Investigating the Effect of Team-Size on the Nature of Communication

Author:
Jeroen Brands
S1985663

Examination Committee:
Henny Leemkuil
Hannie Gijlers

FACULTY OF BEHAVIOURAL, MANAGEMENT AND SOCIAL SCIENCES
UNIVERSITY OF TWENTE

Enschede, 29th August 2019

Table of Contents

Acknowledgement

For my parents, who enabled me to study for as long as I needed.

Abstract

In a time where information is readily available, the focus of learning has shifted from knowledge to skills capable of implementing this knowledge. These skills are referred to as 21st century skills. Four competencies, the four C's, foster the practice of these skills. Critical thinking and creativity, the personal skills, and communication and collaboration, the interpersonal skills. Hence, teamwork cannot be circumvented when researching 21st century skills. This study focused on the effect of team size on the nature of communication when collaborating on solving a problem that requires critical thinking and creativity. This problem was developed through a design-based board game as research has shown the effectiveness of games for teaching 21st century skills. Sixty students from the Behavioural, Management and Social Sciences faculty at the University of Twente were split into ten groups of two and ten groups of four. While letting these teams solve the developed problem, their communication was recorded. It was found that team size had no significant effect on performance. However, bigger teams did communicate more than smaller teams. Finally, some elements of communication, which were classified as 'higher order communication', were found to positively influence communication and collaboration. These skills can be improved by implementing measures that stimulate and regulate higher order communication, resulting in improved overall performance.

Keywords: 21st century skills, communication, group-size, performance, higher order communication

Introduction

The digital age, a technological era we only recently entered, is classified as a new historical period, just like the stone, bronze and iron age. A radical shift from the industrial revolutions, associated with the invention of the transistor: this enabled the creation of informational processing units (Castells, 1996). Shifting from a material-based to an information-based society requires a new way of working and thus a new way of thinking. This is what we call a paradigm shift (Kuhn, 1962). The digital age is characterized by digitalizing information. This entails that most information can easily be accessed online by anyone who has access to the internet. Hence, the need to remember facts, which used to be the primary way of learning for many years, has become less important than possessing the skills to implement this knowledge (Voogt & Pelgrum, 2005). According to Binkley et al. (2012) these necessary skills have been described as 21st century skills. This is a construct well-accepted within the international education world (Voogt & Roblin, 2012). 21st century skills refer to higher level cognitive, interpersonal, and intrapersonal skills. These skills are increasingly relevant in public education aimed at a global economy (Voogt & Roblin, 2012). Withal, those who study 21st century skills do not argue that 21st century skills should replace teaching discipline-specific content. Instead, it is an addition, adhering the necessity of applying discipline-specific knowledge in practise (Hayse, 2018). Moreover, 21st century skills are beneficial to students long after they graduate both in their private life as in their work field (Kaufman, 2013). Nevertheless, even though scholars developed frameworks and definitions of these skills, no consensus has emerged yet (Binkley et al., 2012). Defining 21st-century skills is difficult, partly due to these skills being vague (Kickmeier & Albert, 2012).

The World Economic Forum (WEF), in line with the Partnership for 21st Century Learning (P21), postulates four competencies (which P21 refers to as Learning and Innovation Skills) for the practice of 21st century skills[1][2]. These competencies are:

- Critical Thinking and Problem Solving
- Creativity
- Communication
- Collaboration

They can be split into two categories: personal skills (critical thinking and creativity) and interpersonal skills (communication and collaboration). These four competencies are thus the backbone of the teaching of the future. Therefore, it seems appropriate to study these competencies

[1] WEF: http://widgets.weforum.org/nve-2015/appendices.html retrieved 26/07/2019
[2] P21: https://www.imls.gov/assets/1/AssetManager/Bishop%20Pre-Con%202.pdf retrieved 26/07/2019

in depth. Academics have found that to teach 21st century skills, other methods than traditional teacher-centered education – where the students focus on listening to the teacher and work alone on assignments – are more useful (Romero, Usart, & Ott, 2015). One of the methods of teaching 21st century skills that seems to work well is teaching through serious games (van Laar, van Deursen, van Dijk, & de Haan, 2017; Sourmelis, Ioannou, & Zaphiris, 2017; Qian & Clark, 2016). Hence, it makes sense that there is an increase of interest within the communicative and educational field to study the relation between game-based learning and communication (Wouters & van Oostendorp, 2013).

Hayse (2018) found that three out of four 21st century learning competencies (communication, collaboration and critical thinking) are a necessity to play board games: board games enable learners to practice these three competencies. Insomuch as Hayse studied *if* board games can be used as a tool, he did not study the relation between these competencies and games, nor which type of game work best. Qian & Clark (2016) found that design-based games – games that are designed specifically with the goal of learning or increasing performance, most often in collaboration – tend to work better than entertainment games that have created for entertainment and commercial purposes only.

In order to prepare students for the 21st century and to make them adaptive to the future, it is vital for educators and educational scientists to understand what the 21st century skills are, how they work and how they should be taught. To do so a better understanding of these skills in themselves is required. Only then can scientists begin to fathom the complex relation between these skills. Hence, this study will focus on investigating what the differences in nature of communication are between different group sizes. This will be done using a design-based board game.

Theoretical Framework

Now that it is clear which skills are important for the future, they will need to be conceptualized. By analysing each of the skills, multiple dimensions appear. Therefore, it seems wise to start with outlining a few important definitions. These are required in order to understand the model that follows at the end of this framework. Moreover, some of the terms that will be discussed are used in common language with multiple connotations that might be ambiguous, and hence require clarification. Others are a variation of existing words that are used to imply a slightly different meaning.

Definitions

Critical Thinking, Problem Solving and Creativity.

The WEF describes critical thinking or problem solving as the ability to identify, analyse and evaluate situations, ideas and information to formulate responses and solution. Furthermore, they describe creativity as the ability to imagine and devise new, innovative ways of addressing problems, answering questions or expressing meaning through the application, synthesis or repurposing of knowledge. Hence, critical thinking, problem solving, and creativity are about addressing problems.

Wechsler et al. (2018) also concluded that problem solving, critical thinking and creativity are three correlated, interdependent variables. However, it should be noted that these three ideas are so interrelated that separating them as variables does not seem wise. Thinking is intrinsically creative and critical (Lipman, 2003). Hence, to solve problems, especially complex ones, one must come up with a way of solving, where thinking of a way of solving problems is intrinsically critical and creative. If absolutely no creativity is required, then there would be a standardised solution to every problem. Then problem solving would be about finding this standardised solution instead of using critical thinking to come up with a creative way to solve it. Hence, due to the inseparability of these three variables, and due to the scope of this research, this research will assume these three components as one and will thus not be studied in depth.

Team size and performance. Research indicates that a team's adaptability and creativity have the potential to be greater when increasing the amount of team members (Gladstein, 1984). What this means is that increasing the amount of team members increases the overall problem-solving capacity of a team: they can offer more complex and innovative solutions to problems. However, they often carry more and bigger responsibilities, their failure can have a bigger impact, and the breakdown of a larger team is more likely to occur than of a smaller team (Alderfer, 1977). This implies that having more team members should increase the potential of a team and makes

them more capable of solving more complex problems, but also increases the chance a team breaks down. Two factors why are poor planning and communication (Salas, Sims, & Burke, 2005). What ensures the success of a team is what they call teamwork. In order to define the pillars of teamwork Salas et al. (2005) did a literature review and found that shared mental models and communication are at the core of necessary requirements for teamwork and hence team performance. Therefore, these two elements require to be included in any study regarding teams. However, we should keep in mind that the amount of effort it takes to create shared models increases when the amount of mental model sharers increases. Hence, we should not disregard team size as an important variable in a team's success.

According to a study done by Lowry, Roberts, Romano, Jr., & Cheney (2006) smaller groups – with three members as opposed to six – establish and maintain higher levels of communication quality. However, this research was conducted in a computer-based environment. Moreover, members of smaller teams participated more actively on their team, were more committed to their team, were more aware of the goals of the team, had greater awareness of other team members, and were in teams with higher levels of rapport (Bradner, Mark, & Hertel, 2005). However, this research was done with teams of four versus teams of eighteen. Furthermore, when teams become larger, each team member has less direct interaction; therefore, and as consequence of this reduced interaction, students on larger teams would be less prone to mastering the learning objectives associated with an exercise and hence will perform at a lower level (Cochrane & Eversole, 2012). Therefore, smaller sized teams, due to higher levels of interaction and lesser degrees of social loafing/disengagement would perform at higher levels than larger size teams. This is in line with a recent research done by Swanson et al. (2019) who found that smaller groups, with five or less students, were more effective than larger groups. Finally, an analysis done Cochrane & Eversole (2012) failed to support a hypothesis that a correlation exists between performance and team sizes of two, three or four students. However, they mention that this study was conducted in laboratories, and these results may thus not be generalizable.

All in all, research postulates that smaller groups perform better than bigger groups. However, what is defined as a smaller group differs per study. The smallest groups that were measured in existing literature on the effect of team size on performance were groups of four. This leaves us with the question of what the difference in performance is between groups of four and even smaller groups such as dyads.

Collaboration.

The WEF postulates collaboration as the ability to work in a team towards a common goal. Collaboration is about participation: if no one is participating in collaboration, there is no

collaboration. Understanding the world is only possible through interpretation (Nietzsche, trans. 1998). Each perspective gives a slightly different interpretation of the world, which connects different ideas and experiences. Therefore, increasing the total amount of possible perspectives equals increases the possible ways of looking at and understanding things. This means that to have the best possible dialogue, all participants should share their view. Even when disagreeing, hearing other perspectives might help one's understanding of what should be done. Therefore, the best possible scenario would entail equal participation among all participants.

Quantity and Heterogeneity of Participation. Quantity of participation (QOP) is to what extend learners contribute to discourse (Weinberger & Fischer, 2006). For discourse to exist, learners must contribute something. The amount of contribution can be expressed in quantity of participation. Quantity of participation can be expressed in different ways, but it comes down to counting the total amount of speech acts, whether it be the overall number of words and sentences or something more specific.

Heterogeneity of participation (HOP) is how much each individual contributes to the discourse compared to the rest of those in the discourse. Where quantity of participation is absolute, heterogeneity of participation is relative. If one person is holding a monologue and the rest of the group only replies affirmative occasionally the heterogeneity is high. If each person in a group contributes equally to the dialogue, the heterogeneity is low. As mentioned, sharing multiple perspectives is beneficial to create more insight and thus more options. As performance is based on the ability to solve problems, and sharing perspectives increases the ability to so, one would expect a correlation between group heterogeneity and group performance. If participation is not equal, this indicates a mismatch of power: one person follows the other because they are insecure, less dialogically skilled, have less prior knowledge, or there is an external motive in the relation between the participants. These unequal power relations can result in multiple problems in terms of normativity, truthfulness and dedication (Habermas, trans. 1981).

Quick consensus seeking. The problem to be considered in this study is a part of truthfulness, which will be referred to as quick consensus seeking: agreeing with arguments put forward in order to avoid confrontation (Asterhan & Schwarz, 2016). If the participants are afraid to share their opinion and simply agree with their team just to avoid confrontation, the outcome of the dialogue is sub-optimal at best (Leitão, 2000; Keefer, Zeitz, & Resnick, 2000). In general, quick consensus building may be detrimental to individual knowledge acquisition, when learners disregard other forms of consensus building in favour of quick consensus building (Weinberger & Fischer, 2006). Hence, if participants focus on avoiding confrontation instead of focusing on exchanging ideas

and mental models, the outcome of their teamwork is expected to be sub-optimal. Therefore, in order to achieve the best possible result, it is expected to be necessary that learners are actively and equally engaged in the dialogue, with the intend of sharing what they think is – or could be – the solution.

Communication.

Communication is defined by the WEF as the ability to listen to, understand, convey and contextualize information. To understand these pillars of communication, two main factors can be distinguished. First, something is communicated with the goal of realizing some effect: this will be referred to as the argumentive dimension. Second, the communication should be truthful and correct. This will be referred to as the epistemic dimension.

Argumentive, a term first coined by Kuhn & Wadiya (2003), is used to describe a specific type of argumentation. To prevent misunderstanding and to achieve a higher level of understandability, distinctions in language are useful and necessary (Austin, 1975). Since argumentation has a double connotation (namely both positive – with the goal to understanding something more clearly and to share perspectives – and negative – with the goal to achieve 'winning' a dialogue) a distinction is suitable. Kuhn & Wadiya (2003) describe argumentive as skilful argumentation that is aimed at elaborating arguments. Moreover, the goal of argumentive dialogue is never to beat your opponent: it is always to ensure fruitful dialogue that stimulates understanding and an open exchange of ideas. Argumentive interaction can be recognized by "participants show[ing] a willingness to actively revise or change their own views in response to persuasive arguments" (Keefer, Zeitz, & Resnick, 2000). However, that does not mean that participants cannot argue against each other's ideas. By facing critique, learners are pushed to test multiple perspectives on the subject matter or to find more and better arguments for their positions (Chan, Burtis, & Bereiter, 1997). Instead, learners should integrate each other's perspective, synthesize their ideas and arguments, and jointly try to make sense of a task (Nastasi & Clements, 1992). When learners build consensus in a conflict-oriented manner, where they are not afraid to disagree, they share their arguments and engage on a high level of reasoning. They are then required to pinpoint certain aspects of their peers' arguments, increasing the need to either modify them or to present feasible alternatives, resulting in a high level of communication, and hence performance (Weinberger & Fischer, 2006).

The epistemic dimension refers to the truthfulness of what is said. Epistemology, coming from the Greek words for 'knowledge' and 'to study', concerns itself the nature, justification and rationality of belief of knowledge. In other words, when can we say something is true? Hence, the epistemic dimension of knowledge construction is not only about the quantity, but mostly about the quality of contributions (Weinberger & Fischer, 2006). The epistemic dimension refers to how the

learners work on the knowledge construction task they have been given in order to ensure correctness (Fischer, Bruhn, Gräsel, & Mandl, 2002). This dimension can be considered a necessary check to ensure that learners do not inhibit or express untruthful knowledge. Moreover, depending on the task, this dimension might foster knowledge as well. Knowing when something is correct depends on the framework; to think from that framework, in itself, the framework must also be correct, otherwise the thoughts coming from that framework in relation to the problem are incorrect by default.

The framework of everything that is necessary to understand a problem will be referred to as the problem space. Hence, the construction of the problem space is an epistemic activity that is required to understand a problem (Weinberger & Fischer, 2006). In this study problem space refers to all rules and mechanics, both within and outside, the game. When engaging in the discourse, going beyond the concrete level of the problem space may foster acquisition of knowledge on complex problems (Hogan, Nastasi, & Pressley, 2000). In other words, to understand and learn from complex problems, going beyond the concrete limitations of the question presented is necessary. Therefore, epistemic activities related to the problem space are considered to be one of the three main domains.

However, while some epistemic activities are vital to comprehend and solve complex problems, others increase the difficulty, or even make solving the problem at hand impossible. If one misunderstands the problem space, solving the problem within will prove difficult. Therefore, a distinction seems suitable. Positive epistemic speech acts (PESA) refers to all epistemic speech acts that benefit those operating within the problem space, whereas negative epistemic speech acts (NESA) refers to the speech acts that impede the construction of a proper problem space.

Models

Now that it is clear what constitutes the skills that will be studied, a model would greatly enhance the understanding of how these skills are related to each other. However, to the best of the author's knowledge, a model for 21st century skills does not exist. However, each of the skills separately have been studied for decades. Hence, some models can be found, but will need to be adjusted and updated to fit the current study. Moreover, many of these skills have been treated as concepts, but have not been studied as being a skill. Therefore, this study will propose a new model based on existing research.

Weinberger & Fischer (2006) model.

The model postulated by Weinberger & Fischer (2006) can be found in full detail in Appendix A. The authors developed this framework to analyse the process dimensions of knowledge construction in a computer-supported collaborative learning (CSCL). However, the present study will look at the nature of offline, face-to-face communication and the relation with performance. Hence, it needs to be adjusted in several ways. First, it will need to be adjusted to a non-computer environment, as it has been developed for text-based online communication. Therefore, the model will be adjusted to fit face-to-face verbal communication. Second, this model convolutes some aspects of communication. The present study will not focus on the micro-dimensions. Thirdly, some dimensions and subdimensions do not apply to the present study as they presuppose theoretical knowledge required for knowledge construction which has been left out intentionally. This is more relevant for using communication for studying science for example. Finally, as the goal of this study is to stay as objective as possible, all variables that require the researcher's personal judgment have been adjusted where possible, or removed. All that remains will be synthesized with an adjustment of the following model.

Van der Meij, Albers & Leemkuil (2011) model.

To examine whether people benefit more from playing a commercial game in dyads rather than alone, Van der Meij, Albers & Leemkuil (2011) classify communication using a distinction between higher and lower order communication that can be found in the table below. However, in the present study the focus lies on the effect of group size on the nature of communication, whereas Ver der Meij et al. (2011) focus on the differences in communication between individuals and dyads. Hence, this model is not fully in line with the proposed goal, and needs to be adjusted: first, it will be adjusted to fit measuring communication in depth; second, to account for differences in group sizes.

Van der Meij et al. (2011) describe lower order communication as the speech acts the first and second level verbalisations. These concern: describing, asking questions and proposing. Higher order communication, the third and fourth-level verbalisations, include: relating to prior knowledge, explaining/arguing and predicting and deliberating. Below the model is displayed, where a double line is drawn according to the original authors's distinction between higher and lower communication.

Activity	*Definition*	*Example*
Fourth-level verbalisations		
Relating to prior knowledge	Indicating having seen or done something before	Just like in a real city
Explaining/arguing	Accounting for and reasoning out ideas	Building a factory will bring us more money but it's not good for the environment You know what I would have done. Put these energy things on the other side of the river. Then people will not be bothered
Third-level verbalisations		
Predicting and deliberating	Forecasting future situations and contemplating effects.	If we do this, our environment is likely to become better I don't think that they want to live there, if you build industry
Second-level verbalisations		
Proposing an action	Suggesting what actions to take, or responding to such a suggestion.	We are going to buy light industry We could build windmills at sea Eco-roofs are also possible. Shall we do that?
First-level verbalisations		
Describing	Telling what happened in an event	We have no money Look, this has increased
Asking about the interface	Inquiring about a feature of the interface	What is this, forest?

Figure 1. Communication Distinction

Firstly, as mentioned, arguing can be both positive and negative. Classifying all arguing communication as higher order is too hasty as negative arguing is possibly not beneficial. The same goes for predicting: if participants do not share the reason behind their prediction, it is difficult to understand for their fellow team members why this prediction has been made and whether it will become true. Moreover, Van der Meij et al. (2011) classify inquiries about the interface as the lowest level of verbalisations. However, these are necessary to ensure shared mental models, and might thus be more useful than anticipated beforehand.

Therefore, a new distinction between higher (HOC) and lower order communication (LOC) has been developed for this study. Lower order communication refers to all speech acts that do not include justifications; higher order communication differs in that it does require a justification. As mentioned, in order for teams to succeed they need to first make sure that they share the same mental model. Justifying actions based on these mental models would thus reinforce the mutual understanding of these models. That way the other participants would understand why certain communications have been made and helps them align their thoughts about what steps to take, saving both brain computing power and time.

Synthesized New Model.

Based on the previously regarded models, a new model has been synthesized. This new model has been developed to test the effect of team size on the 21st century skills, primarily communication. Hence, it separates the four C's, and connects them to the models covered. The model looks as follows:

1. Critical thinking
    a. Performance Score
2. Collaboration
    a. Participation dimension
        i. Quantity of participation
        ii. Heterogeneity of participation
3. Communication
    a. Epistemic dimension
        i. Positive Epistemic Speech Acts
            1. Questions about the problem space
            2. Correct construction of problem space
        ii. Negative Epistemic Speech Acts
            1. Incorrect construction of problem space
            2. Non-epistemic activities
    b. Argumentive dimension
        i. Lower Order Communication
            1. Questions
            2. Statements
        ii. Higher Order Communication
            1. Arguments
            2. Counterarguments
            3. Integrations

It has general usage: it can be applied in many different scenarios. However, there are a couple of assumptions that need to be fulfilled for this model to be applicable. First, it is a collaborative model that is used for testing team communication. Second, the teams need to have a task that requires critical thinking. Third, as mentioned, it has been developed for offline, face-to-face communication only. Finally, it requires one to compare different team sizes to provide useful data.

How it Comes Together

Figure 2. How it Comes Together

The figure above illustrates how the model should be understood. The collaboration skill is expressed in participation. Participation is split into two measurable variables: quantity of participation and heterogeneity of participation. The communication skill is split into two: epistemic and argumentive. The epistemic dimension is split into the measurable variables of positive and negative epistemic speech acts. The argumentive dimension is split into the measurable variables of lower and higher order communication. The effect of each of these variables on score has been studied. As mentioned, since critical thinking is required to get a good score, the score is considered an indication of how well teams critically thought.

Research Questions

The main research question of this study is: *What is the effect of team-size on the nature of communication?* In order to adequately answer this question, numerous sub-questions have been formulated below. These questions can be split into two categories: differences between smaller and bigger groups, and the effect of the measured variables on performance. Just looking at the differences between bigger and smaller groups, without any regard of the underlying connections

between the (sub)variables was deemed unwise. Hence, the effect of these variables on each other was also measured, as to ensure the validity and reliability of this study.

The differences between smaller and bigger groups:

- Do bigger groups have a game score?
- Do bigger groups have more quantity of participation?
- Do bigger groups have more heterogeneity?
- Do bigger groups have more lower order communication?
- Do bigger groups have more higher order communication?
- Do bigger groups have more positive epistemic speech acts?
- Do bigger groups have more negative epistemic speech acts?

The effect of the measured variables on performance:

- Do groups with more quantity of participation have a higher game score?
- Do groups with more heterogeneity of participation have a higher game score?
- Do groups with more lower order statements have a higher game score?
- Do groups with more higher order statements have a higher game score?
- Do groups with more positive epistemic speech acts have a higher game score?
- Do groups with more negative epistemic speech acts have a higher game score?

Hypotheses.

It is expected that bigger groups will get a higher score because they have more perspectives. Having more people will make communication more difficult, hence requiring more communication in general to bring across the same potential message to more people. Hence, it is expected that bigger teams will have more quantity of participation. Moreover, bigger groups make social loafing and freeriding easier (Cochrane & Eversole, 2012). Likewise, quick consensus seeking is also easier in bigger teams as the pressure to stand up and deny claims is bigger, making it inherently more difficult to do so. Hence, it is expected that the level of heterogeneity in bigger groups is higher. Bigger groups will also need to communicate more and on a higher level to bring the same message across to more people. Explaining something to one person is easier than explaining it to a hundred different ones. Hence, both the quality and quantity of communication should be high for everyone to understand. Therefore, it is expected that bigger groups will have more higher order

communication and positive epistemic speech acts, and less lower order communication and negative epistemic speech acts.

More communication entails that more ideas will be shared (Nietzsche, trans. 1998). This leads to more perspectives being shared between team members. Since it is expected that more perspectives would lead to a better and more in-depth understanding, this should lead to a higher score as well. As mentioned, it is expected that equal participation should lead to more perspectives being shared. By increasing the number of perspectives, it gives all team members more options to consider. When picking the best option, more options should lead to the possibility of having better options as well, which in turn, when picked, could potentially lead to a higher performance. Thus, it is expected that a higher level of heterogeneity results in a lower game score.

This whole paper is structured around the assumption that quality communication should lead to quality performance. Hence, it is expected that higher order statements and positive epistemic speech acts will result in a higher game score, whereas negative epistemic speech acts should lead to a lower game score. This also leads to believe that, based on what has been described above, a bigger team size should lead to a higher performance of 21st century skills: by having more team members, they should be able to perform better.

Method

Goal

The goal of this study is to analyse the differences in communication between groups of two and groups of four. This will be achieved by forming a total of twenty teams: ten teams of two and ten teams of four. These teams will all play the same game under the same conditions.

Respondents and Sampling

Only students studying at the department of Behavioural, Management and Social Sciences (BMS) on the University of Twente have participated in this study (n = 60). The reasoning behind this was that their study backgrounds are similar. This made it safe to assume that their levels of English, and communication in general, are similar. There were no other requirements since no prior knowledge was necessary to play this game.

Materials

The Game.

The game used in this experiment was a modified version of the game called Isis & Osiris (2001) which was developed by Michael Schacht. The original game is a competitive game that revolves around memory. The modified version is a collaborative game that revolves around teamwork, strategy and critical thinking. The goal of the game is to reach the highest score possible. This could only be done through collaboration. Thus, each team strove to reach the highest possible score they could attain.

Each player had, depending on the team size, ten (or five) tokens and eight (or four) tiles. Each turn a player chose either a tile or a token. These tokens represented the players. The tiles added or subtracted a score from the player. They stayed faced down until picked up so that no one could see what number was on them. Whenever a player picked up a tile, it was mandatory to play that tile that turn. The order of the stack was also not to be changed. Only the tiles that laid directly (horizontally or vertically) next to the player could change their score. Diagonally connected tiles did not affect the players in any way. The scores on the tiles ranged from minus four to plus four. Both the plus and minus ones were in the game two times. Both the twos were present five times. The threes were present two times and the fours once.

Figure 3. A tile.

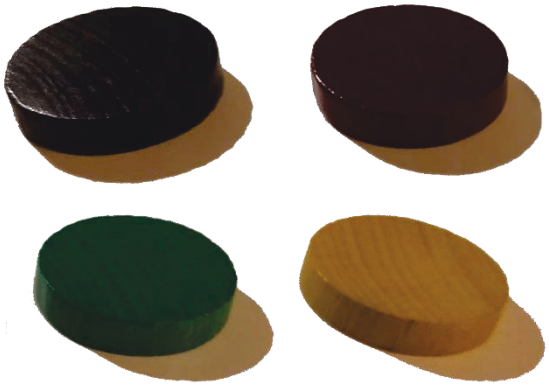


Figure 4. The Four Different Tokens

The youngest player started the game. He would place either a tile or a token on the board as he, and potentially his team, saw fit. The next player, following the clock, would then play their turn by repeating this process. This process would repeat until all playing pieces were used and the board would be filled. Every piece had to be placed on the field, filling the field completely and leaving no pieces left. When all pieces have been placed on the field a picture was made. The researcher then gave the participants the option to calculate their score, but counted the score afterwards regardless.

Figure 5. The Playing Field

Figure 6. Example of Lower Score

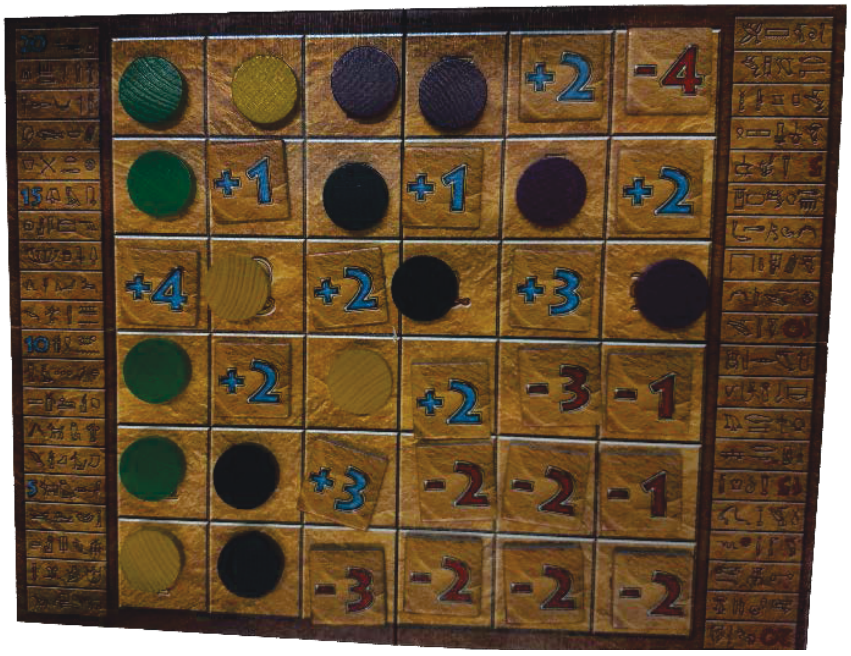


Figure 7. Example of Higher Score

Figure 6 shows an example of a team that did not score well. They connected multiple times to the higher minus values (minus three, minus four), and three tiles to the minus four, which already results in minus twelve. This will have a big impact on their maximum potential score. Moreover, they only connected two tiles to the plus four, leaving a minus four deficit when regarding only the highest tiles (plus and minus four). Figure 7 shows an example of a team that did score well. They did not connect the minus four once, connected to the plus four three times, placed most of the minus

tiles in the corner and walled them off with positive tiles. Even though this placement is not fully optimized, which would require the plus three and four to have four connections each, this group scored relatively higher by only taking four minus points, as opposed to the previous example where the group took more than thirty minus points. In the table below a summary of the rules is present.

Table 1. Rules

1) The whole field needs to be filled at the end of the game.
2) The players have 15 minutes to finish the game.
3) Each turn a player can play either a token or a tile.
4) You can flip a tile when you want to play it, but it has to be played when flipped.
5) You are not allowed to adjust the order of the tiles in the stack.
6) The youngest player starts.
7) Each time a player finishes their turn, the next player clockwise has to play.
8) Tiles only count when a token is connected to it horizontally or vertically.
9) Diagonally connected tiles do not count.
10) The researcher will warn all players when there are only five minutes left.
11) The colour of the tokens does not matter.
12) Individual score does not matter.
13) Communication is allowed and advised while playing.
14) No communication before the game has begun.
15) The researcher will not answer any questions before or during the game related to the problem space or strategy.

The Video.

Before the game the participants were shown a 4-minute video that explained the rules of the game. The goal of using a video was two-fold. First, to introduce the participants to the game. Since this was a modification of the original game, the original rules and rule book do not apply, and hence the participants needed something else to understand the rules. Second, this video was intended to be an objective baseline, providing all participants with equal knowledge about the game and its rules, hence equal opportunities to perform well. When explaining rules, even when reading from a paper, some changes will be there every instance. Therefore, it was decided that a video was the most suitable tool of introducing the game to the participants. The video immediately instructed the participants not to communicate during the video and before the game began. In order to prevent the eager participants from getting a head-start, communication was only allowed during the 15-minutes of playing the game. Firstly, the aim of the game was explained. Secondly, the basic game mechanics were explained. Thirdly, it was stressed that it was a collaborative game where individual scoring was not counted. Fourthly, the boundaries of the game were explained. Fifthly, the procedure of the game was explained. Sixthly, the scoring mechanism was explained. Seventhly,

some hints were provided. Finally, it was mentioned that the researcher will not answer any questions regarding the mechanics of the game.

Audio Recorder.

As the goal of this study is to say something about verbal communication, audio files will be used. Since the audio files were to be uploaded to AmberScript, an online program that transcribes speech-to-text automatically, it was deemed important to have high-quality audio. To ensure this, a good audio recorder was used. The name of the audio recorder used in this study is Zoom H4nPro.

Research Tactics.

What the participants did not know was that the order of the stacks of tiles was the same every game, and that the stacks – or two stacks in the case of four players – mirrored the other remaining stack(s). What that means is that the first tile would be a plus one for player one and minus one for player two. The next tile for player one was then a minus one, and for player two a plus one. The mirroring of the tiles was intended to give the players more structure and to provide a way to figure out what would come next, without making it obvious, as to stimulate the critical thinking factor. The exact same order of the stacks in every game was done to keep each game exactly the same for each team so that the score would be objective.

Moreover, in the video the participants were instructed that they were not allowed to ask the researcher any questions regarding the game before and during the game. This was done with the intent to prevent some groups from having more information than others. Moreover, since the goal of this study was to understand more about communication in relation to performance, a maximized amount of communication was preferable. In order to achieve this, part of the game was modelled so that participants had to figure out the rules together with their teammates. As mentioned previously, the video briefly went through all the rules. Figuring out problem space and making sure that every group member understood it was intended by the researcher to be the first move the teams had to make in order to successfully play the game and to achieve a high score.

Furthermore, whenever participants were making mistakes, the researcher did not intervene. The performance of teams was fully up to the teams and was not meddled with. Again, to remain as objective as possible, the researcher did not intervene or answer any questions.

Procedure

Upon entering the room, the participants received the consent form (Appendix D). They were asked to take a few minutes to read it carefully and sign with their name and autograph. The researcher then kept these forms in a safe place inaccessible to others. After all participants filled the

consent forms, they were asked to watch the video. Afterwards, the participants received their tiles and tokens and were clearly instructed not to change the order of the tiles. The researcher then repeated that all communication had to be performed in English and asked if the participants were ready to start. When all were ready the audio recorder was turned on and the participants started their communication and began playing the game for a total of fifteen minutes. When five minutes were left the researcher warned the participants about the remaining time. When the board was filled the researcher took a photo of the filled board and notified the participants that the experiment was finished. The score chart per team can be found in Appendix E.

Coding, Scoring and Data Analysis

Out of the twenty groups that participated, all data was valid and usable. The audio files were uploaded to AmberScript.com, the speech-to-text transcriber mentioned before. Since everything in this study had been done in English, including the consent form and the video, this was the best approach in terms of consistency. After the transcribing was finished, the documents were uploaded and coded in NVivo, a coding program that allows social scientists to quickly and accurately code large pieces of text. The data was coded with a coding scheme that was transposed from the model – the coding book, including two examples per sub-level can be found in Appendix B.

General.

Half statements were not coded. Sometimes, in the participants' eyes, it seemed obvious what the they meant to say so they did not finish their sentence, and sometimes they got interrupted by their fellow participants. All of these statements were not coded in order to keep the interpretation and subjectiveness of the researcher to a minimum. Moreover, statements were coded with the full meaning of an utterance: grammatically incorrect sentences or sentences with the wrong usage of the exact words such as tokens and tiles were also coded. Hence, statements were coded like they were intended to be understood. For example, an utterance like 'yes' in reply to 'should I place it here?' was considered as 'yes, you should place it there'. Statements were thus generously but consistently coded throughout.

Scoring.

Every variable will be measured on the ratio scale, as they have a meaningful zero point – zero times the speech act of – and will be counted in number of speech acts (e.g. zero correct constructions of problem space, two arguments, three integrations). Therefore, every variable can be seen as 'what is the number of', e.g. in the participation dimension, the quantity of participation would refer to 'what is the total number of speech acts per person'.

Critical Thinking.

Firstly, by adding a competitive element where the best team in their category wins a prize (cinema tickets, but this was not mentioned to the participants), the teams will be motivated to perform optimally and will turn to any source of help, which in this case is limited to their follow participants only. Secondly, by showing the participants the rules on the problem space before the start of the experiment, their mental models can align. However, since the rules are shown briefly and only once, the participants will need to communicate at the start of the game in order to formulate their mental models out loud and to ensure they share the same understanding of the problem space, hence creating a shared mental model.

Since it is difficult to measure critical thinking in itself, the game was modified in such a way that the participants did not know anything about possible solutions. They merely knew what was allowed, what was not allowed, and how to gain points. Therefore, they had to use their critical thinking skills and creativity to solve the problem at hand. Hence, critical thinking has not been measured as it is difficult to know which cognitive processes are on-going, but is considered to be well-reflected in the overall performance score.

Participation.

Only one part of the participation dimension was coded directly from the recorded speech: the quantity of participation. This was done by counting the total number of speech acts by a team. The other part of the participation dimension was heterogeneity. Heterogeneity was measured by counting the amount of speech acts per person in a team and then calculating the standard deviation based on the mean. The higher the standard deviation, the higher the heterogeneity of participation.

Epistemic.

The epistemic dimension was measured in four different ways: epistemic questions, true statements, false statements and off-topic speech acts. These were divided into two categories: positive and negative epistemic activities.

As mentioned, all speech acts related to the problem space are epistemic. An example of an epistemic question would be: “when can we turn the tiles around?”. The rules of the game stated that one can, each turn, choose a tile or token. When the tile was chosen, it had to be picked up from the top of the pile and then immediately turned around. However, it was also mentioned that once a tile has been turned, it had to be played in that turn. The players were thus not allowed to put it back. All questions related to the rules or mechanics of the game thus were epistemic questions.

All epistemic statements were either true or false. Some participants mentioned rules about individual score. The rules clearly explained that this was a collaborative game and that the colour of the tokens did not matter, and that individual scores were not counted and non-existing. Statements mentioning things about individual scores were thus coded as 'incorrect construction of problem space'. Statements such as "we need to fill the whole board" or "we only have fifteen minutes to complete the game" were flagged as 'correct construction of problem space' as these rules were true and enforced without exception.

Off-topic statements were all statements that said something about everything not relevant to the game. These types of statements were considered confounding for the other players as it took their attention away from the game and took away from the total amount of game time. For example, one participant mentioned that they usually did not pay attention to game rules. Another participant blamed their team member for not having the right understanding of the rules, while they both saw the same video and rules. Hence, these types of statements are thought to decrease performance and were thus considered to be negative.

Argumentive.

The argumentive dimension was measured in five different ways: questions, statements, arguments, counterarguments and the integration of counterarguments. These were also divided in two categories: higher and lower order communication.

Questions in this category related to what the next move should be, e.g. "should I place it here?" and "should I pick a tile or a token now?". They generally held no meaning as nothing was expressed.

Statements indicated what the next move should be, without any base or argument. Examples include statements such as: "place it here" and "put it in the corner". Moreover, statements that confirmed argumentive questions such as replying "yes" to the question "should I put a token here?" were also considered argumentive statements as they also did not have any base or justification, but did express a direction of play. However, only the first confirmation of a question or proposed tactic was coded as a statement as some teams replied "yes" to a question more than five times. By coding these types of statements only once the difference between statements and group size was kept to an objective minimum, as bigger groups replied positively or negatively more often due to the amount of people sitting at the table.

Arguments, counterarguments and integration constituted the higher order statements as mentioned before. Arguments indicated a strategy or next move and provided a justification for this. An example would be "Do not put it near the edge because then you cannot place a tile or token next

to it". Another example would be "We should place it there because then we can get the points from that tile more often".

Counterarguments were all statements that rejected either arguments or argumentive statements. One participant mentioned "we should put this tile here". This tile was a plus one, the lowest positive tile. Their team member responded with "No we shouldn't because this is a low positive, and if we put it there [in the middle of the field] it would block the rest of the tiles. Therefore, we should place the higher positives in the middle and the lower positives at the edge so we can connect more tokens to the higher tiles, increasing our overall score". This was considered a good counterargument as it corrected a statement, steered the players to a better strategy and helped the rest of the team understand how to get a better score. The first participant then replied with "yes, you're right", indicating a change of behaviour based on a (counter) argument: argumentive integration. Integration was considered the highest form of argumentive behaviour because the participant adjusted their strategy based on an upgraded and renewed understanding of what constituted a good strategy. Hence, one could argue that this type of statements is where you can see learning occur. Nevertheless, whereas counterarguments could oppose both statements and arguments, integration could only happen after arguments or counterarguments. It was considered that participants did not necessarily need to voice their opinion before they could integrate knowledge from their fellow participants. Therefore, in order to account for unverbalized cognitive processes, integration could also happen after an argument from their peers, not necessarily when their own arguments got countered.

Data Analysis.

After the data was successfully coded, an interrater reliability test was done by a colleague of the same study. Cohen's kappa coefficient was calculated at 0.78. This was deemed acceptable. Afterwards, the data was transposed to SPSS. Here the data of the twenty groups was analysed. An independent sample t-test was used to see if a significant difference could be found between group size and the previously mentioned variables. The effect of each of the other variables on the team score was measured using the Pearson product-moment correlation coefficient. To describe the strength of the correlation the recommendation by Evans (1996) for the absolute value of r has been used:

- .00-.19 "very weak"
- .20-.39 "weak"
- .40-.59 "moderate"
- .60-.79 "strong"

EFFECT OF TEAM-SIZE ON COMMUNICATION

- .80-1.0 “very strong”

Results

In order to answer the research question '*What is the effect of team-size on the nature of communication?*', the relation between collaboration, communication and critical thinking was studied. Moreover, the effect of group size, between teams of two and teams of four, upon each of these variables was calculated. The sub-questions mentioned previously help provide the necessary insight into the data required to answer the main research question.

Table 3 shows the absolute data regarding the variables present in the sub-questions. The mean and standard deviation per amount of team members of team score, quantity of participation, heterogeneity of participation, positive and negative epistemic speech acts, and higher order statements can be found. Below table 4 the findings for each of the sub-questions have been noted. For a visual representation a scatterplot of each of the correlations can be found in Appendix C.

Table 2. Mean Totals per Amount of Team Members

Participants	Team Score M (SD)	Quantity of Participation M (SD)	Heterogeneity of Participation M (SD)	Positive Epistemic Speech Acts M (SD)	Negative Epistemic Speech Acts M (SD)	Higher Order Statements M (SD)
2 (n = 10)	36 (19.7)	152.6 (45.9)	8.1 (7.8)	10.2 (7.6)	5 (7.7)	14.8 (7.3)
4 (n = 10)	49 (12.7)	206.6 (47.8)	21 (6.3)	17.2 (7.5)	3.1 (5.1)	22.5 (12.7)
Total	42.5 (17.5)	178.6 (53.3)	14.6 (9.6)	13.7 (8.1)	4.1 (6.4)	18.7 (10.8)

The differences between smaller and bigger groups

Firstly, a t-test was conducted to check if bigger groups got a higher score. There was an insignificant difference; t (18)=1.750, p = 0.100. Bigger groups did not get a higher game score. Secondly, a t-test was conducted to check if bigger groups got a higher quantity of participation. There was a significant difference; t (18)=2.579, p = 0.019. Bigger groups did have more quantity of participation. Thirdly, a t-test was conducted to check if bigger groups got a higher level of heterogeneity. There was a significant difference; t (18)=4.054, p < 0.001. Bigger groups did have more heterogeneity. Fourthly, a t-test was conducted to check if bigger groups got more lower order communication. There was no significant difference; t (18)=0.568, p = 0.577. Bigger groups did not have more higher order communication. Fifthly, a t-test was conducted to check if bigger groups got more higher order communication. There was no significant difference; t (18)=1.665, p = 0.113. Bigger groups did not have more higher order communication. Sixthly, a t-test was conducted to check if bigger groups got more positive epistemic speech acts. There was no significant difference; t (18)=2.072, p = 0.053. Bigger groups did not have more positive epistemic speech acts. Finally, a t-test was conducted to

check if bigger groups got more negative epistemic speech acts. There was no significant difference; $t(18)=0.651$, $p = 0.523$. Bigger groups did not have more negative epistemic speech acts.

The effect of the measured variables on performance

Firstly, there was a weak positive significant correlation between the two variables ($r = .390$, $p = .045$). It was observed that more overall speech acts resulted in a higher score. Secondly, there was a very weak positive not significant correlation between the two variables ($r = .187$, $p = .215$). Groups that had a higher level of heterogeneity did not score higher. Thirdly, there was a weak positive not significant correlation between the two variables ($r = .278$, $p = .118$). The groups that had more lower order statements did not significantly higher than the groups with less lower order statements. Fourthly, there was a moderate positive significant correlation between the two variables ($r = .486$, $p = .015$). The groups that had more higher order statements scored significantly higher than the groups with less higher order statements. Fifthly, there was a weak negative not significant correlation between the two variables ($r = -.217$, $p = .179$). Having more positive epistemic statements did not result in a significantly higher game score. Finally, there was a strong negative significant correlation between the two variables ($r = -.704$, $p < 0.001$). Having more negative epistemic statements resulted in a strongly significant lower game score.

Discussion

Participation and Performance

The aim of the study was to investigate the effect of team-size on the nature of communication. No higher performance was observed in bigger groups, which was unexpected. Groups of four did not score significantly higher than groups of two. However, they did have a lower standard deviation, which indicates that they are, on average, more reliable or stable when performing a team task: high standard deviation would imply that the teams in that category are relatively unreliable because it is less predictable if they will perform well or not. For example, if teams of four would all score forty points, and teams of two would score from zero to eighty points, having a team of four would be more reliable since one could presuppose what score they will get.

What was expected is that the quantity of participation is significantly higher for groups of four over the dyads. However, in absolute terms the increase seems logical. Twenty-six percent more speech acts in fifteen minutes, while doubling the amount of team members, seems like a logical increase. Moreover, the results show that an increase in quantity of participation also increases the overall score. This is in line with the idea that many hands make light work: sharing more perspectives increases the options and hence make a better outcome possible. However, it is still up to the participants to make the correct choice. Having more choices does not make it easier to choose one of those options, nor the correct one. Moreover, in groups of four the heterogeneity of participation was also higher. It was observed that in teams of four one person usually took the lead. The data also confirmed this with an almost trifold increase in heterogeneity for groups of four. It was predicted that a higher heterogeneity would decrease the overall score. This was not the case. Some teams needed a leader in order to perform well and started to play a lot better when one person took the lead. In these situations, the leader understood the rules and carried their team to play well. Their team only challenged the leader when the rest of the team really thought their leader was wrong. This resulted in a high number of heterogeneity, but incorrectly reflects whether teams were engaging in quick consensus seeking or not. Overall, it seems that this variable is not suitable for measuring quick consensus seeking in teams of four.

Communication

As expected, more higher order statements correlate to a higher score: these are the verbalizations that are considered the highest form of communication. Based on the ideas that communication stimulates team performance and that these types of utterances work well to share one's perspective, this shows that adding context and sharing perspectives are good ways to help a

team perform. This makes sense: adding a reason as to why a certain move has to be made provides the rest of the team with the related thoughts behind it. This gives them a base on which to judge the move. This saves both time and thinking power, as opposed to them having to figure everything out from scratch. Moreover, it prevents ambiguity by eliminating the possibility of assuming other reasons to do a specific move. Hence, higher order statements are effective and clarifying. Moreover, lower order statements did not increase or decrease team performance. However, one could argue that for every lower order statement made, a team loses time to make higher order statements, which have shown to be effective in increasing the performance of a team. However, 'what ifs' are impossible to measure.

An increase in the number of positive epistemic speech acts (PESA) was also expected to lead to a significant increase in score. As stated by Hogan et al. (2000), the problem space has to be constructed in the right way in order to understand the problem. However, in this study that is not the case. One team of two had zero PESA and still managed to place in the top twenty-five percent, whereas the lowest scoring team of four had a lot of PESA, relative to the average. An explanation for this result can be that the construction of the problem space mainly happens in one's head. Therefore, counting the utterances alone is not sufficient to measure the understanding of the problem space. This could be solved by adding a multiple-choice test on the rules of the problem space after the game was finished. This would then serve as an indicator of whether each participant understood the problem space, even if they did not communicate about it. By asking the participants specific questions about the game, for example 'How many tiles are there on the board after the game has ended?' or 'Who starts?' you can figure out if they understood the problem space. If so, the need to communicate about these specific topics could be lower.

Unlike the PESA, the negative epistemic speech acts (NESA) have a strong negative correlation, indicating that the more incorrect and off-topic statements had been made, the worse a team performed. The literature stated that PESA would positively affect performance and said nothing about NESA. This study thus challenges the outcomes of the study done by Hogan et al. (2000) as both results about problem spaces postulated by Hogan et al. (2000), about both PESA and NESA, did not match the conclusions drawn. However, a distinction between PESA and NESA had not been made in the study done by Hogan et al. (2000). If the authors tested the variable without distinction, it could have been the case that NESA was actually the third variable influencing the total result the most, hence making the result – in a sense – unclear. Hence, future research on the relation between PESA and NESA is advised.

Conclusion

The results show that groups of four only scored significantly higher on quantity and heterogeneity of participation. As mentioned, this was to be expected. What was not expected was that they did not score higher and did not have more PESA. The only two variables significantly increasing the score were quantity of participation and higher order statements. Hence, it seems that the best way to increase a team's critical thinking skills, and thus overall performance, is by having them focus on providing arguments, counterarguments and integration.

Within the collaborative skill dimension only the quantity of participation seems to make a difference on performance. It makes sense that more overall collaboration also increases the collaboration skill. Thus, to improve one's collaboration skills, one should work more on collaborative tasks. This is in line with the general conception of 'practice makes perfect'.

Within the communication skill dimension only higher order statements improved performance. Hence, this would lead us to also conclude that only a few, specific forms of communication lead to better communication skills in general. Thus, to improve this skill, learners should work on improving their argumentive skills. This can be done by following training on how to argue or debate, a class on reasoning fallacies, or adding scripting to dialogic situations. This is in line with what Van der Meij, Veldkamp, & Leemkuil (2019) recently found, namely that scripting leads to more dialogical acts of a deeper communicational level.

Within the critical thinking skill dimension, it seems that, apart from what was previously mentioned, one could also argue that scores largely depend on individual capacity. This agrees with Trower & Moore (1996) who found that a higher team size does not significantly increase team scores per se, but is heavily dependent on the combined total of individual capacity. Within this research, a few indicators of this idea were also present. Some teams did not need to communicate much because every member understood the rules just by watching the video. They saw structures, connections and creative opportunities immediately, whereas other players kept staring at the best option and decided to do something else based on something that they did not communicate. However, to figure out one's individual capabilities is very difficult, if not impossible. There are some tests, but these test only specific dimensions, such as an IQ test. The question is, which specific dimensions should one be good at the solve a specific problem? This differs per type of performance, and is difficult to predict beforehand. Hence, the only way to circumvent this problem is by increasing the total number of participants drastically, downplaying the differences between individuals.

Lastly, it seems that many of these skills are interconnected. This research implies that improving one's communication skills would also lead to improve one's collaborative skills. By

improving these two skills, one could theoretically improve team performance as well. Therefore, it seems that research on 21st century skills is useful to understand more about them, and in turn can be used to improve the overall performance of individuals.

In conclusion, it seems that the effect of team size on 21st century skill performance is relatively small. Teams of four do not perform better than teams of two in terms of score, did not have more higher order statements (which is believed to be the key to increasing collaboration and the quality of communication), nor is their quality of communication in general better. However, it should be noted that these teams differ in that they are freshly formed and do not work on a complicated task that required a lot of conceptual (prior) knowledge. Therefore, these results do not apply to organizational or highly specialised teams.

Limitations, reflection and for the future

General game experience.

In general, practice makes perfect. This goes for games as well. Game experience is unaccounted for in this study. However, experience is not a rule of thumb within the gaming world: some people play the same game casually for years and still are not very good at it, while others are more competitively orientated and play on the highest level within months. Within this study most groups of which the members play a lot of games were positioned in the middle of the score list, so it seems safe to say that experience was not a factor that influenced the outcome.

Moreover, the overall quality of speech acts was not measured. Some speech acts, in the same category, are a lot more valuable than others. A question, for example, can take many forms. Some have been formulated in a way that they are valuable to the overall understanding of the game; other questions were repeated over and over and did not add any value. However, one does not know why people say something in a specific situation: many ongoing processes in their head together form an expression, but why this expression is uttered in a specific way or on a specific time will stay unclear to us as long as we have not figured out the exact workings of the human brain. Which cognitive processes are going on inside is unbeknownst to researchers, and can therefore not be analysed and used in studies. This is a limitation for the social sciences in general that should be accounted for in some way.

Finally, unguided communication and measuring learning is a difficult combination due to various factors such as repetition and understanding. To have the participants communicate as much as possible, in an honest way, the game had to be new to them. Therefore, using a pre- and post-test, which in most cases would be ideal to measure learning, was not a possibility. A pre-test would require the participants to play the game multiple times. However, this would impede the communication severely, as most utterances on the best strategy or problem space would be communicated in the pre-test, leaving barely any communication other than what every person thinks is the right solution based on what they saw after finishing the first round. Hence, it is advised for future research to consider this difficult duality when studying learning and or through communication.

Future research.

The relation between positive and negative constructions of problem spaces has proven to be problematic and under-researched. This variable is important in constructing epistemic knowledge and should thus be understood well if one wants to measure utterances on problem spaces. This is in line with what was mentioned previously, namely that in this study only the

quantity of statements has been considered; one would ideally also consider the quality of those statements. To the best of the researcher's knowledge, such a tool does not yet exist – at least not one of sufficient quality to be used in scientific research. Transcribing the documents revealed that even though some speech acts went into the same category, they were of very different quality. It would thus be inaccurate to see them as equally valuable.

Finally, literature mostly speaks of existing teams that have worked together for a longer period of time. These teams' dynamics are very different from freshly formed teams like the ones that participated in this study. Therefore, it would be advised to sort out the differences between these two before conducting a study on teams and performance, which to the best of the author's knowledge has not been done before.

References

Alderfer, C. P. (1977). Improving organizational communication through long-term intergroup intervention. *Journal of Applied Behavioral Science, 1*(3), 193-210.

Asterhan, C. S., & Schwarz, B. B. (2016). Argumentation for learning: Well-trodden paths. *Educational Psychologist, 51*(2), 164-187.

Austin, J. L. (1975). *How to do things with words.* Cambridge: Harvard University Press.

Binkley, M., Erstad, O., Herman, J., Raizen, S., Ripley, M., Miller-Ricci, M., & Rumble, M. (2012). Defining Twenty-First Century Skills. In P. Griffin, B. McGaw, & E. Care, *Assessment and Teaching of 21st Century Skills* (pp. 17-66). Dordrecht: Springer.

Bradner, E., Mark, G., & Hertel, T. D. (2005). Team size and technology fit: Participation, awareness, and rapport in distributed teams. *IEEE Transactions on Professional Communication, 48*(1), 68-77.

Burbules, N. C. (1993). *Dialogue in Teaching: Theory and Practice.* London: Teachers College Press.

Buskes, C. (2006). *Evolutionair Denken.* Amsterdam: Nieuwezijds.

Castells, M. (1996). *The information age: economy, society and culture.* Oxford: Blackwell.

Chan, C. K., Burtis, P. J., & Bereiter, C. (1997). Knowledge building as a mediator of conflict in conceptual change. *Cognition and Instruction, 15*(1), 1-40.

Cochrane, P., & Eversole, B. (2012). Student laboratory team performance as related to team size. *Journal of Technology, Management, and Applied Engineering, 28*(4), 2-8.

Engeström, R. (1995). Voice as communicative action. *Mind, Culture and Activity, 2*(3), 192-215.

Ennis, R. H. (2018). Critical thinking across the curriculum: A vision. *Topoi, 37*, 165-184.

Fischer, F., Bruhn, J., Gräsel, C., & Mandl, H. (2002). Fostering collaborative knowledge construction with visualization tools. *Learning and Instruction, 12*, 213-232.

Gladstein, D. L. (1984). Groups in context: A model of task group effectiveness. *Administrative Science Quarterly, 29*, 499-517.

Habermas, J. (trans. 1981). *The theory of communicative action. Volume I: Reason and the rationalization of society and Volume II: Lifeworld and system: A critique of functionalist reason.* Boston: Beacon Press.

Hayse, M. (2018). Tabletop games and 21st century skill practice in the undergraduate classroom. *Teaching Theology & Religion*, 288-302.

Hogan, K., Nastasi, B. K., & Pressley, M. (2000). Discourse patterns and collaborative scientific reasoning in peer and teacher-guided discussions. *Cognition and Instruction, 17*(4), 379-432.

Huizinga, J. (1949). *Homo ludens: A study of the play-element in culture.* London, Boston & Henley: Routledge & Kegan Paul.

Iordanou, K. (2010). Developing argument skills across scientific and social domains. *Journal of Cognition and Development, 11*(3), 293-327.

Janhonen, S., & Sarja, A. (2000). Data analysis method for evaluating dialogic learning. *Nurse Education Today, 20*, 106-115.

Kaufman, K. J. (2013). 21 ways to 21st century skills: Why students need them and ideas for practical implementation. *Kappa Delta Pi Record*, 78-83.

Keefer, M. W., Zeitz, C. M., & Resnick, L. B. (2000). Judging the quality of peer-led student dialogues. *Cognition and Instruction, 18*(1), 53-81.

Kickmeier, M., & Albert, D. (2012). A Domain Model for Smart 21st Century Skills Training in Game-Based Virtual Worlds. *12th International Conference on Advanced Learning Technologies*, (pp. 680-681). Rome.

Kuhn, T. S. (1962). *The structure of scientific revolutions.* Chicago: University of Chicago Press.

Leitão, S. (2000). The potential of argument in knowledge building. *Human Development, 43*, 332-360.

Lipman, M. (2003). *Thinking in Education.* Cambridge: Cambridge University Press.

Lowry, P. B., Roberts, T. L., Romano, Jr., N. C., & Cheney, P. D. (2006). The impact of group size and social presence on small-group communication: Does computer-mediated communication make a difference? *Small Group Research, 37*(6), 631-661.

Moore, G. C., & Benbasat, I. (1991). Development of an instrument to measure the perceptions of adopting an information technology innovation. *Information Systems Research, 2*(3), 192-222.

Morreale, S. P., & Pearson, J. C. (2008). Why communication education is important: The centrality of the discipline in the 21st century. *Communication Education, 57*(2), 224-240.

Nastasi, B. K., & Clements, D. H. (1992). Social-cognitive behaviors and higher-order thinking in educational computer. *Learning and Instruction, 2*, 215-238.

Nietzsche, F. (trans. 1998). *On the Genealogy of Morality.* Leipzig.

Plato. (360 BC). *Symposium* (Vol. 3). Athens.

Qian, M., & Clark, K. R. (2016). Game-based Learning and 21st century skills: A review of recent research. *Computers in Human Behavior, 63*, 50-58.

Romero, M., Usart, M., & Ott, M. (2015). Can serious games contribute to developing and sustaining 21st century skills? *Games and Culture, 10*(2), 148-177.

Rosas, R., Nussbaum, M., Cumsille, P., Marianov, V., Correa, M., Flores, P., . . . Salinas, M. (2003). Beyond Nintendo: design and assessment of educational video games for first and second grade students. *Computers & Education, 40*(1), 71–94.

Salas, E., Sims, D. E., & Burke, S. (2005). Is there a "big five" in teamwork? *Small Group Research*, 555-599.

Sourmelis, T., Ioannou, A., & Zaphiris, P. (2017). Massively Multiplayer Online Role Playing Games (MMORPGs) and the 21st century skills: A comprehensive research review from 2010 to 2016. *Computers in Human Behavior, 67*, 41-48.

Swanson, E., McCulley, L. V., Osman, D. J., Lewis, N. S., & Solis, M. (2019). The effect of team-based learning on content knowledge: A meta-analysis. *Active Learning in Higher Education*, 39-50.

Trower, J. K., & Moore, K. K. (1996). A study of the impact of individual goals and team composition variables on team performance. *ACM SIGCPR Conference*, (pp. 206-213). Waco.

Van den Bossche, P., Gijselaers, W., Segers, M., & Kirschner, P. (2006). Social and cognitive factors driving teamwork in collaborative learning environments: Team learning beliefs and behaviors. *Small Group Research, 37*(5), 490-521.

van der Meij, H., Albers, E., & Leemkuil, H. (2011). Learning from games: Does collaboration help? *British Journal of Educational Technology, 42*(4), 655-664.

Van der Meij, H., Veldkamp, S., & Leemkuil, H. H. (2019). Effects of scripting on dialogues, motivation and learning outcomes in serious games. *British Journal of Educational Technology*.

van Laar, E., van Deursen, A. J., van Dijk, J. A., & de Haan, J. (2017). The relation between 21st-century skills and digital skills: A systematic literature review. *Computers in Human Behavior, 72*, 577-588.

Voogt, J. M., & Pelgrum, H. (2005). ICT and curriculum change. *Human Technology; an Interdisciplinary Journal on Humans in ICT Environments, 1*(2), 157-175.

Voogt, J., & Roblin, N. (2012). A comparative analysis of international frameworks for 21st century competences: Implications for national curriculum policies. *Journal of Curriculum Studies, 44*, 299-321.

Vygotsky, L. (1970). *Thought and language.* Cambridge: M.I.T. Press.

Wechsler, S. M., Saiz, C., Rivas, S. F., Vendramini, C. M., Almeida, L. S., Mundim, M. C., & Franco, A. (2018). Creative and critical thinking: Independent or overlapping components? *Thinking Skills and Creativity*, 114-122.

Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers and Education, 46*(1), 71-95.

Wouters, P., & van Oostendorp, H. (2013). A meta-analytic review of the role of instructional support in game-based learning. *Computers & Education, 60*(1), 412–425.

Appendices

Appendix A

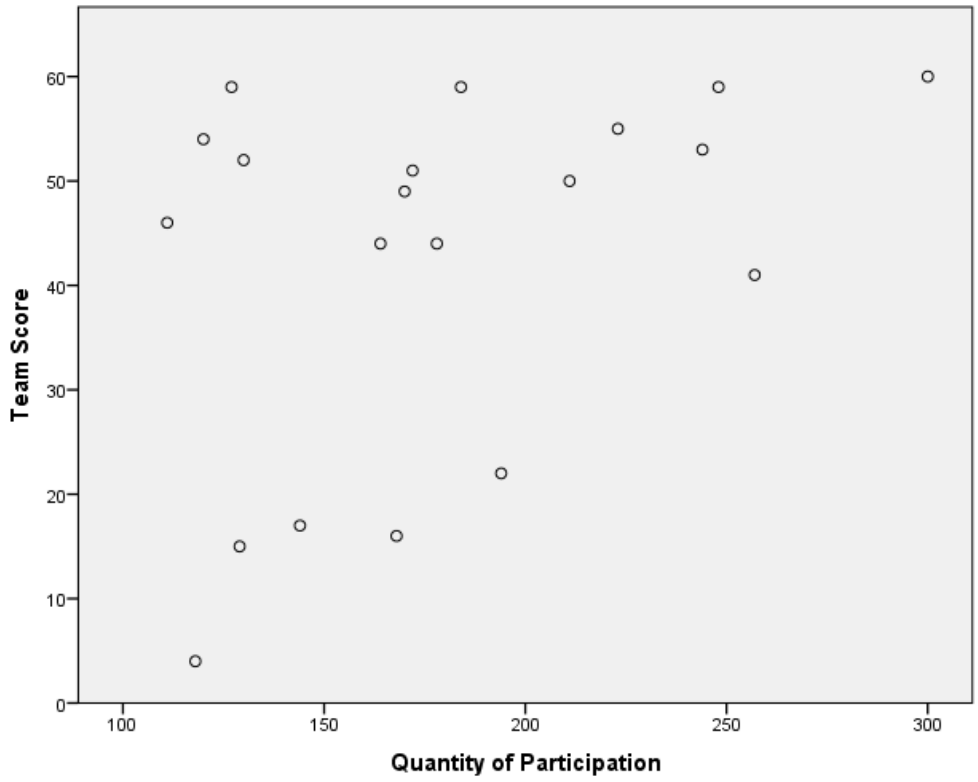
Dimension	Category	Description
Participation	**Quantity of participation**	Entering a CSCL environment and contributing to online discourse
	Heterogeneity of participation	(Un-)Equal participation of learners in the same group
Epistemic	**Construction of problem space**	Learners relate case information to case information within the problem space with the aim to foster understanding of the problem
	~~Construction of conceptual space~~	Learners relate theoretical concepts with each other and explain theoretical principles to foster understanding of a theory
	~~Construction of adequate relations between conceptual and problem space~~	Applying the relevant theoretical concepts adequately to solve a problem. Learners relate theoretical concepts to case information. A number of concept-case-relations may need to be constructed to adequately solve a complex problem
	~~Construction of inadequate relations between conceptual and problem space~~	Applying theoretical concepts inadequately to the case problem. Learners may select the wrong concepts or may not apply the concepts according to the principles of the given theory
	~~Construction of relations between prior knowledge and problem space~~	Applying concepts that stem from prior knowledge rather than the new theoretical concepts that are to be learned
	Non-epistemic activities	Digressing off-topic
Micro-level Argumentive	**Simple claim**	Statements that advance a position without limitation of its validity or provision of grounds that warrant the claim
	~~Qualified claim~~	Claim without provision of grounds, but with limitation of the validity of the claim (with qualifier)
	~~Grounded claim~~	Claim without limitation of its validity, but with the provision of grounds that warrant the claim
	~~Grounded and qualified claim~~	Claim with grounds that warrant the claim and a limitation of its validity
	Non-argumentative moves	Questions, coordinating moves, and meta-statements on argumentation
Macro-level argumentive	**Argument**	Statement put forward in favour of a specific proposition
	Counterargument	An argument opposing a preceding argument, favouring an opposite proposition
	Integration (reply)	Statement that aims to balance and to advance a preceding argument and counterargument
	Non-argumentative moves	Questions, coordinating moves, and meta-statements on argumentation
Social modes	~~Externalization~~	Articulating thoughts to the group
	~~Elicitation~~	Questioning the learning partner or provoking a reaction from the

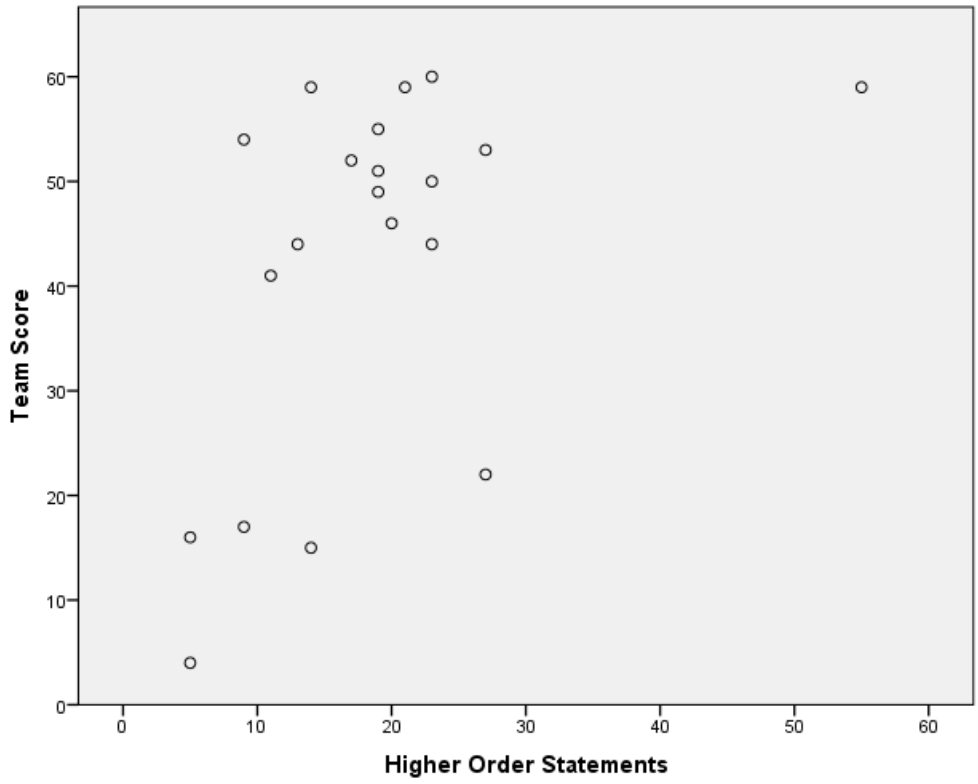
		learning partner
	Quick consensus building	Accepting the contributions of the learning partners in order to move on with the task
	~~Integration oriented consensus building~~	Taking over, integrating and applying the perspectives of the learning partners
	~~Conflict-oriented consensus building~~	Disagreeing, modifying or replacing the perspectives of the learning partners

Appendix B

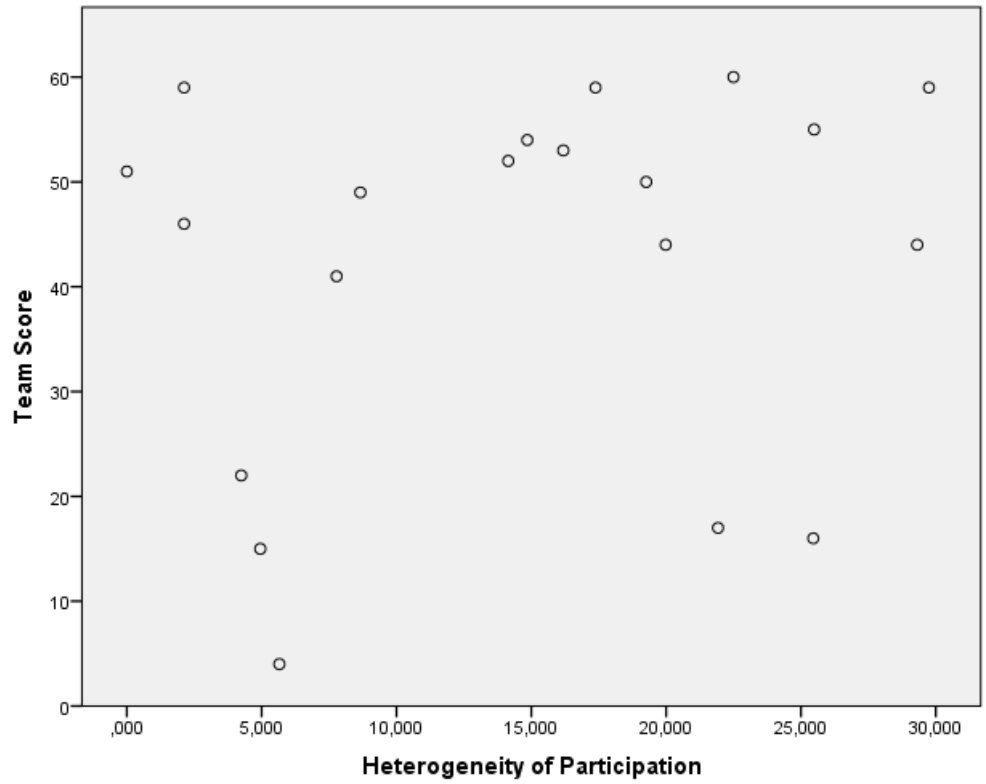
Dimension	Values	Definition	Example 1	Example 2
Participative	Quantity of participation	The number of speech acts per person	X	X
Epistemic	Questions on construction of problem space	Questions about the rules or mechanics of the game	“So, the colors don't matter?”	“How many minus ones do we have?”
	Correct construction of problem space	Correct speech acts on the rules or mechanics of the game	“Because we have ten minus tiles.”	“But you don't want to connect everything; there are also negative points.”
	Incorrect construction of problem space	Incorrect speech acts on the rules or mechanics of the game	“So that's my score then.”	“We should have placed the tokens first.”
	Non-epistemic activities	Overall off-topic speech acts	“The problem with games that usually I just tried to play, and I don't listen to the rules.”	“Should have paid more attention.”
Argumentive	Questions	Questions about the next move	“Should put one next to it?”	“Shall I play this one?”
	Statements	Speech acts on next move without justification	“Oh, we can place it here.”	“So you just place it here I guess.”
	Arguments	Speech acts on next move with justification	“I think we should count first because they have to be next to one of those that we spread them out even equally.”	“The tile is useful because you can make the patterns and then afterwards you fill in the blanks instead of already determining where your grit is going to be.”
	Counterarguments	Rejection of arguments	“No. I'd like to see what the points are before.”	“But then it would have been better if you put a coin there because then it also had two contact points for each.”
	Integration	A change in behaviour because of a(n) (counter)argument	“Yeah otherwise you don't know where to put the tiles.”	“Oh yeah.”

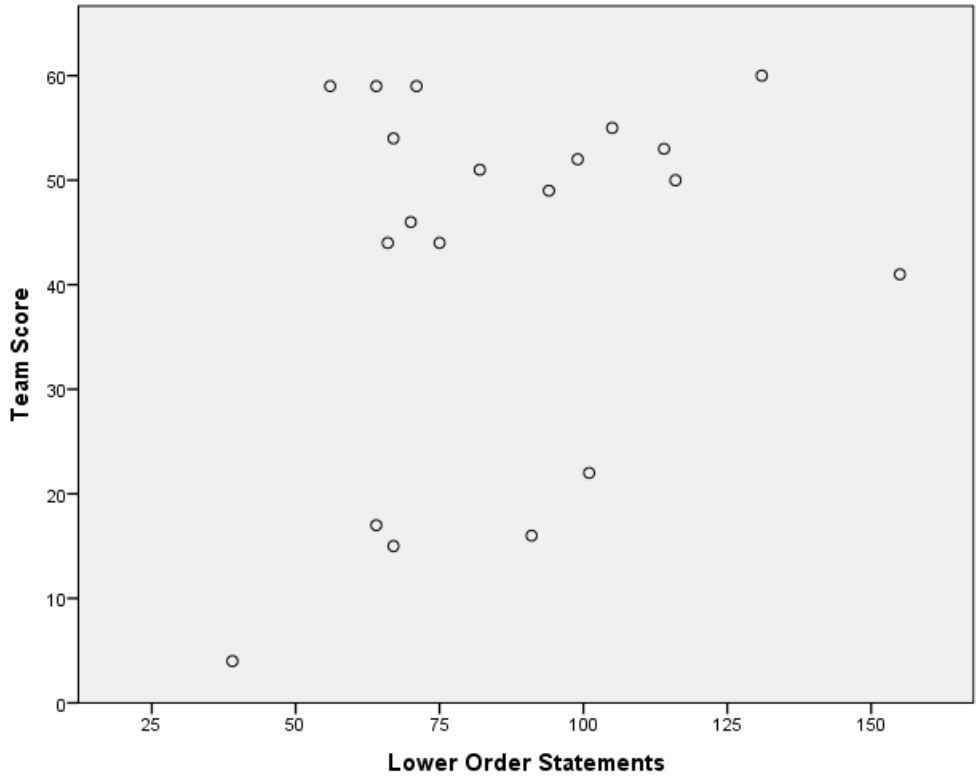
Appendix C

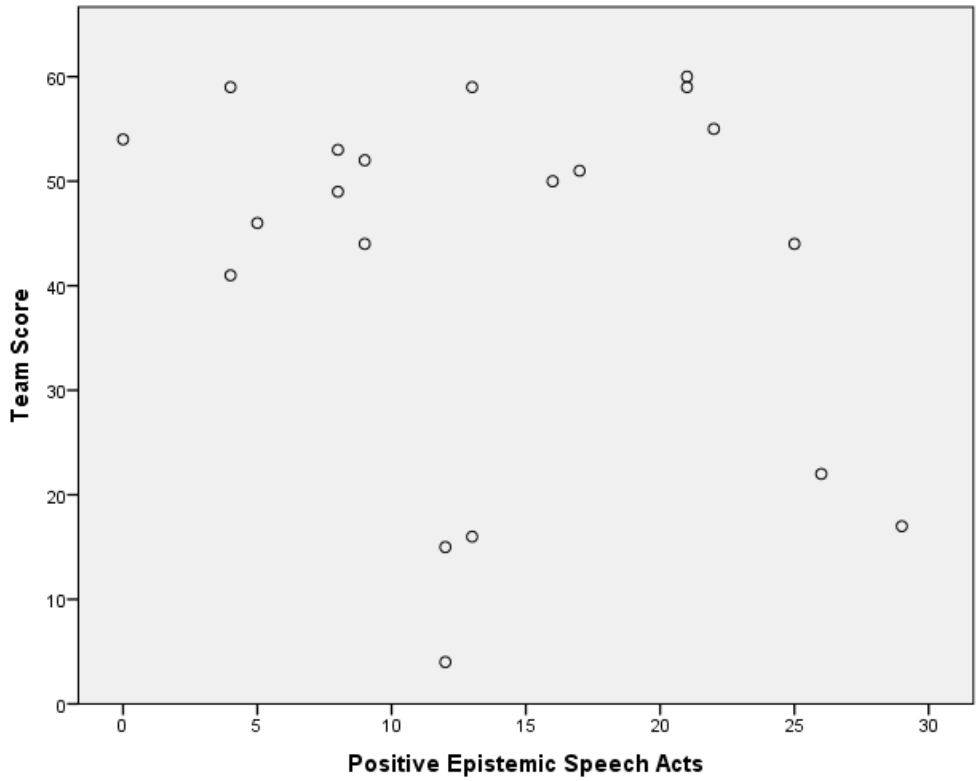

Team Score
Quantity of Participation
0
10
20
30
40
50
60
100
150
200
250
300

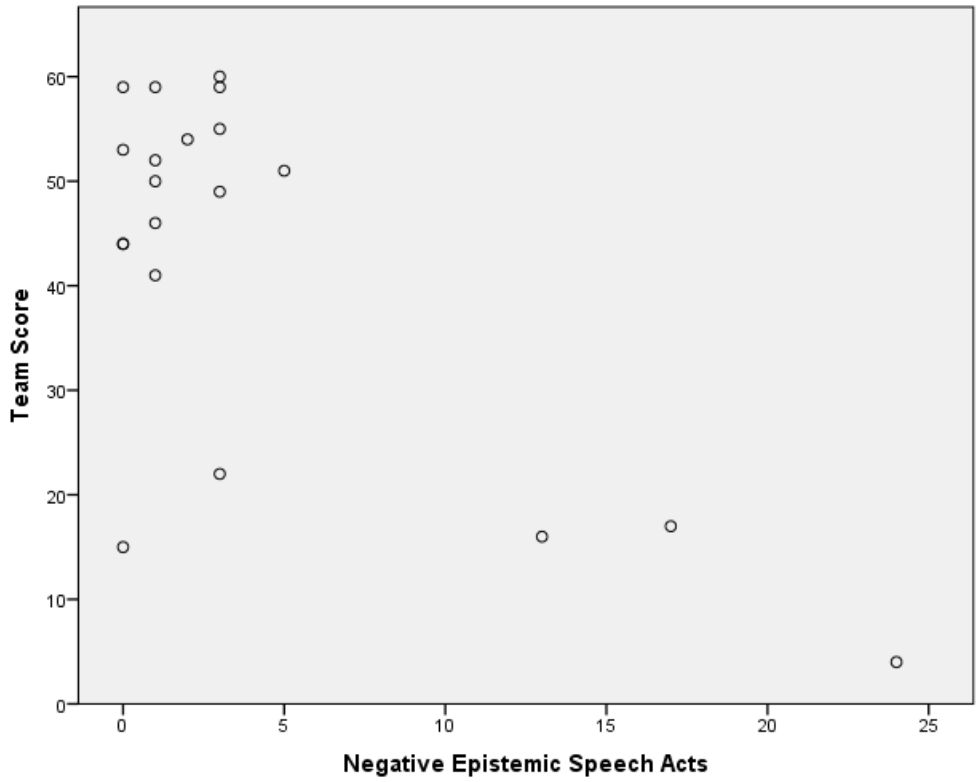

Team Score
Higher Order Statements
0
10
20
30
40
50
60

EFFECT OF TEAM-SIZE ON COMMUNICATION






Team Score
60
50
40
30
20
10
0
0
5
10
15
20
25
30
Positive Epistemic Speech Acts


Team Score
60
50
40
30
20
10
0
0
5
10
15
20
25
Negative Epistemic Speech Acts

Appendix D

INFORMED CONSENT FORM

Project Title
An investigation into the effect of team-size on learning through dialectic game participation

Purpose of the Study
This research is being conducted in order to measure the effect of team-size on learning through communication while playing a game. I am inviting you to participate in this research project.

Procedures
You will participate in an experiment where you will watch a video and then play a board game, lasting approximately 30 minutes total. Your communication will be recorded using an audio recorder. You must be at least 16 years old.

Potential Risks and Discomforts
There are no obvious physical, legal or economic risks associated with participating in this study. You do not have to answer any questions you do not wish to answer. Your participation is voluntary, and you are free to discontinue your participation at any time.

Potential Benefits
Participation in this study does not guarantee any beneficial results to you. However, as a result of participating you may get a prize for being the best team in your category by finding the relatively best solution.

Confidentiality
Your privacy will be protected to the maximum extent allowable by law. No personally identifiable information will be reported in any research product. Within these restrictions, results of this study will be made available to you upon request.
Your name will be replaced by a pseudonym during the transcription of the audio recordings. After this transcription the audio recordings will be deleted, ensuring no one will have access to any of your personal data apart from the researcher.
As indicated above, this research project involves making audio recordings of your presentation. The audio recordings, forms, and other documents created or collected as part of this study will be stored in a secure location.

Right to Withdraw and Questions
Your participation in this research is completely voluntary. You may choose not to take part at all. If you decide to participate in this research, you may stop participating at any time. If you decide not to participate in this study or if you stop participating at any time, you will not be penalized or lose any benefits to which you otherwise qualify. The data you provided before you stopped participating however will be processed in this research; no new data will be collected or used.
If you decide to stop taking part in the study, if you have questions, concerns, or complaints, or if you need to report an injury related to the research, please contact the primary investigator Jeroen Brands, j.brands@student.utwente.nl

Statement of Consent
Your signature indicates that you are at least 16 years of age; you have read this consent form or have had it read to you; your questions have been answered to your satisfaction and you voluntarily agree that you will participate in this research study.

I agree to participate in a research project led by Jeroen Brands. The purpose of this document is to specify the terms of my participation in the project.

1. I have been given sufficient information about this research project. The purpose of my participation in this project has been explained to me and is clear.
2. My participation in this project is voluntary. There is no explicit or implicit coercion whatsoever to participate.
3. Participation involves being recorded by a researcher from the Educational Science & Technology department. The experiment will last approximately 30 minutes. I allow the researcher to take written notes during my participation. I also allow the audio recording of the participation.
4. I have the right not to answer any of the questions. If I feel uncomfortable in any way during the session, I have the right to withdraw.
5. I have been given the explicit guarantees that the researcher will not identify me by name or function in any reports using information obtained from this interview, and that my confidentiality as a participant in this study will remain secure.
6. I have been given the guarantee that this research project has been reviewed and approved by Henny Leemkuil and by the BMS Ethics Committee. For research problems or any other question regarding the research project, the Secretary of the Ethics Commission of the faculty Behavioural, Management and Social Sciences at University Twente may be contacted through ethicscommittee-bms@utwente.nl.
7. I have read and understood the points and statements of this form. I have had all my questions answered to my satisfaction, and I voluntarily agree to participate in this study.

_____	_____	_____
Name of the participant	Signature	Date
_____	_____	_____
Name of the investigator	Signature	Date

Appendix E

Score chart

2 players	4 players
Team Onwijs Score = 59	Team Fancy Ancy Score = 17
Team De Zonnebloemen Score = 22	Team Patio 24 Score = 50
Team Sirius Score = 15	Team Phoenix Score = 44
Team Black Stories Score = 46	Team Domum Matris Vestrae Score = 55
Team Valencia Score = 52	Team Work In Progress Score = 59
Team P-Kringeltje Score = 54	Team Aaaa Score = 53
Team Russia Score = 16	Team Jeremiah Score = 44
Team India Score = 4	Team Inertus Score = 49
Team Stenna Score = 41	Team Piratio Score = 59
Team EST Score = 51	Team Penny Lane Score = 60