected with acceptability, another criterion that is used to analyze measurement methods (Cook, 2016). Acceptability is the extent to which the assessment process and results are found to be credible by the participants (Norcini et al., 2011). A subtype of face validity that includes the perceptions of the participants is perceived predictive validity, which can be described as the extent to which participants think the measure is able to predict who will perform better on the measured concept.

H3a: Participants will have positive perceptions regarding the face validity of the serious game.

H3b: Participants will have positive perceptions regarding the perceived predictive validity of the serious game.

Additionally, fairness will be studied. Fairness is the extent to which the measurement method provides equal opportunities for all the participants (Cook, 2016). One way of analysing the fairness of a method is by evaluating the perceptions and reactions of the participants to the selection method. The reactions of participants to a selection method are important for several reasons. For example, they can influence the pursuit or acceptance of job offers (Smither et al., 1993). Participants can experience a loss of belief in efficacy, when they feel that they have been negatively assessed on tasks that measure unchangeable abilities or attributes. Perceived unfairness may also lower motivation to perform well on the test which can lead to biased or inaccurate scores (Kluger & Rothstein, 1993). Likewise, research has suggested that perceived fairness can be influenced by the extent to which participants are allowed to demonstrate their knowledge or skills on a selection method (Gilliand, 1993). When participants feel they have the opportunity to perform, they are more likely to perceive the method as fair regardless of the outcome of the assessment (Gilliand, 1993).

H3c: Participants will have positive perceptions regarding the fairness of the serious game.

Previous studies have shown that applicant reactions towards gamified assessments and serious games are more favourable when compared with traditional pen and paper tests (Dubbelt et al., 2014; Georgiou et al., 2019). We, therefore, expect that the overall perceptions of the serious game will be positive. The overall face validity of the game will be analyzed through scores on face validity, perceived predictive validity, fairness perception and opportunity to perform. The corresponding hypotheses are formulated as followed:

H3d: Participants will have positive perceptions of the opportunity to perform in the serious game.

The last criterion that will be used to analyze the game is usability. Usability can be described as the extent to which participants find the assessment method easy to use (Cook, 2016). Because the serious game is in an early stage of development it is useful to know if the participants find the game easy to use. Moreover, the results of the study might be negatively influenced if participants consider the game difficult to use. We formulate:

H4: Participants will have positive perceptions regarding the usability of the serious game.

In summary, to analyze the newly developed *Team Trust Game* we will evaluate the construct validity, congruent validity, overall face validity and usability. Furthermore, we will look at the relationship between the cooperation scores of the serious game and other components that are believed to influence cooperation such as trust, knowledge sharing, extraversion and agreeableness. Next, we will discuss the procedure of the study in the method section. After the method section, the results of the study will be presented. Lastly, the discussion will include a further review of the findings together with the limitations of the study and suggestions for future studies.

Method

Procedure/participants

Due to the COVID restrictions imposed, a convenience sample was used as the method for finding participants. This means that a sample of the population was taken by asking people who were easy to contact and reach to participate in the study. The first group of people that were asked consisted of friends, family and colleagues. A short text message was sent to them that explained the study briefly. Some individuals forwarded this message to members of their own network. With the help of this snowball sampling method, new potential participants were added to the sample. People who were interested in participating were asked for their email address. These people were then sent a research invitation which included additional information on the study. The Dutch version of this e-mail can be found in Appendix B and the English version in Appendix C. Taken together, 120 persons were invited to participate in the study.

Afterward, a separate email was sent that included a link to Flowsparks, the game platform where the research was conducted. By following the link, the participants were able

to create an account on Flowsparks which allowed them to participate in the study. A short text reintroducing the study was presented in the home screen and after clicking the continue button the participants were able to follow the study step by step. The first step of the study was a questionnaire that included background questions. Participants were asked for permission to use their results for research at the end of this questionnaire. Without permission, it was not possible to proceed with the research. Of the 120 persons who were invited to participate in the study, 95 gave their consent and were able to officially start the study (response rate = 79.2%).

For the second step, the participants had to fill in self-rating questionnaires. This questionnaire measured the five components of cooperation (coaching, collaboration, empathy, loyalty and integrity) as well as the personality traits extraversion and agreeableness through self-rating. After completing this questionnaire, the participants were allowed to play the serious game. 93 participants of the 95 managed to complete the self-rated questionnaire (response rate = 97.9%).

The third step of the study was playing the *Team Trust Game*, which was provided in English. If the game was played in one sitting, participants took an hour on average to complete the game. However, since the study was conducted from home the participants were able to pause and resume the game at a later time. In addition, participants were able to track their progress in the game through an indicator that stated the completed percentage. At any stage of the study, it was possible to discontinue participation without further explanation.

93 persons started the game and 79 persons managed to complete the game (response rate = 84.9%). 16 persons did not complete the game for unknown reasons. Of the 79 participants, 46 were male (58.2%) and 33 were female (41.8%). 58% of the participants belonged to the age category below 25 years, 16% to 25-34 years, 9% to 35-44 years, 9% to 45-54 years and 8% to 55 years and above. For 18% of the participants, the highest level of education was high school, for 25% college, for 27% a bachelor's degree and for 30% a master's degree. The fourth step of the study included a short evaluation questionnaire that consisted of 23 items. This questionnaire included items to measure face validity, fairness perceptions, perceived predictive validity, opportunity to perform and user experience. 77 of the 79 participants finished the evaluation questionnaire (response rate = 97.5%).

For the last part of the study, participants were asked to invite a colleague or person who they generally collaborated with for the 360-degree questionnaire. The participants were able to fill-in the email of people they work with or had worked with in the past. These people were then sent an email with a link to the 360-degree questionnaire, where they were asked to rate the cooperation qualities of the participant. There were 62 participants out of the 77 that

had a colleague, friend or relative (response rate = 80.5%) answer questions about them. The amount of evaluators per participants ranged from 1 to 3 with a mode of one co-worker per participant.

Material

Background variables. Before participants started the game, they were asked to fill in a questionnaire that included some background questions. People were asked about their gender (male, female, rather not tell), age (<25 years, 25-34, 35-44, 45-54, 55 and above) and their highest level of education (high school, college, bachelor, masters, other)

Cooperation. Cooperation was measured in the following three ways. Before playing the game, cooperation was measured with a self-rated competency questionnaire that included items on the five components of cooperation (coaching, collaboration, empathy, integrity and loyalty). The self-rated cooperation questionnaire consisted of 20 items, with four items for each component. Items were rated on a 5-point Likert scale ranging from 1 (NO!) to 5 (YES!). For the component coaching an example of an item is: "Guides subordinates and colleagues when they are familiarizing themselves with the organization". An example of an item for the component collaboration is: "Puts the combined results before one's own results". For the cooperation component empathy, an example of an item is: "Respects and accepts others as they are in emotional situations". An example of an item for the component integrity is: "Is open and honest; will not withhold information to which the other party is entitled". Lastly, an example for the component loyalty is: "Does not abandon team members when they are under attack". Reliability analysis showed an alpha coefficient of .82 for the self-rated cooperation scale.

Next, cooperation was measured with the *Team Trust Game*. After playing through the game, participants were scored on the five components of cooperation. These scores were computed by adding the pre-established scores on a cooperation component that were linked to the responses in the game. Table 3 shows an example of a game scene that measures the cooperation component empathy, including response options and pre-established scores.

After playing the game, cooperation was again measured with a 5-point Likert scale ranging from 1 (NO!) to 5 (YES!). However, this time the questionnaire was filled in by colleagues (other-rated cooperation). Reliability analysis showed an alpha coefficient of .78 for the other-rated cooperation scale.

Table 3. *Example of game scene that measures Empathy*

Game scene dialogue	Response options	Score on cooperation component
MEETING ROOM - DAY (1:1) <i>JEAN-MARC looks a bit nervous.</i>	A. How could you have been so careless?	Empathy + 0
JEAN-MARC: What's going on here?	B. I'm not very pleased with the way you handled this, we've emphasized secrecy so much.	Empathy + 25
PLAYER: A "multi-purpose miracle"? You are the only one that uses that term so you must have been the source.	C. I understand this must be very awkward to you. How did it happen?	Empathy + 100
JEAN-MARC: It must have been a slip of the tongue between the talks. How could I know that someone would post it online?	D. You've stirred the water here, I've had Virtanen shouting at me.	Empathy + 50

Trust and Knowledge Sharing. Trust and Knowledge Sharing were measured with the *Team Trust Game* as the additional components of cooperation. Scores on these components were given based on the participants' behaviour and responses in the serious game. For example, participants received scores on knowledge sharing if they had chosen a response that shared relevant information with the characters in the game. Likewise, participants received scores on trust if they had chosen a response that showed belief in the team. Additionally, participants received an overall score on trust and knowledge sharing on the other-rated questionnaire. Participants were given a single rating ranging from 1 to 10.

Extraversion and Agreeableness. The personality traits extraversion and agreeableness were measured with a self-rated questionnaire. Participants were asked to rate 12 items on both extraversion and agreeableness on a Likert scale that ranged from 1 (NO!) to

5 (YES!). An example of an item for extraversion was: "Finds it easy to get on with new people". For agreeableness, an example of an item was: "Allows others to stand in the limelight". The items on extraversion and agreeableness had an alpha coefficient of .72 and .70 respectively.

Face validity. Seven items were used to measure face validity. These items were adapted from the face validity scale by Smither, Reilly, Millsap, Pearlman, and Stoffey (1993) which had an alpha coefficient of .86. Reliability analysis showed an alpha coefficient of .79. Participants were asked to rate these items on a 5-point Likert scale ranging from 1 (NO!) to 5 (YES!). An example of an item that measured face validity is: " I could see the relationship between the Game and what is required in my current job".

Perceived predictive validity. Three items were used to measure perceived predictive validity. These items were adapted from the predictive validity scale by Smither et al., (1993) which reported an alpha coefficient of .83. Reliability analysis showed an alpha coefficient of .60. Participants were asked to rate the items on a 5-point Likert scale ranging from 1 (NO!) to 5 (YES!). An example of an item that measured perceived predictive validity is: "I am confident that the outcomes on the Game can predict how well an employee will perform their job".

Fairness perception. Four items were used to measure the fairness perceptions of the participants. These items were taken from the fairness perceptions scale by Kluger and Rothstein (1993) which reported an alpha coefficient of .81. Reliability analysis showed an alpha coefficient of .71. Participants were asked to rate these items on a 5-point Likert scale ranging from 1 (NO!) to 5 (YES). An example of an item that measured fairness perception is: "I believe that this type of measurement can predict whether I will be a successful employee".

Opportunity to perform. Four items were used to measure the opportunity to perform. These items were taken from the Selection Procedural Justice Scale (SPJS) by Bauer, Truxillo, Sanchez, Ferrara & Campion (2001). The article stated that the SPJS had alpha coefficients ranging from .73 to .92 (Bauer et al., 2001). Reliability analysis showed an alpha coefficient of .76. Participants were asked to rate these items on a 5-point Likert scale ranging from 1 (NO!) to 5 (YES!). An example of an item that measured the opportunity to perform is: "This type of measurement gives people the opportunity to show what they can really do".

Usability. Five items were used to measure usability. These items were taken from the Usability Metric for User Experience that was developed by Finstad (2010). The article mentioned an alpha coefficient of .94, but reliability analysis showed an alpha coefficient of .61. Participants were asked to rate these items on a 5-point Likert scale ranging from 1 (NO!) to 5 (YES!). An example of an item that measured usability is: "The structure of the information in the Game was clear".

Analyses

Game score analysis

The Team Trust game was developed on an online learning platform named Flowsparks (<https://www.flowsparks.com/nl/>). This platform is able to create a simulation-style serious game based on a script. By uploading media files, the platform is able to create a visual display of the game. These files can then be linked to the script, in order to make the right picture appear with its corresponding scene or moment. The program now provides a playable storyline, where dialogue and scenes follow each other. A last notable feature of the program is the implementation and recording of game scores.

As stated before, the game consists of scenes where the player gets to interact with the characters. It's during these interactions that the components of cooperation are measured. In the development phase of the script, different scores were assigned to the responses that players are able to select during interactions. A response can consist of 0, 25, 50, 75 or 100 points for a specific component of cooperation (coaching, collaboration, empathy, integrity, or loyalty). For example, a player will receive 100 points on *empathy* if the most empathetic response is chosen. Likewise, a player will receive 0 points on this component if the least empathetic response is chosen. For these components, the difference in scores is based on the content of the response that is given to the characters in the game.

The scores on *Trust* and *Knowledge Sharing* are measured in a similar way. For example, the highest scores are linked to the responses that show the greatest degree of trust towards the team. For knowledge sharing, the highest scores are linked to the responses that share most of the acquired information. For these two components, the difference in scores is based on the behavioural degree of the response towards the characters.

After completing the Team Trust Game it is possible to determine a total score on a cooperation component. This is done by adding the scores of the responses that are related to that component. The next step is conducting the research, using the total scores of the cooperation components

For this research, a composite score of cooperation was made by adding the total scores of the components *Coaching*, *Collaboration*, *Empathy*, *Integrity* and *Loyalty*. These components are derived from the job competency framework made by Kurz and Bartram (2002). The scores on Trust and Knowledge Sharing are not combined, since it is expected that both are separately related to cooperation (Pangil & Chan, 2014; Alsharo, 2017).

Results

Game result analysis

An outlier analysis was conducted before testing the hypotheses. For hypotheses 1 and 2 the construct and congruent validity was measured by calculating the correlations between the game scores and the scores on both the self- and other-rated questionnaires. Regression analyses were used to further investigate the relationship between these variables. For hypotheses 3 and 4, a one-sample *t*-test was carried out to see if the five mean scores were statistically different from the value '3'. The value '3' was chosen since the response scale on the criteria ranged from 1 to 5 with the last being the most positive. By analyzing if the means were significantly higher than '3' we could conclude if the participants have positive views on the face validity, perceived predictive validity, fairness perception, opportunity to perform and usability.

Table 4. Descriptive statistics and correlations of variables

	Mean	SD	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.
1. Game Cooperation	332.75	44.95	-										
2. Self-rated Cooperation	4.03	.35	.09	-									
3. Other-rated Cooperation	4.10	.33	.01	.12	-								
4. Trust	136.71	52.17	.12	.13	.07	-							
5. Knowledge Sharing	212.66	74.02	.08	-.05	-.25*	-.04	-						
6. Extraversion	3.48	.62	.16	.24*	.04	0.15	0.05	-					
7. Agreeableness	3.90	.37	.20	.56*	.27*	0.11	-0.02	0.17	-				
8. Face validity	3.47	.65	-.04	-.10	.01	.18	-.09	-.01	-.24*	-			
9. Perceived predictive validity	3.20	.68	-.07	.00	.15	.02	-.06	-.01	.18	.56**	-		
10. Fairness perception	3.51	.53	-.04	.07	.17	-.13	-.01	.00	.07	.62**	.75**	-	
11. Opportunity to perform	3.24	.65	-.12	.04	.12	.03	-.11	-.14	.09	.47**	.64**	.59**	-
12. Usability	3.97	.55	-.02	.05	.18	-.02	.03	-.16	.05	.38**	.29*	.42**	.41**

Note. $N = 79$ for Game Cooperation, Self-rated Cooperation, Trust, Knowledge sharing, extraversion and Agreeableness. $N = 77$ for Face validity, Perceived predictive validity, Fairness perception, Opportunity to perform and Usability. $N = 62$ for Other-rated Cooperation, Other-rated Trust and Other-rated Knowledge sharing. For Other-rated Trust and Other-rated Knowledge sharing a report grade was given on a 1-10 rating scale.

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

Descriptive statistics

The means, standard deviations and correlations of the variables used in this study can be found in Table 4. The table presents three cooperation scores (game cooperation, self-rated cooperation and other-rated cooperation). These three cooperation scores were calculated by taking the mean of the cooperation component scores (coaching, collaboration, empathy, loyalty and integrity). The data were checked for outliers before testing the hypotheses. One potential outlier was found in the data set. This outlier was related to a low Game cooperation score with a Z-score of $Z = -2.95$. First it was checked if this participant fully completed the *Team Trust Game*, to make sure that the score was valid. After this was confirmed, the play time of this participant was analyzed. There were no irregularities regarding the play time of this participant. Therefore, it was decided to not remove this participant from the dataset.

Inter-rater reliability

As mentioned before, the other-rated cooperation score was computed by taking the mean of the cooperation component scores (coaching, collaboration, empathy, loyalty and integrity). In this study, only nine participants were rated on the other-rated questionnaire by at least two persons, which is why the inter-rater reliability was not further investigated.

Hypotheses testing

The first hypothesis H1 stated that the Game cooperation score and the self-rated cooperation score would be positively correlated. Results, as can be seen in Table 4, show a non-significant correlation that is close to zero between Game cooperation and self-rated cooperation. Thus, Hypothesis 1 was not supported.

Hypotheses H1b and H1c looked into the relationship between game cooperation and the personality traits extraversion and agreeableness. H1b stated that the game cooperation score would be positively correlated with self-rated extraversion. Results showed a small and insignificant correlation between these variables. H1c stated that the game cooperation score would be positively correlated with self-rated agreeableness. Results showed a small and insignificant correlation between the game cooperation score and self-rated agreeableness. Based on the results no support was found for both H1b and H1c. The descriptive table provided some additional findings relating to the two personality traits. First, as shown in table 4, a moderate correlation was found between self-rated extraversion and self-rated cooperation.

Second, moderate to high correlations were found between self-rated agreeableness and both self- and other-rated cooperation.

Hypothesis 2a expected a positive correlation between the self-rated cooperation score and the other-rated cooperation score. No significant correlation was found between these two variables, as can be seen in Table 4. Therefore, Hypothesis 2a was not supported. Hypothesis 2b expected a positive correlation between the game cooperation scores and the other-rated cooperation score. This hypothesis was not supported, as no significant correlation was found between these two variables.

The relationship between the game scores and the other-rated questionnaire was investigated further in H2c and H2d. Hypothesis 2c stated that the trust score measured by the game would be positively correlated with trust scores on the other-rated questionnaire. Results showed a non-significant correlation between trust measured by the game and trust rated by others. Therefore, hypothesis 2c was not supported. For hypothesis 2d a positive correlation was expected between the knowledge sharing score measured by the game and the knowledge sharing score on the other-rated questionnaire. No support was found for hypothesis 2d as results showed a non-significant correlation that's close to zero. The descriptive table provided two additional findings relating to the knowledge sharing score measured by the game. First, as shown in table 4, a moderate negative correlation was found between knowledge sharing measured by the game and other-rated cooperation. Second, a moderate negative correlation was found between knowledge sharing measured by the game and other-rated trust.

For hypothesis 3a to hypothesis 3d a one-sided *t*-test was conducted to determine if the mean scores of the five variables measured with the evaluation questionnaire were significantly different than the midrange value 3. Hypothesis 3a stated that the participants would have positive perceptions regarding the face validity of the serious game, which was measured with the scores of the face validity items. The face validity mean score ($M = 3.47$, $SD = .65$) was significantly higher than the midrange value of 3, $t(76) = 6.32$, $p < .001$. This result supports hypothesis 3a, meaning that the participants had positive perceptions regarding the face validity of the serious game.

Hypothesis 3b stated that the participants would have positive perceptions regarding the perceived predictive validity of the game. This was measured with the scores of the perceived predictive validity items. The mean score for perceived predictive validity ($M = 3.20$, $SD = .68$) was significantly higher than the midrange value of 3, $t(76) = 2.61$, $p = .01$. This result gives support for hypothesis 3b.

Hypothesis 3c focused on fairness perception. It was expected that participants would have positive perceptions regarding the fairness of the serious game. This was measured with the scores on the fairness items. The mean score for fairness perception ($M = 3.51$, $SD = .53$) was significantly higher than the midrange value of 3, $t(76) = 8.36$, $p < .001$. This result supports hypothesis 3c, indicating that the serious game was perceived as fair by the participants.

Hypothesis 3d stated that the participants would have positive perceptions of the opportunity to perform in the serious game. The mean score for opportunity to perform ($M = 3.24$, $SD = .65$) was significantly higher than the midrange value of 3, $t(76) = 3.31$, $p = .001$. This result supports hypothesis 3d, indicating that the participants had positive perceptions of the opportunity to perform in the serious game.

The fourth and final hypothesis focused on the usability of the serious game. It was expected that the participants would have positive perceptions regarding the usability of the serious game. This was measured with the scores on the usability items. The mean score for usability ($M = 3.97$, $SD = .55$) was significantly higher than the midrange value of 3, $t(76) = 15.36$, $p < .001$. This result gives support for hypothesis 4.

The descriptives table presented some additional findings relating to hypothesis 3 and hypothesis 4. As shown in table 4, high correlations were found between face validity, perceived predictive validity, fairness perception, opportunity to perform and usability.

Discussion

Effective cooperation is relevant for both employees and organizations as it can lead to positive outcomes like increased intrinsic motivation and team performance (Tauer & Harackiewicz, 2004; Alsharo et al., 2017). Traditional measurement methods for cooperation, like self-assessments, can be considered out of date as they are found to be repetitive and prone to faking (Harris & Schaubroeck, 1988; Shute & Rahimi, 2017). Recent studies have suggested that serious games can be used to effectively measure interpersonal competencies (Kato & de Klerk, 2017; Georgiou et al., 2019). However, to our knowledge, no serious game currently exists to measure cooperation. In this study, we developed and researched a serious game that intends to measure cooperation skills as an interpersonal competency based on the job-competency framework by Kurz & Bartram (2002). Furthermore, additional components of cooperation were added in the study based on the literature on virtual teams and personality

traits (Pangil & Chan, 2014; Bell, 2007). The serious game was then assessed by analyzing the construct validity, face validity, opportunity to perform and usability.

Findings

First, the construct validity of the serious game was analyzed by comparing the outcomes of the serious game as a new measurement method with a traditional, but well established self-rating questionnaire. Results showed a small non-significant correlation between the cooperation scores of the serious game and the cooperation scores of the self-rated questionnaire. This result indicates that the serious game and the self-rated questionnaire do not measure cooperation as the same construct. This result is not in line with previous studies that found support for the construct validity of gamified assessment and serious games using the same approach (Dubbelt et al., 2015; Georgiou et al., 2019; Barends et al., 2021).

It is possible that the self-rated questionnaire and the serious game do not measure cooperation as the same construct. For example, it could be that the self-rated questionnaire measures the intention to cooperate rather than actual cooperation behaviours. This could explain why the current study was not able to find a significant correlation between the self-rated questionnaire and the serious game. Previous research has already suggested that, within an assessment context, the combination of a self-rated questionnaire and serious game results can provide insight into potential differences between intention and behaviour (Georgiou et al., 2019). These insights can be used for the development and training of a competency, as specific feedback can be communicated with the assessee regarding their intentions and behaviours.

Additionally, the finding could be explained by the characteristics of the sample. Our sample did not solely consist of people who are currently working. Due to convenience sampling, because of COVID restrictions at the time of the research, students were also included in the sample. The serious game was initially designed to measure interpersonal cooperation skills by presenting a specific working environment. It could be possible that students do not have enough work experience to be familiar with the simulated working environment which could influence the way they are able to demonstrate their cooperation skills. Similarly, this could be the case for blue-collar workers that were also included in the sample. It could be that blue-collar workers have different experiences with cooperating compared with the portrayed scenario of the serious game.

Next, we looked into the relationship between cooperation and the personality traits extraversion and agreeableness. Previous research suggests that individuals with higher levels of extraversion and agreeableness are likely to be better at cooperating (Barrick & Mount,

2005; Bell, 2007). More specifically, these personality factors contribute to the interpersonal interaction between team members. Therefore, a theoretical concept of cooperation was made in the introduction where the personality traits extraversion and agreeableness function as antecedent components of cooperation. The cooperation scores of the serious game were compared with the scores on both self-rated extraversion and self-rated agreeableness. Contrary to the expectations, results were not able to show a significant correlation between the game scores and the personality traits. This finding indicates that for this study participants with higher levels of extraversion and agreeableness did not receive higher scores on cooperation as measured by the serious game. Analysis showed that a big portion of our sample scored high on both extraversion and agreeableness. Therefore, the findings could possibly be explained by a ceiling effect, as it becomes more difficult to find a correlation when a group scores high on one of the variables.

Another possible explanation for this finding is that the serious game is not able to provide real interpersonal interactions. For example, it might be difficult to detect behaviours related to extraversion and agreeableness as it is currently not possible for participants to initiate interactions or provide their own responses. It should be noted that, in line with previous research, results showed significant correlations between the self-rated cooperation scores and the scores on both self-rated extraversion and agreeableness (Barrick & Mount, 2005; Bell, 2007). These results indicate that participants with higher degrees of extraversion or agreeableness received higher scores on the self-rated cooperation questionnaire.

There was a significant correlation between agreeableness and other-rated cooperation, indicating that participants with higher levels of agreeableness received higher cooperation scores from others. As such, these results of the study support the idea that the personality traits extraversion and, even more so, agreeableness, are connected with an individual's innate ability to cooperate with others.

The second aspect analyzed was congruent validity. First, we looked into the relationship between the cooperation scores of the self-rated questionnaire and the other-rated questionnaire. Results showed no significant correlations between the cooperation scores on the self-rated questionnaire and the cooperation scores on the other-rated questionnaire. This finding indicates that these questionnaires did not measure cooperation as the same construct, even when the items of these questionnaires were identical.

Subsequently, the relationship between the cooperation scores of the serious game and the cooperation scores of the other-rated questionnaire was analyzed to investigate if the serious

game assessment of cooperation skills was similar to the assessment from coworkers. Again, the results were not able to show significant correlations between the cooperation scores of the serious game and the cooperation scores of the other-rated questionnaire indicating that they do not measure cooperation in the same manner. We expected to find a correlation between these measures as they have in common that they are able to assess shown behaviour and are less likely to be influenced by biases like faking and social desirability (Dubbelt et al., 2015).

These findings could possibly be explained by the sample for the “other-rated questionnaire”. Analysis showed that this sample did not solely consist of colleagues, which could have influenced the cooperation scores of the other-rated questionnaire, but rather that the rater were Evaluations made by colleagues should be more accurate since colleagues have experienced the actual cooperation abilities of the participants while working with them.

Furthermore, the study investigated the relationship between the scores of the serious game and the components trust and knowledge sharing. Trust and knowledge sharing were included in our theoretical concept of cooperation as determinant factors of cooperation. In this study, the game-measured scores on trust and knowledge sharing were compared with the scores on trust and knowledge sharing of the other-rated questionnaire. Results were not able to show significant correlations between the game scores and the other-rated questionnaire scores. This finding is not in line with previous studies that showed that both trust and knowledge sharing are able to positively affect cooperation in effective teams (Alsharo et al., 2017; Pangil & Chan, 2014).

One explanation for the findings relating to trust could be that the serious game did not provide a real opportunity to establish and develop trust. Previous research has indicated that face-to-face interactions are important to developing trust as they can provide non-verbal cues and informal conversations (Morrison-Smith & Ruiz, 2020; Pangil & Chan, 2014). It could be that it is difficult to develop and show trust in a simulated serious game where the characters rarely change facial expressions. An explanation for the finding regarding knowledge sharing could be that people are not familiar with the concept of knowledge sharing within a team. It could be that the term knowledge sharing has a negative connotation as it could be seen as giving away important information.

The overall face validity of the serious game was analyzed to investigate if the participants had positive perceptions of the serious game. Results show that participants had positive perceptions regarding the face validity of the serious game. This finding indicates that

the serious game was perceived as a credible measurement tool for cooperation. Additionally, results show that participants had positive perceptions regarding the perceived predictive validity of the serious game. This result indicates that participants thought the serious game was able to differentiate between those who are better at cooperating and those who are worse.

Subsequently, the scores on perceived fairness were analyzed. Results showed that participants had positive perceptions regarding the fairness of the serious game. This finding indicates that the serious game was able to provide equal opportunities for all the participants (Cook, 2016). Furthermore, results show that participants had positive perceptions regarding the opportunity to perform in the serious game. This finding indicates that participants were able to demonstrate their cooperation skills in the context of the serious game. These findings are in line with previous research that showed that gamified assessments and serious games receive favourable reactions from participants (Dubbel et al., 2015; Georgiou et al., 2019). Taken together, it can be concluded that participants had positive perceptions of the overall face validity of the serious game.

The last criterion that was used to analyze the serious game concerns the usability of the serious game. Results show that participants had positive perceptions regarding the usability, which indicates that the game was easy to use which is a relevant finding for the future development of the serious game.

Limitations

This research was conducted in the midst of the COVID-19 pandemic. The pandemic influenced this research in several ways. First, it meant that the serious game had limited possibilities to be developed. The research team had planned to integrate videos with real actors into the game to make the game more dynamic. This, however, was not possible to realize due to the restrictions on meeting people.

A second limitation that the pandemic put on the study regarded searching for and finding participants. There was no possibility to personally invite people to participate in the study. This is why convenience sampling was applied in this study, which could have influenced the results of the study. For example, it could have been better to specify requirements for the sample before conducting the study with regard to work experience. This also applies to the other-rated sample that did not solely consist of colleagues. Future studies could focus on the relationship between game cooperation scores and colleague-rated

cooperation scores, perhaps by asking groups of people that work at the same company to participate in both playing the game and evaluating each other.

A general limitation of the study is the fact that the serious game was still in an early stage of development. For example, the serious game was only available in English at the time the study was conducted. Furthermore, there are moments in the game where participants had to read multiple paragraphs of text. This could have influenced the results of the study as some participants are not as fluent in the English language as others. When taking into account that the serious game contained specific terminology that matched the storyline, it could be that the serious game was too complex for some participants.

A last limitation of the study is a lack of available measures for cooperation. For example, it was not possible to compare the types of validity that were analyzed in this study with other cooperation measures that already established these types of validity. As a result, this study was limited to the validity analysis based on participants' perceptions.

Future research

In terms of future research, it would be useful to expand our understanding of cooperation. The theoretical concept of cooperation that was used in this study is mainly based on the framework made by Kurz & Bartram (2002). Future studies can further explore the conceptualization of cooperation that can be used in order to effectively measure it. For example, there are components of cooperation that were not included in this study like listening and respect. With the help of technological developments, future studies might be able to integrate into a serious game.

Additionally, future studies should elaborate on the concept of cooperation by looking into the literature on teamwork, collaboration and collaborative problem-solving. These topics have in common that they all study the phenomenon of people working together. It could therefore be helpful to find similarities and differences between studies regarding relevant components of cooperation. By integrating the components of these various studies, a well-accepted concept of cooperation can be made and used in future research. This could improve the availability of literature on cooperation, which currently is lacking.

Another avenue for future research relates to working from home and the impact it has on components of cooperation. More specifically, it will be useful to know if working from home affects the opportunity to establish a trusting work environment where knowledge

sharing occurs. There are less face to face interactions when people work from home which might make it more difficult for people to get a reasonable impression of their colleagues. Moreover, there are fewer moments to engage in informal conversations. As trust and knowledge sharing have been found to positively influence cooperation, it might be interesting to explore the factors that play a role in the development of these components in a changing work environment.

Finally, future research should focus on developing serious games for behavioural competencies other than cooperation. Serious games have the potential to become great tools for assessment in terms of the measurement of the competency as well as the experience of playing the game. If using serious games for the measurement of behavioural competencies becomes the golden standard in the future, it will likely bring improvements on criteria like reliability and validity.

Conclusion

In this study, we aimed to develop a serious game to measure interpersonal cooperation skills. A theoretical framework of cooperation was made including multiple components that influence cooperation. During the study, participants filled out a self-rated questionnaire before playing the serious game. The results of this study showed that there was insufficient support for the construct and congruent validity of this specific serious game. However, additional results of the study indicate that participants had positive perceptions on the overall face validity and usability of the serious game. We hope that the current study will stimulate further investigation of the use of serious games for the measurement of competencies like cooperation skills within an assessment context.

References

- Alsharo, M., Gregg, D., & Ramirez, R. (2017). Virtual team effectiveness: The role of knowledge sharing and trust. *Information & Management*, 54(4), 479–490.
<https://doi.org/10.1016/j.im.2016.10.005>
- Armstrong, M. B., Ferrell, J. Z., Collmus, A. B., & Landers, R. N. (2016). Correcting Misconceptions About Gamification of Assessment: More Than SJTs and Badges. *Industrial and Organizational Psychology*, 9(3), 671–677.
<https://doi.org/10.1017/iop.2016.69>
- Barends, A. J., De Vries, R. E., & Van Vugt, M. (2021). Construct and Predictive Validity of an Assessment Game to Measure Honesty–Humility. *Assessment*, 107319112098561.
<https://doi.org/10.1177/1073191120985612>
- Barrick, M. R., Mount, M. K., & Judge, T. A. (2001). Personality and performance at the beginning of the new Millennium: What do we know and where do we go next? *International Journal of Selection and Assessment*, 9(1 & 2), 9–30.
<https://doi.org/10.1111/1468-2389.00160>
- Barrick, M. R., Parks, L., & Mount, M. K. (2005). Self-monitoring as a moderator of the relationships between personality traits and performance. *Personnel Psychology*, 58(3), 745–767.
<https://doi.org/10.1111/j.1744-6570.2005.00716.x>
- Bartram, D. (2005). The Great Eight Competencies: A Criterion-Centric Approach to Validation. *Journal of Applied Psychology*, 90(6), 1185–1203.
<https://doi.org/10.1037/0021-9010.90.6.1185>
- Bauer, T. N., Truxillo, D. M., Sanchez, R. J., Craig, J. M., Ferrara, P., & Campion, M. A. (2001). Applicant reactions to selection: Development of the selection procedural

justice scale (SPJS). *Personnel Psychology*, 54(2), 387–419.

<https://doi.org/10.1111/j.1744-6570.2001.tb00097.x>

Bell, S. T. (2007). Deep-level composition variables as predictors of team performance: A meta-analysis. *Journal of Applied Psychology*, 92(3), 595–615.

<https://doi.org/10.1037/0021-9010.92.3.595>

Brett, J. F., & Atwater, L. E. (2001). 360° feedback: Accuracy, reactions, and perceptions of usefulness. *Journal of Applied Psychology*, 86(5), 930–942.

<https://doi.org/10.1037/0021-9010.86.5.930>

Breuer, C., Hüffmeier, J., & Hertel, G. (2016). Does trust matter more in virtual teams? A meta-analysis of trust and team effectiveness considering virtuality and documentation as moderators. *Journal of Applied Psychology*, 101(8), 1151–1177.

<https://doi.org/10.1037/apl0000113>

Campion, M. A., Medsker, G. J., & Higgs, A. C. (1993). Relations between work group characteristics and effectiveness: Implications for designing effective work groups. *Personnel Psychology*, 46(4), 823–847.

<https://doi.org/10.1111/j.1744-6570.1993.tb01571.x>

Campion, M. A., Fink, A. A., Ruggeberg, B. J., Carr, L., Phillips, G. M., & Odman, R. B. (2011). Doing competencies well: Best practices in competency modeling. *Personnel Psychology*, 64(1), 225–262.

<https://doi.org/10.1111/j.1744-6570.2010.01207.x>

Christian, M. S., Edwards, B. D., & Bradley, J. C. (2010). Situational judgment tests: Constructs assessed and a meta-analysis of their criterion-related. *Personnel Psychology*, 63(1), 83–117.

<https://doi.org/10.1111/j.1744-6570.2009.01163.x>

Cook, M. (2016). *Personnel Selection* (6th Edition). Wiley.

- De Jong, B. A., Dirks, K. T., & Gillespie, N. (2016). Trust and team performance: A meta-analysis of main effects, moderators, and covariates. *Journal of Applied Psychology, 101*(8), 1134–1150. <https://doi.org/10.1037/apl0000110>
- Dubbelt, L., Oostrom, J. K., Hiemstra, A. M. F., & Modderman, J. P. L. (2014). Validation of a Digital Work Simulation to Assess Machiavellianism and Compliant Behavior. *Journal of Business Ethics, 130*(3), 619–637. <https://doi.org/10.1007/s10551-014-2249-x>
- Feitosa, J., & Salas, E. (2020). Today's virtual teams: Adapting lessons learned to the pandemic context. *Organizational Dynamics, 50*(1), 100777. <https://doi.org/10.1016/j.orgdyn.2020.100777>
- Finstad, K. (2010). The Usability Metric for User Experience. *Interacting with Computers, 22*(5), 323–327. <https://doi.org/10.1016/j.intcom.2010.04.004>
- Georgiou, K., Gouras, A., & Nikolaou, I. (2019). Gamification in employee selection: The development of a gamified assessment. *International Journal of Selection and Assessment, 27*(2), 91–103. <https://doi.org/10.1111/ijsa.12240>
- Gilliland, S. W. (1993). The Perceived Fairness of Selection Systems: An Organizational Justice Perspective. *The Academy of Management Review, 18*(4), 694. <https://doi.org/10.2307/258595>
- Gkorezis, P., Georgiou, K., Nikolaou, I., & Kyriazati, A. (2020). Gamified or traditional situational judgement test? A moderated mediation model of recommendation intentions via organizational attractiveness. *European Journal of Work and Organizational Psychology, 30*(2), 240–250. <https://doi.org/10.1080/1359432x.2020.1746827>

Harris, M. M., & Schaubroeck, J. (1988). A meta-analysis of self-supervisor, self-peer, and peer-supervisor ratings. *Personnel Psychology*, *41*(1), 43–62.

<https://doi.org/10.1111/j.1744-6570.1988.tb00631.x>

Johnson, J. W., & Ferstl, K. L. (1999). The effects of interrater and self-other agreement on performance improvement following upward feedback. *Personnel Psychology*, *52*(2), 271–303.

<https://doi.org/10.1111/j.1744-6570.1999.tb00162.x>

Kato, P. M. & Klerk, S. (2017). Serious Games for Assessment: Welcome to the Jungle. *Journal of Applied Testing Technology*, *18*(1), 1-6.

Kim, Y. H., Kwon, H., Lee, J., & Chiu, C. Y. (2016). Why Do People Overestimate or Underestimate Their Abilities? A Cross-Culturally Valid Model of Cognitive and Motivational Processes in Self-Assessment Biases. *Journal of Cross-Cultural Psychology*, *47*(9), 1201–1216.

<https://doi.org/10.1177/0022022116661243>

Klein, C., DeRouin, R. E., & Salas, E. (2008). Uncovering Workplace Interpersonal Skills: A Review, Framework, and Research Agenda. *International Review of Industrial and Organizational Psychology 2006*, 79–126. <https://doi.org/10.1002/9780470696378.ch3>

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

<https://doi.org/10.1007/bf02230391>

Koole, S. L., Jager, W., Van den Berg, A. E., Vlek, C. A. J., & Hofstee, W. K. B. (2001). On the Social Nature of Personality: Effects of Extraversion, Agreeableness, and Feedback about Collective Resource Use on Cooperation in a Resource Dilemma. *Personality and Social Psychology Bulletin*, *27*(3), 289–301.

<https://doi.org/10.1177/0146167201273003>

- Kurz, R., & Bartram, D. (2002). Competency and Individual Performance: Modelling the World of Work. *Organizational Effectiveness*, 227–255.
<https://doi.org/10.1002/9780470696736.ch10>
- Lievens, F., & Sackett, P. R. (2006). Video-based versus written situational judgment tests: A comparison in terms of predictive validity. *Journal of Applied Psychology*, 91(5), 1181–1188.
<https://doi.org/10.1037/0021-9010.91.5.1181>
- Lievens, F., Peeters, H., & Schollaert, E. (2008). Situational judgment tests: a review of recent research. *Personnel Review*, 37(4), 426–441.
<https://doi.org/10.1108/00483480810877598>
- Malhotra, A., Majchrzak, A., & Rosen, B. (2007). Leading Virtual Teams. *Academy of Management Perspectives*, 21(1), 60–70.
<https://doi.org/10.5465/amp.2007.24286164>
- McCrae, R. R., & Costa, P. T. (1997). Personality trait structure as a human universal. *American Psychologist*, 52(5), 509–516.
<https://doi.org/10.1037/0003-066x.52.5.509>
- Morrison-Smith, S., & Ruiz, J. (2020). Challenges and barriers in virtual teams: a literature review. *SN Applied Sciences*, 2(6).
<https://doi.org/10.1007/s42452-020-2801-5>
- Mount, M. K., Barrick, M. R., & Stewart, G. L. (1998). Five-Factor Model of personality and Performance in Jobs Involving Interpersonal Interactions. *Human Performance*, 11(2–3), 145–165.
<https://doi.org/10.1080/08959285.1998.9668029>

- Naul, E., & Liu, M. (2019). Why Story Matters: A Review of Narrative in Serious Games. *Journal of Educational Computing Research*, 58(3), 687–707.
<https://doi.org/10.1177/0735633119859904>
- Neubert, J. C., Mainert, J., Kretschmar, A., & Greiff, S. (2015). The Assessment of 21st Century Skills in Industrial and Organizational Psychology: Complex and Collaborative Problem Solving. *Industrial and Organizational Psychology*, 8(2), 238–268.
<https://doi.org/10.1017/iop.2015.14>
- Norcini, J., Anderson, B., Bollela, V., Burch, V., Costa, M. J., Duvivier, R., Galbraith, R., Hays, R., Kent, A., Perrott, V., & Roberts, T. (2011). Criteria for good assessment: Consensus statement and recommendations from the Ottawa 2010 Conference. *Medical Teacher*, 33(3), 206–214.
<https://doi.org/10.3109/0142159x.2011.551559>
- Nunamaker, J. F., Reinig, B. A., & Briggs, R. O. (2009). Principles for effective virtual teamwork. *Communications of the ACM*, 52(4), 113–117.
<https://doi.org/10.1145/1498765.1498797>
- Oostrom, J. K., Born, M. P., Serlie, A. W., & Van der Molen, H. T. (2011). A Multimedia Situational Test With a Constructed-Response Format. *Journal of Personnel Psychology*, 10(2), 78–88. <https://doi.org/10.1027/1866-5888/a000035>
- Pangil, F., & Chan, J. M. (2014). The mediating effect of knowledge sharing on the relationship between trust and virtual team effectiveness. *Journal of Knowledge Management*, 18(1), 92–106.
<https://doi.org/10.1108/jkm-09-2013-0341>

Porter, C. O. L. H., Hollenbeck, J. R., Ilgen, D. R., Ellis, A. P. J., West, B. J., & Moon, H. (2003). Backing up behaviors in teams: The role of personality and legitimacy of need.

Journal of Applied Psychology, 88(3), 391–403.

<https://doi.org/10.1037/0021-9010.88.3.391>

Rios, J. A., Ling, G., Pugh, R., Becker, D., & Bacall, A. (2020). Identifying Critical 21st-Century Skills for Workplace Success: A Content Analysis of Job Advertisements.

Educational Researcher, 49(2), 80–89.

<https://doi.org/10.3102/0013189x19890600>

Shute, V., & Rahimi, S. (2017). Review of computer-based assessment for learning in elementary and secondary education. *Journal of Computer Assisted Learning, 33*(1), 1–19.

<https://doi.org/10.1111/jcal.12172>

Simmons, C. H., Wehner, E. A., Tucker, S. S., & King, C. S. (1988). The Cooperative/Competitive Strategy Scale: A Measure of Motivation to Use Cooperative or Competitive Strategies for Success. *The Journal of Social Psychology, 128*(2), 199–205.

<https://doi.org/10.1080/00224545.1988.9711363>

Sliter, K. A. (2015). Assessing 21st Century Skills: Competency Modeling to the Rescue. *Industrial and Organizational Psychology, 8*(2), 284–289.

<https://doi.org/10.1017/iop.2015.35>

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

<https://doi.org/10.1111/j.1744-6570.1993.tb00867.x>

Staples, D. S., & Webster, J. (2008). Exploring the effects of trust, task interdependence and virtualness on knowledge sharing in teams. *Information Systems Journal*, 18(6), 617–640.

<https://doi.org/10.1111/j.1365-2575.2007.00244.x>

Tauer, J. M., & Harackiewicz, J. M. (2004). The Effects of Cooperation and Competition on Intrinsic Motivation and Performance. *Journal of Personality and Social Psychology*, 86(6), 849–861.

<https://doi.org/10.1037/0022-3514.86.6.849>

Westera, W. (2019). Why and how serious games can become far more effective: Accommodating productive learning experiences, learner motivation and the monitoring of learning gains. *Education Technology & Society*, 22(1), 59–69.

https://drive.google.com/file/d/1yTQBRopaRgNo99__BVuGg1aoh_Utx3Ry/view

Whetzel, D., Sullivan, T., & McCloy, R. (2020). Situational Judgment Tests: An Overview of Development Practices and Psychometric Characteristics. *Personnel Assessment and Decisions*, 6(1).

<https://doi.org/10.25035/pad.2020.01.001>

Woods, S. A., Ahmed, S., Nikolaou, I., Costa, A. C., & Anderson, N. R. (2019). Personnel selection in the digital age: a review of validity and applicant reactions, and future research challenges. *European Journal of Work and Organizational Psychology*, 29(1), 64–77.

<https://doi.org/10.1080/1359432x.2019.1681401>

Wouters, P., Oostendorp, H. V., Boonekamp, R., & Spek, E. V. D. (2011). The role of Game Discourse Analysis and curiosity in creating engaging and effective serious games by implementing a back story and foreshadowing. *Interacting with Computers*, 23(4), 329–336.

<https://doi.org/10.1016/j.intcom.2011.05.001>

Appendix A

Uitnodiging onderzoek

Beste lezer,

Momenteel ben ik druk bezig met mijn afstudeeronderzoek voor de studie Positieve Organisatiepsychologie aan de Erasmus Universiteit Rotterdam. Het onderzoek gaat over samenwerkingsvaardigheden en in hoeverre deze gemeten kunnen worden met een simulatie.

Deze simulatie, de *Team Trust Game* (TTG), is een serious game die momenteel nog in ontwikkeling is. In het spel gaat het om een samenwerkingsopdracht. De verhaallijn, bestaande uit diverse scènes, wordt onderbroken zodat u keuzes kunt maken, waarna het spel zich verder ontwikkelt. De manier van spelen en de antwoorden in het spel geven een indicatie in hoeverre u samen wil werken (intentie) en uw daadwerkelijke gedrag.

Het onderzoek bestaat uit de volgende onderdelen:

1. Een vragenlijst waarmee we een indicatie van uw samenwerkingsbereidheid kunnen krijgen.
2. De Team Trust Simulatie.
3. Een korte evaluatie vragenlijst over de simulatie.
4. Een vragenlijst die u naar een aantal van uw collega's of anderen waar u mee samenwerkt kunt doorsturen.

In totaal neemt het onderzoek maximaal een uur in beslag. Daarbij is het ook mogelijk om te pauzeren, stoppen of herstarten. Het is belangrijk om te weten dat alle antwoorden anoniem en vertrouwelijk verwerkt en behandeld zullen worden.

Ik hoop ontzettend dat u mij kunt helpen met dit onderzoek waarmee ik kan afstuderen. Als u verdere vragen of opmerkingen heeft dat kunt u contact met mij opnemen middels 432625kk@student.eur.nl.

Met vriendelijke groet,

Karch Kooijman

Appendix B

Research invitation

Dear reader,

At the moment I am busy with my graduation research for Positive Organizational Psychology at Erasmus University Rotterdam. The research is about collaboration skills and to what extent these can be measured with a simulation.

This simulation, the *Team Trust Game* (TTG), is a serious game that is currently under development. The game involves a collaboration assignment. The storyline, consisting of various scenes, is interrupted so that you can make choices, after which the game continues to develop. The way of playing and the answers in the game give an indication of how far you want to work together (intention) and your actual behavior.

The research consists of three parts:

1. A questionnaire with which we can get an indication of your willingness to cooperate.
2. The Team Trust Simulation.
3. A questionnaire that you can send to some of your colleagues or others you work with.

The research takes a maximum of one hour to complete. It is also possible to pause, stop or restart. It is important to know that all responses will be processed and treated anonymously and confidentially.

I very much hope that you can help me with this research with which I can graduate. If you have any further questions or comments, you can contact me by phone at 0653984253 or by email at 432625kk@student.eur.nl.

Kind regards,

Karch Kooijman

Appendix C - Description of story, character, and game design in the *Team Trust Game*

Story. The game follows a scenario where the player gets assigned as the new Project Manager for the *Vaccine Defense Consortium*. This consortium consists of a team that developed a working vaccine for an unknown virus. However, the team experiences difficulties with the production of the vaccine. This is because some team members have different priorities when it comes to the quality or quantity of the vaccine. What follows is that team members start to distrust each other, which has a negative effect on the way they cooperate. The player will follow the production process of the vaccine and interact with either the team as whole, or with members separately. Through these interactions, the player gets to know more about the personality, motivations or even personal problems of each team member. Considering these factors, the player gets the opportunity to restore trust within the team so that they can find a solution to the production problem together.



Figure 1. Introduction screen Team Trust Game.

Character profiles. Each character within the game has been allotted with a specific personality profile. The two personality traits that are linked to effective cooperation are *extraversion* and *agreeableness* (Bell, 2007). These personality traits follow from the five-factor model of personality as described by Mcree & Costa (1997). Extraverted individuals tend to be more social, talkative, assertive and active (Barrick et al., 2001). Agreeable individuals tend to be more cooperative, flexible, tolerant and forgiving (Barrick et al., 2001). Each character that is encountered in the game is designed to show a certain amount of extraversion or agreeableness. Based on the Big Five framework the characters will also show

signs of *conscientiousness*, *emotional stability* and *openness to experience* (McCrea & Costa, 1997). The combination of personality traits in the characters lead to a more diverse and realistic workplace simulation.

The first main character is Sigrid Virtanen and she is introduced as the Intelligence Officer that reports back to the higher ups. The character has a background in the Finnish army and she expects the player to deliver results. Moreover, she wants the player to share all the relevant information regarding the *Vaccine Defense Consortium*. Based on the dialogue, graphic design and role of the character, certain personality traits of Sigrid can be detected. The character is designed to show low levels of *agreeableness* and can be seen as intolerant and merciless (Barrick et al., 2001). The lack of agreeableness in this character puts pressure on the player which makes her difficult to work with. Other notable personality traits in this character are a high degree of *conscientiousness* and a low degree of *openness to experience*. Previous research has shown that some personality traits can be detected from visual features of the face (Kramer & Ward, 2010). In this case, the character is graphically designed with a stern look and a serious face. Figure 2 shows the graphic design of Sigrid Virtanen. While playing the game, the player must decide if this character can be trusted and if they should share all relevant information regarding the team with her.

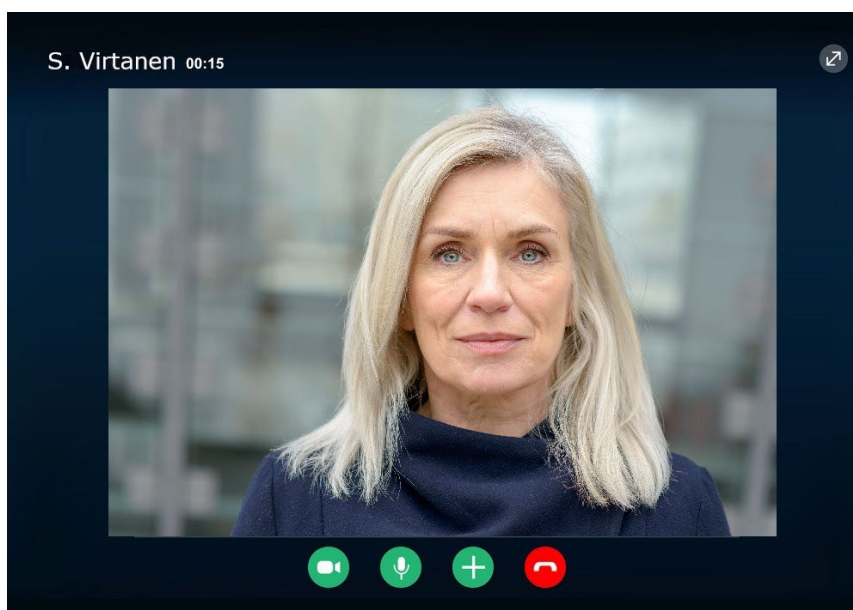


Figure 2. In game character design of Sigrid Virtanen

The second character that is introduced is Gemma Quevedo and she is one of the team members of the *Vaccine Defense Consortium*. This character is the Head of Business Development and is responsible for the sales department for the vaccine. The character is designed as a thriving business woman who is eager to make steps in her career. For Gemma, a successful production process of the vaccine means she will be able to grow her career. However, for this to happen she also needs to find common ground with her other team members. Based on the dialogue, graphic design and role of the character, certain personality traits of Gemma can be detected. High levels of extraversion can be found in this character, as Gemma is sociable, talkative, assertive and active (Barrick et al., 2001). Low levels of *agreeableness* can be found in the character. The character is difficult to cooperate with because she tends to prioritize her own ideas over the ideas of others (Barrick et al, 2001). The character is self-confident, curious and focused on achievement. Figure 3 shows the graphic design of Gemma Quevedo. During the game, the player must convince Gemma to take the ideas of other team members into consideration, so that they can succeed as a team.



Figure 3. In game character design of Gemma Quevedo

The next member of the *Vaccine Defense Consortium* is dr. Bernhard Breier, the Head of Engineering. This character is designed as an experienced and well respected doctor who has helped create and produce the vaccine. However, due to his conservative and cynical

demeanour, he has trouble working together with the other team members. Based on the dialogue, graphic design and role of the character, certain personality traits of Bernhard can be detected. The character shows a low degree of agreeableness as the cynical attitude of this character leads to being less flexible and forgiving to the ideas and actions of other characters (Barrick et al., 2001). Furthermore, the character tends to act reserved when interacting with the other team members, which indicates a low degree of extraversion. The personality traits of Bernhard are also found through his closed facial expressions and posture. Figure 4 shows the graphic design of Bernhard Breier. The serious game investigates if the player has the skills to make Bernhard and the other characters willing to cooperate with each other.



Figure 4. In game character design of Dr. Bernhard Breier.

The fourth character is Sophia Bisset, the Political Advisor in the *Vaccine Defense Consortium*. Sophia's role in the team is to make sure that international protocols regarding the production of the vaccine are being followed. The character shows a higher degree of agreeableness. Contrary to the previous character Bernhard, Sophia is willing to cooperate with the other members as long as they adhere to the rules and regulations. Figure 5 shows Sophia in the early stage of the game with a serious and professional attitude. These characteristics make the character look agreeable and conscientious (Kramer & Ward, 2010). Perhaps the most notable personality trait in this character is a lower degree of *emotional stability*. Later in the

game Sophia is found to be dealing with a personal situation that causes her to be stressed, worried and sensitive. These characteristics are fitting for the lower degree of emotional stability in the character (Kramer & Ward, 2010). The game asks the player to be empathetic towards Sophia, and to encourage her so she can focus on her tasks at hand.



Figure 5. In game character design of Sophia Bisset.

The fifth and final character in the *Vaccine Defense Consortium* is Jean-Marc van de Heijden, a professor in microbiology who has been responsible for creating the vaccine. Jean-Marc is characterized by his enthusiastic, assertive and outgoing attitude, indicating a high degree of extraversion (Barrick et al., 2001). The second personality trait in this character is a moderate to degree of *agreeableness*. This is because the character shows signs of being cooperative and flexible (Barrick et al., 2001). Based on these two personality traits, it would be expected that this character is easy to cooperate with (Bell, 2007). The personality traits of the character are also transferred in his graphic design, which can be found in Figure 6. This character is designed with open and expressive facial features that match his personality traits (Kramer & Ward, 2010). During the game, the player must figure out how to handle the outgoing nature of Jean-Marc, especially when he is found to be distracted with other projects. The game investigates if the player coaches Jean-Marc to remind him that his behaviours can have negative consequences for the team.



Figure 6. In game character design of Jean-Marc van de Heijden

Game environment design. The game environment of the Team Trust game was designed to complement the scenarios in the story and interactions with the characters. Background pictures of buildings, meeting rooms, offices and laboratories were added to scenes to make the game play realistic and engaging. For example, the environment puts emphasis on the diplomacy in the story by showing a conference building followed by a closed meeting room. Figure 7 shows an example of this closed meeting room.



Figure 7. Game environment design of the meeting room

The implementation of virtual communication devices is another example of complementary game design. During the game, the player gets to interact both face-to-face and virtually with the other characters. Pictures of computers, templates of Skype screens and a smartphone device were edited in the game to establish the virtual side of cooperation. Furthermore, the player receives notifications and separate messages from characters on this smartphone. These messages are sent in the form of texts, e-mails or even voicemail. The player is able to manually open these messages that contain additional and important information to the story.

Lastly, video montages were made and edited in the game with the intention to show a sequence of events related to the story. For example, a video montage can be shown right before the player gets to interact with a character. This video contains multiple pictures of the character with different attitudes and emotions, as shown in Figure 8. This informs the player that something significant could have happened to the character. This information can then be brought up during the interaction with the character. The video montages can be shown after interactions as well, to provide a follow-up on the interaction or to introduce a new scene.



Figure 8. Example of a montage moment as part of game environment design