DOCTORAL SCHOOL OF EDUCATION
FACULTY OF ARTS
UNIVERSITY OF SZEGED

ANITA PÁSZTOR-KOVÁCS

THE ASSESSMENT OF COLLABORATIVE PROBLEM SOLVING SKILLS

Summary of the PhD dissertation

Supervisor:

Gyöngyvér Molnár DSc

Professor of Education

Szeged

2018

THE SUBJECT AND STRUCTURE OF THE DISSERTATION

Teamwork offers numerous advantages in the case of solving complex problems: increased coverage of knowledge, skills and ideas becomes available through the members of a team (Finnegan & O'Mahoney, 1996). The potential of teamwork has been widely recognized and used in the workforce in recent decades; therefore, the ability to effectively solve problems in collaboration with others represents a continuously growing value (Brannick & Prince, 1997; Binkley et al., 2012; Cohen & Bailey, 1997; McGrath, 1997; National Research Council, 2011). Consequently, it is a highly significant competence to effectively work in groups and collaboratively create solutions to particular problems in the world today. Thus, there is an urgent necessity of developing Collaborative Problem Solving (CPS) skills in an educational context (Neubert, Mainert, Kretzschmar, & Greiff, 2015). In addition, to be able to diagnose the level of CPS time to time, sufficient instruments are necessary (Csapó, Lőrincz, & Molnár, 2012).

Recently several major research projects have targeted to develop an effective tool for assessing CPS skills. The measurement of the construct was aimed at the Assessment and Teaching of 21st Century Skills (ACT21S) project, by the experts of the Educational Testing Service, furthermore, CPS was the fourth, cross-curricular competence assessed in PISA 2015 (Griffin & Care, 2015; OECD, 2013, 2017; Hao, Liu, von Davier, & Kyllonen, 2017). Yet, numerous basic researches would be still necessary until a valid and reliable, commonly accepted measurement technology emerges as many serious methodological issues are waiting for solutions.

In the first three chapters of the dissertation we give a detailed review of the different perspectives of CPS. The traditions of publishing empirical researches would demand to begin the expose of the theoretical background with the definitions of CPS. Nevertheless, a different line is followed in our review, as we believe, one has to be familiar with the background of its two great components, collaboration and problem solving, to be able to interpret its models and descriptions. We stress the diversity of the traditional collaboration and problem solving assessment methods as one reason for the complexity of measuring CPS. Afterwards the so far existing CPS instruments are getting reviewed.

In the empirical chapters of the dissertation the development and the testing of two different instruments are discussed. First the developmental process of an interactive CPS instrument is presented which enables dyads to collaborate via many innovative communicational options (Chapter 5 and 6). Then a questionnaire for exploring collaborative skills (ColSQ – Collaborative Skills Questionnaire) is described followed by the results of two studies in which we investigated the functioning of the questionnaire (Chapter 7). The last chapter of the dissertation outlines, in our point of view, the probable future perspectives of the research field based on our theoretical considerations and empirical data.

THEORETICAL BACKGROUND

One aspect which makes the assessment of CPS challenging is that the construct itself in its current interpretation is quite new. Therefore, it has no solid empirical background. So far there are four theoretical models to describe CPS (Hesse, Care, Buder, Sassenberg, & Griffin, 2015; Liu, Hao, von Davier, Kyllonen, & Zapata-Rivera, 2015; OECD, 2013; O'Neil, Chuang, & Chung, 2003; lásd Pásztor-Kovács, 2015; Pásztor-Kovács, Pásztor, & Molnár, 2018a, 2018b). There is one core similarity in all the four models: they contain two major elements representing a social or collaborative and a cognitive or problem solving aspect of the construct. However, they name different skills and subskills tied to the social and cognitive components, and these subskills also differ in their arrangement; they are ordered in a hierarchy as well as in a matrix. There was no empirical study which could have validated any of the four models so none of them can be considered to cover the structure of CPS with full certainty. In the lack of empirical data, the quality of the relation between the two major components is also unclear. The matrix models indicate a tight connection, while the hierarchical models do not signify any relation between them. Consequently, the question arises whether there is a connection between the two components, and if so, in what way there is and how strong it is. If we, for instance, assume that general intelligence predicts both components, it is also reasonable to hypothesize that there is at least a minor correlation between them.

Several researches have shown the connection between problem solving skills and general intelligence (pl.: Greiff et al., 2013; Stadler, Becker, Gödker, Leutner, & Greiff, 2015; Wüstenberg, Greiff, & Funke, 2012). However, we have not found any research exploring the relation between collaborative skills and general intelligence. Nevertheless, the relationship between emotional intelligence and collaborative skills were the focus of plenty of researches (pl.: Feyerherm, & Rice, 2002; Goleman, 1995, 1998; Jordan & Troth, 2004; Prati, Douglas, Ferris, Ammeter, & Buckley, 2003; Salovey & Mayer, 1990). These studies found a strong connection between the two constructs. If we accept the theory that emotional intelligence is independent of those mental abilities which are assessed by traditional intelligence tests [in Gardner's (1983) multiple intelligence model for example these are the elements of linguistic, logical-mathematical and visual-spatial intelligence], we cannot have the assumption that the collaborative component, similarly to problem solving is correlated to general intelligence. Thus we cannot hypothesize even a small correlation between the collaboration and problem solving components.

The completely different assessment traditions of the two components raise another barrier to the creation of CPS tests. The measurement of problem solving skills by objective tests relies on a stable empirical basis. The assessment of collaboration processes on the other hand, rest on qualitative tools like natural observation, interview, self-reported questionnaires and peer-rated scales. We are not aware of such an objective test which would assess one's collaborative skills based on his/her achievement in a given group. This means the measurement of the collaborative component with objective methods is itself problematic. The development of such an instrument which gives feedback about the unified construct on individual level relying on the results of a single objective test seems extremely challenging (Pásztor-Kovács, 2013a, 2016a; Pásztor-Kovács, Magyar, Hülber, Pásztor, & Tongori, 2013; Pásztor-Kovács, Pásztor, & Molnár, 2018a, 2018c).

The target of collaborative assessments tends to be the whole team instead of the individuals. Even if group members' behaviours are in the focus of a given study, it is not an aim to evaluate a member's collaborative skills in global (Greiff, 2012; Pásztor-Kovács, 2013c). A possible reason for the lack of research on the individual level in group assessments is the complexity of isolating one's achievement from one's peers in a group, as group members strongly depend on each other's characteristics in a teamwork situation. To create a clear picture of an individual's CPS profile, the ideal procedure would actually be to test him/her in multiple groups with different compositions and thus make it possible to manifest his/her CPS skills with team members with highly different skill sets (Hao et al., 2017; Rosen, 2017).

In the case of summative assessments, where comparability is a core expectation and time constraints are also frequent, this "mapping out" method of CPS skills is clearly not feasible. To produce comparable data, test takers should receive the exact same stimuli within the test, meaning everyone should solve the same problems in teams made up of members with similar characteristics. This design may seem impossible to achieve at first sight; technology, however, offers a creative and so far unique solution: the application of computer agents as collaborators. In a technology-based assessment environment where the collaborating peer is not another person but a conversational agent, it becomes possible to develop a standardized test environment, as agents can generate their reactions from the same pre-programmed set of responses to every test taker (Krkovic, Pásztor-Kovács, Molnár, & Greiff, 2014; OECD, 2013).

There has only been one example so far of a standardized assessment of CPS skills that uses computer agents, namely, the PISA 2015 summative assessment. Nevertheless, the OECD decision on Collaborative Problem Solving measurement in PISA 2015 prompted several significant studies on creating pioneer instruments of CPS with embedded computer agents (Krkovic, Wüstenberg, & Greiff, 2016; Hao et al., 2017; Rosen & Foltz, 2014). There have been some concerns about the aforementioned so called Human-Agent (H–A) design. Proponents of the Human–Human (H–H) assessment line question the validity of results produced by H–A instruments. They point out the obvious differences between human and computer agents and therefore the unrealistic quality of the test environments (Graesser, Forsyth, & Foltz, 2017; Krkovic, Mustafic, Wüstenberg & Greiff, 2018; Scoular, Care, & Hesse, 2017; Rosen, 2017). Indeed, it is completely unexpected for a computer agent to accurately imitate all human behaviours, such as displaying the wide range of emotions or the sometimes rather irrational thinking that humans are known for. Nevertheless, the option of creating a standardized test environment and thus ensuring comparability is such a crucial advantage of H–A designs in summative assessments that we believe that commonly accepted, reliable technology should rely on a H–A approach as it is developed. In our view, the main question is in what ways we can maximize the validity level of H–A designs.

Communication between collaborators contains core information about the problem-solving process and participants' CPS skills, so it is a key issue in what way it is realized and what solutions can be found to evaluate it (Hesse et al., 2015). There are many methodological questions to consider when we decide on the communication design of a CPS instrument. Technology offers a number of advantages for the field of educational assessment (Csapó, Ainley, Bennett, Latour, & Law, 2012). Automated coding, for example, is such a practical feature of technology that it would be extremely resource- and time-consuming not to take advantage of it in large-scale assessments. In the case of CPS, everyday educational practice would also require instruments which generate results that can be coded automatically, as teachers are not necessarily experts on methods of analysing human discourse.

In the case of open-ended communication, which is without doubt the most valid way of exchanging ideas, it is again quite challenging, however, to exploit this potential as content analysis, the traditional method for analysing interactions at the current stage of technology, cannot be implemented with the full elimination of human rating (Care, Griffin, Scoular, Awwal, & Zoanetti, 2015). The solution to handling the complicated case of automated coding was to eliminate the option of conversing freely in the rest of the CPS instruments. More specifically, group members can only talk by exchanging a set of pre-defined messages, which are previously assigned to different skills, so automated data coding can be developed and implemented with this pre-assignment.

What we have learned so far from recent researches was that pre-defined message exchange could be an efficient way to interact with the aim of problem solving (Chung, O'Neil, & Herl, 1999; Hsieh & O'Neil, 2002; Krkovic et al., 2016; OECD, 2017; O'Neil, Chung, & Brown, 1997; Rosen & Foltz, 2014). However, besides its effectiveness for problem solving and automated coding, pre-defined message exchange also proved to have its limits, as it may lead to frustration; participants may be disturbed at not being able to express themselves if the messages fail to cover every possible scenario for talk (Krkovic et al., 2014; Pásztor-Kovács, 2016a, 2016b, 2017a; Pásztor-Kovács et al., 2018a, 2018b). To conclude, one of the main tasks of this research area at this point is to maximize the flexibility and the convenient use of these restricted communication options and thus increase the validity level of the interactions realized in a restricted way.

RESEARCH AIMS

Our ideas about the necessary developmental stages of such a CPS instrument which can ensure a standard test environment and automated data coding in the most valid way can be summarized in this process model:

(1) the development should be started by setting up the H–H version of the instrument, which permits open-ended discussion;

(2) pre-defined messages and further restricted communication options should be based on an analysis of data gathered via the H–H instrument, which permits open-ended discussions;

(3) after eliminating free chat, restricted communication options, including pre-defined messages, should be tested in several further H–H tests;

(4) if the restricted ways of communicating can be considered well-established, the response set of the computer agent to be embedded should be strictly built on the H–H interactions that emerge through restricted communication options;

(5) the automated coding scheme in the H–A instrument should also be based on an analysis of students' interactions using the restricted communication options in the H–H version.

The long-term aim of our research is to create a H–A CPS instrument suitable from the age of 13 by proceeding along the stages we defined as ideal for maximizing the validity level of the assessment tool. Till the submission of the dissertation we have reached the third stage, namely the creation of such a Human-Human instrument within which students can interact by restricted communication options only. The development of a H–H instrument should not only be considered as a necessary step in the project of creating a H–A assessment tool. Our aim is to make the H–H version a valuable tool for formative CPS assessments in its own right (Pásztor-Kovács, 2013b, 2013d).

By the first version of our CPS instrument we have observed the collaboration of four-member teams solving analytical, decision-making type of problems with the condition of using free chat (Magyar, Pásztor, Pásztor-Kovács, Pluhár, & Molnár, 2015; Pásztor-Kovács, 2014a, 2014b, 2016a, 2016b). Following the stricter definition of collaboration, the tasks of the test were not interdependent. The results of the trial provided important conclusions. Relying on them some major changes have been implemented on the instrument.

By using the revised version of the assessment tool, we investigated the collaboration of dyads solving interdependent, interactive problems based on the MicroDYN model. In this new version many restricted communication options were developed in line with our long-term research goals (Pásztor-Kovács, 2017b; Pásztor-Kovács et al., 2018a, 2018b, 2018c). We created the content of the pre-defined messages based on the results of our in-house studies. Several innovative options were ensured for the interaction, we developed visual channels besides the verbal ones to exchange information so that the flexibility and by that the validity of restricted communication could increase.

In the second part of our research a self-reported questionnaire was developed to explore the collaborative component of CPS (Pásztor-Kovács & Pásztor, 2017). The aim of the development of ColSQ was twofold. For future external validation of the CPS assessment tool suitable instruments are necessary. In case of the problem solving component, this is easily achievable with the application of the individual test based on the MicroDYN problems. For the exploration of different aspects of the collaborative component many methods are available, mostly qualitative ones. However, in most cases these instruments do not observe the individuals but the whole groups. Also, if the assessment tool measures on individual level, it is a peer or an expert who gives the evaluation. This is again not in line with our research plan, which is to create such a design where students first solve a problem solving and a collaborative test, then the results of the thirdly registered CPS test can be compared to the first two ones.

We have found two instruments in the literature which, on one hand, explore collaborative skills in global (and not referring to a given teamwork task post hoc), and which are, on the other hand, self-reported and by that fit to our research plan for a validation study (Cumming et al., 2015; Kasik, 2013). Nevertheless, these instruments in their content hardly cover the collaborative component of the CPS models. We decided to develop a new instrument, a self-reported questionnaire based on one of the CPS models which measures collaborative skills in general and not referring to one case.

The second aim of developing the questionnaire was to learn more about the structure of CPS. None of the CPS models have been tested empirically so far so we also intended to fill this gap with our questionnaire which was based on the collaborative component of the ATC21S model (Hesse et al., 2015).

Within a large-scale online assessment, we had the opportunity to deepen our knowledge about the construct of CPS. Based on the results of our questionnaire, a problem solving test with MicroDYN problems and an inductive reasoning test, we could explore the relations between the collaboration and the problem solving component and additionally their relation to inductive reasoning skills.

RESEARCH QUESTIONS AND HYPOTHESES

In our research we were looking for answers to the following research questions:

RQ1: On which phase (the knowledge acquisition or the knowledge application) of the MicroDYN problems do students achieve better in the collaborative problem solving test?

RQ 2: Is it possible to solve the collaborative problem solving test with the application of restricted communication options only?

RQ 3: Does the Collaborative Skills Questionnaire prove to be reliable?

RQ 4: Does the theoretical model on which the questionnaire is based on appear behind the factor structure of the questionnaire?

RQ 5: Which results do the scores of the inductive reasoning test (as an indicator of general intelligence) show a closer connection with: the results of the Collaborative Skills Questionnaire or the problem solving test?

RQ 6: In which direction is there and how strong is the relationship between the results of the Collaborative Skills Questionnaire and the problem solving test?

The following hypotheses were phrased connected to research questions, based on the research results presented in the literature review part of the dissertation:

H_1: Students (similarly to the pattern experienced in the individual problem solving test) achieve better on the knowledge acquisition than on the knowledge application phases of the problems in the collaborative problem solving test.

H_2: It is possible to solve the collaborative problem solving test with the application of restricted communication options only.

H_3: The Collaborative Skills Questionnaire proves to be reliable.

H_4: The theoretical model on which the questionnaire is based on appears behind the factor structure of the questionnaire.

H_5: The scores of the inductive reasoning test (as an indicator of general intelligence) show a closer connection to the results of the problem solving test than to the results of the Collaborative Skills Questionnaire.

H_6: There is a weak positive correlation at most between the results of the Collaborative Skills Questionnaire and the problem solving test.

DEVELOPMENT OF THE FIRST VERSION OF OUR COLLABORATIVE PROBLEM SOLVING INSTRUMENT AND THE RESULTS OF ITS TRIAL

In the first version of the CPS instrument in the eDia system analytical, decision making problems were applied of which development was inspired by the PISA 2003 problem solving assessment tasks (Molnár, 2006). Students had to understand the problem situation, consider the possible outputs of the situation and the limiting aspects, finally, choose the best one from the offered solutions. The complexity of the problems could have been increased by the growing number of the limiting aspects and the growing difficulty of identifying these aspects. We developed domain-general problems to control the variable of prior knowledge in team members' achievements. The problem situations were such scenarios which could happen in students' daily lives in the target age group (e.g. choosing a birthday present, a summer camp or a school with the consideration of given parameters). Following the stricter definition of collaboration, the problems were not independent, students had the same pieces of information by getting the same instruction.

In the trial of the instrument our aim was to find out its reliability and the time required to solve the test, and also, to learn about the attitudes of the test takers toward the instrument. Furthermore, in line with our long-term research aims, we were focusing on the questions whether the four-member team design is eligible, if a discussion of the problems emerges in the lack of interdependence, and whether decision making problems are appropriate for our future research plans.

71 BSc students were participating in the study. They were collaborating in seventeen groups which had three, four or five members. At the beginning of the test students learned about the way they could give solutions to the problems jointly and they could try out the chat function. After a pilot task four problems were provided containing 11 items. At the end of

the test we asked students to answer three close-ended questions offering a five-point scale and also one open-ended question to give their opinions about the instrument.

The reliability index of the 11 items was Cronbach's α=.78. The average score on the test was 8.91 (SD=2.50) out of 11, the four problems required approximately 30 minutes to solve (Mmin=29.99; SD=5.96). Significant positive correlation was found between the first and the second (r=.66, p<.01), furthermore between the second and the third problems (r=.47, p=.05). The fourth problem did not correlate with any other ones. More than half of the students (59.7%) reported to enjoy the joint problem solving. Most of them found the problems moderately difficult (70.1%) and evaluated team work successful (86.5%). 37 participants gave their opinion about the test in the last, open-ended question. The answers could be organized in six categories. The biggest category contained responses expressing general attitudes about the test. 93% of these answers mirrored positive opinions.

From psychometrical aspects the test was basically satisfying, so were the attitudes toward the instrument. However, from the viewpoint of our long-term research aims, the results of the study indicated the need of some major changes. For example, in the answers given to the open-ended question it was clearly outlined that the exchange of messages was slow. Also, it was difficult to follow and to search back what kind of messages and particularly, solution plans students had sent to each other. This feedback made us conclude that although from the perspective of generalizability we would find teams with bigger number of members working together to be optimal, written chat hardly fits to this condition.

The second conclusion was connected to the design of exposing the problems. We had to face the fact that the presentation of the complete set of information was not a good choice considering our long-term plans because of the lack of interaction. The results indicated that we have to accept the wider definition of collaboration in the future, which allows the condition of interdependence. This condition can ensure the interaction of team members, which is substantial for evaluating students' CPS skills.

The analysis of the interactions made us clear that the type of the problems also needs to be changed. Although there were some objective criterions to consider, students tended to make decisions based on subjective elements like their own experience and opinions. This might have been caused by the familiar, lifelike context. We realized that we have to present such problems which are completely exact, there are no subjective elements related to their solutions, and also they are domain-general.

Furthermore, we found out that the chosen problem type cannot be solved by restricted communication options. Based on the messages students sent to each other we failed to develop such a pre-defined message set which could fairly cover everything that the participants would potentially want to express. We recognized that the possible outcomes of the interactions which are induced by the decision making problems are too many and too uncertain to be suitable for solving them by restricted communication. The conclusion we made was that we needed such problem types in which the problem space is relatively small, the process of the problem solving is well-predictable, so the development of a pre-defined message set connected to it is also easier to realize.

DEVELOPMENT OF THE SECOND VERSION OF THE CPS INSTRUMENT AND THE RESULTS OF ITS TRIAL

Based on the conclusions of the prior study we had implemented, some great changes connected to the instrument. The improved version was suitable for the collaboration of dyads. For the exact, domain-general problems with a relatively small problem space we looked for, we found the interactive MicroDYN problems to be the most appropriate. The so called MicroDYN model was developed by the researchers of the University of Heidelberg.

The Hungarian adaptation of the MicroDYN problems, which assess a knowledge acquisition and a knowledge application phase separately, was tested several times via the eDia platform and proved to be effective (Csapó & Molnár, 2017; Greiff, Wüstenberg, & Funke, 2012; Molnár, 2016a, 2016b, 2017; Molnár & Csapó, 2018; Molnár & Pásztor-Kovács, 2015; Wüstenberg et al., 2012).

Problems were interdependent. To ensure that, in their first, knowledge acquisition phase we created a jigsaw design in which team members have different pieces of information. Half of the information was available only for one member of the dyad, the other half was available for the other member. For creating the solution, the two students had to collaborate. In the knowledge application phase the jigsaw design was not found appropriate for maintaining interdependence. Instead, students had to build consensus on how to manipulate the different variables to come to the solution.

This instrument version was already suitable for developing restricted communication options in it for automated coding in the future. We were trying to create such restricted ways for interacting which could ensure a flexible communication even in the lack of typing in messages freely, which was convenient to use, fairly covered the potential messages of the students and by all of these features increased the validity of the assessment tool. To be able to fulfil this aim we based the content of the pre-defined messages on the results of our prior in house studies in which it was allowed to form messages freely. Furthermore, we developed several new, innovative ways besides the pre-defined messages for restricted communication, for example, options for visual information exchange.

This new version with its innovative elements was tried out in two small-scale studies to find out how well students can handle the new, unorthodox test environment and communicational options. We wanted to explore how challenging it is to solve the problems; whether the innovative options besides pre-defined messages represent a sufficient way for communication; what patterns the participants use for the different options. Furthermore, it was again intended to investigate the test takers' attitudes toward the instrument: whether they find it user-friendly; whether they believe communication is flexible enough via these channels; and how difficult they find understanding the use and later on utilizing the restricted communication options.

In the first study, besides the restricted communication options, we also provided the chance of typing in messages freely, however, we asked the participating students (N=10) to use this option only if they cannot express themselves in any other way. Options for restricted communication managed to substitute free communication by more than 85%. This promising result encouraged us to fully eliminate the chance for free chat in the next study. The freely typed messages in the first study represented a rich input for further improving the list of pre-defined messages.

Eight girls and two boys from the seventh and eighth grades (Mage=13.3; SD=.5) participated in the second study, which was carried out via the eDia platform in the ICT lab of the primary school which the participants attended. The data collection required two periods out of the school day. First, a three-task individual problem-solving test was administered containing MicroDYN problems to make students familiar with their context. Then a four-problem CPS test was provided for the dyads. There was no time limit to solve the problems; however, students were asked to stop working on the test after 50 minutes. In the last section of the data collection, we administered a questionnaire to investigate students' attitudes toward the instrument. It contained two open-ended and eight close-ended questions, out of which seven offered a five-point scale for the answer and one was a dichotomous, yes-no item.

None of the pairs managed to complete the whole test, three of them provided their last solutions to the knowledge acquisition phase of the third problem, and the remaining two only reached the knowledge application phase of the second problem. The knowledge acquisition phases required more interactions and more time to discuss and proved to be much more difficult to complete than the knowledge application phases. In the first two problems, the knowledge acquisition phase basically produced a floor effect, while the knowledge application phase produced a ceiling effect. Our first hypothesis (*H1: Students achieve better on the knowledge acquisition than on the knowledge application phases of the problems in the collaborative problem solving test)* was not supported; the knowledge acquisition phase in every aspect was proved to be more difficult.

The contradictory tendency with the individual problem solving tests is supposedly caused by the different designs in the two phases. The jigsaw design and the broader spectrum of restricted communication options in the knowledge acquisition phase might have posed students a bigger challenge. The finding that students' achievements display a divergent trend in knowledge acquisition and knowledge application on the individual and the collaborative problem solving tests hints at a promising hypothesis: the two tests measure distinct constructs. Nevertheless, the low sample size does not allow us to make strong claims, so further researches need to be conducted to investigate this assumption.

The test was quite time-consuming; one student even commented that more time would have been necessary for him/her to become effective. To make sure that students experience some success during the test, one option is to prolong it. This is not necessarily the best option, however, as it is not always possible to devote more than two periods out of the school day to the test.

Another, more feasible option is to replace the problems with easier ones, for example, problems with two input variables and only one output variable. This solution is also supported from the perspective of the low achievement on the knowledge acquisition tasks. In addition, a time limit might be introduced to these task types to eliminate the ceiling effect on the knowledge application phases. By providing easier problems but also a time limit in the knowledge application phases, we could possibly bring the level of challenge in the two phases closer to the optimal.

The second hypothesis of the dissertation (*H_2: It is possible to solve the collaborative problem solving test with the application of restricted communication options only*) was confirmed. The results indicate that the restricted communication ways of the instrument create a suitable basis for solving the problems collaboratively. Our innovative options were found to be usable and rather popular. Options for visual information exchange were used more frequently in both the knowledge acquisition and application phases of the test than options for verbal information exchange and were also used more and more frequently as students went along the test.

The results from the questionnaire led us to conclude that most of the students enjoyed the testing but felt that their joint work was unsuccessful. Responses to the question "Was it more difficult to solve the problems jointly than alone?" covered a broader spectrum, showing that the new design with the elimination of free chat posed a greater challenge for some students. This tendency is also reflected in the rest of the questions: most participants chose the middle of the scale to express how easy they found it to understand and implement the innovative ways of communicating. Although they tended to answer that they were able to express themselves easily through the messages, they consistently noted that they missed being able to type in messages on their own.

Eight students out of ten said *yes* to the dichotomous question "Was there a message you would have needed but didn't find in pressing the Message button?". We also provided the chance to list any lacking messages. Considering the rate of free messages and restricted ways of communicating used in the prior trial of the test, this result was somewhat unexpected. Either restricted communication proved to be a greater challenge for this sample or the small amount of free communication realized in the previous study was more crucial for students to express themselves than we thought.

To reduce the frustration shown with the lack of free chat, there is great potential in further developing the pre-defined message set. Messages suggested by the students should be added to the current set. The layout should also be reconsidered, as three students recommended such content which was already available. A different arrangement may improve the perspicacity of the message set and help students keep all the messages in mind.

Once we reach a point where students report that they barely miss free communication within the test, the next problem is how to create a coding scheme. In this Human–Human version, we cannot ensure the complete automation of the data coding; however, limiting the interaction space itself makes evaluation much simpler for the H–H instrument. Based on a log file analysis, several indicators can be developed to describe students' CPS skills.

DEVELOPMENT AND TRIAL OF THE COLLABORATIVE SKILLS QUESTIONNAIRE

The self-reported questionnaire was based on the collaborative component of the hierarchical model connected to the ATS21S project (Hesse et al., 2015). We intended to cover 9 social subskills (action, interaction, task completion, adaptive responsiveness, audience awareness, negotiation, self-evaluation, transactive memory, responsibility initiative) assigned to the three skills of participation, perspective taking and social regulation. The questionnaire was pilot tested in a paper-based study (N=96). Based on the results of this data collection we improved the instrument further. In the next, large-scale online measurement the improved version was applied already with 36 items. Four-four items were assessing 9 subskills connected to 3 subscales. 15 out of the 36 items were reverse-scored. Students had to rate on a five-point scale how much the given statements described them.

In the large-scale assessment 1613 8th grade students (M_{age}=14.60 years, SD=.52) from 65 schools gave answers to the online questionnaire via the eDia platform all around the country. 49.7% of the students were boys, 49.6% were girls, .7% of them did not provide data. 1044 students of the sample solved a problem solving test including 10 MicroDYN problems, 1233 students a 45-item inductive test (Pásztor, 2016), too.

The third hypothesis of the dissertation (*H3*: *The Collaborative Skills Questionnaire proves to be reliable*) was supported. Based on the results of the confirmatory factor analysis the questionnaire was reduced to 17 items. The reliability index of this reduced scale (Cronbach's α=.91) and its subscales (Cronbach's α=.69-.85) were acceptable, consequently, the instrument could be considered reliable.

The fourth hypothesis (*H4*: *The theoretical model on which the questionnaire is based on appears behind the factor structure of the questionnaire*) was also confirmed. We were the first ones who empirically tested a CPS model, its collaborative component more precisely. Our results underpinned the idea of the ACT21S expert group about the structure of collaborative skills. The three skills clearly appeared behind the collaborative component, and the subskills were also represented, not on factor level but at least on item level. The only exception was the transactive memory subskill.

Out of the one, three and nine dimensional models the three dimensional, 17-item model was proved to be the most adequate for describing the construct (χ^2=574.08; df=115; p<.01; CFI=.920; TLI=.905; RMSEA=.068). However, with the exclusion of one item the one dimensional model showed the same level of fit as the three dimensional model. Therefore, on one hand, it is necessary to investigate the case further; on the other hand, it is not required by all means to reject one of the models. Both the complete 17-item model and the three subscales had acceptable indices, so it might be worth observing the results of all the four scales after a data collection to get a more detailed profile of one's collaborative skills.

Our fifth hypothesis (*H5*: *The scores of the inductive reasoning test show a closer connection to the results of the problem solving test than to the results of the Collaborative Skills Questionnaire*) was again supported. The inductive reasoning, which can be considered as the indicator of general intelligence, showed significant positive correlation with both the collaboration (r=.19, p<.01) and the problem solving (r=.55, p<.01) component (N=623). However, the correlation between the problem solving and the inductive reasoning test results were significantly stronger (Z_{sz}=5.75 >1.6). Furthermore, the results of the inductive test were better predicting the problem solving test results [β=.55; t(622)=-5.62, p <.01], than the ColSQ results [β=.19; t(622)=31.56; p <.01]. These findings show that general intelligence, unlike its relation to problem solving, does not influence the collaborative achievement firmly, so it is not necessarily required to have a high level of general intelligence for being a great collaborator.

The sixth hypothesis (*H6*: *There is a weak positive correlation at most between the results of the Collaborative Skills Questionnaire and the problem solving test*) was also confirmed. We found a weak positive correlation between the collaboration and the problem solving component (r=.17, p=.01; N=623), which technically disappeared (r=.08) with partialling out the inductive reasoning test variable. Consequently, those theoretical models which represent the two major elements of CPS with no relation between them seem right: our results indicate that the components are practically independent of each other.

SUMMARY

The aim of the dissertation was to describe the complex construct of Collaborative Problem Solving, to review the challenges of defining and assessing CPS skills, and to present our own two instruments for exploring CPS and its collaborative component. Being quite a new research field, the studies presented in the dissertation can be considered as basic researches. The analysis and synthesis of the literature is also in this category: we gave a pioneer review of the methodological challenges and created a five-stage process model for the sake of developing valid CPS assessment tools.

Some of our researches was based on this process model, aiming the final stage of developing a Human-Agent CPS instrument in a long-term, which has its own Human-Human version, creates a standardized test environment and which is suitable for automatic data coding. The developmental process both from researcher and programmer sides proved to be a very complex task with constantly new obstacles rising up. That is why by the time of submitting the dissertation we have reached only the third stage. The third stage represents such a Human-Human CPS instrument which is not only a pre-version of the H-A version but also a valuable tool itself for creating a detailed profile of students' CPS skills.

In the latest version of our CPS instrument, primarily again, we have transformed MicroDYN problems suitable for Human-Human collaboration. Students interact via restricted communication options exclusively. To make this kind of interaction convenient we have developed several new, innovative ways besides the pre-defined messages for restricted communication, for example options for visual information exchange.

In the trial of the instrument students were able solve the problems collaboratively without typing in messages freely, which result means, on one hand, a positive feedback connected to the restricted communication options within our assessment tool. On the other hand, it provides evidence and pioneer example for the assumption that besides the exchange of pre-defined messages, which has been the only option so far to ensure automated data coding, it is worth exploring and implementing other alternatives to make restricted communication more flexible.

A further important output of our research is a 17-item questionnaire based on one of the CPS models which reliably measures students' collaborative skills. It was the first time that a theoretical model of CPS, more specifically its collaborative component has been empirically tested. By confirmatory factor analysis we identified the elements of the ATC21S model behind our items.

We were again the first ones to explore whether there is a relation, and if so, how strong it is between the collaboration and the problem solving component of CPS. The results of our study, in which our collaborative questionnaire, furthermore, a MicroDYN-type problem solving test and an inductive reasoning test were administered, indicate that the two components are practically independent of each other. Considering these findings, we believe it is necessary to rethink whether the new concept of the unified CPS construct is worthy to build upon. The diagnosis and the development of both the collaborative and problem solving skills is a high priority task in education, nevertheless, we suggest assessing these two elements isolated, like we did before the idea of the PISA 2015 CPS assessment.

REFERENCES

Binkley, M., Erstad, O., Herman, J., Raizen, S., Martin, R., Miller-Ricci, M., & Rumble, M. (2012). Defining twenty-first century skills. In P. Griffin, B. McGaw, & E. Care (Eds.), *Assessment & teaching of 21st century skills* (pp. 17–66). New York: Springer.

Brannick, M. T., & Prince, C. (1997). An overview of team performance measurement. In M. T. Brannick, E. Salas, & C. Prince (Eds.), *Team performance assessment and measurement: Theory methods and applications* (pp. 3–16). Mahwah, NJ: Lawrence Erlbaum Associates.

Care, E., Griffin, P., Scoular, C., Awwal, N., & Zoanetti, N. (2015). Collaborative problem solving tasks. In P. Griffin and E. Care (Eds.), *Assessment and teaching of 21st century skills: Methods and approach.* (pp. 85-104). Dordrecht: Springer.

Cohen, S. G., & Bailey, D. E. (1997). What makes teams work: Group effectiveness research from the shop floor to the executive suite. *Journal of Management, 23*(3), 239–290.

Chung, G. K. W. K., O'Neil, H. F., Jr., & Herl, H. E. (1999). The use of computer-based collaborative knowledge mapping to measure team processes and team outcomes. *Computers in Human Behavior, 15*(3–4), 463–494.

Cumming, J., Woodcock, C., Cooley, S. J., Holland, M. J., & Burns, V. E. (2015). Development and validation of the groupwork skills questionnaire (GSQ) for higher education. *Assessment & Evaluation in Higher Education, 40*(7), 988–1001.

Csapó, B., Ainley, J., Bennett, R., Latour, T., & Law, N. (2012a). Technological issues of computer-based assessment of 21st century skills. In P. Griffin, B. McGaw, & E. Care (Eds.), *Assessment & teaching of 21st century skills* (pp. 143–230). New York: Springer.

Csapó, B., Lőrincz, A., & Molnár, G. (2012). Innovative assessment technologies in educational games designed for young students. In D. Ifenthaler, D. Eseryel, & X. Ge (Eds.), *Assessment in game-based learning: foundations, innovations, and perspectives* (pp. 235–254). New York: Springer.

Csapó, B., & Molnár G. (2017). Potential for assessing dynamic problem-solving at the beginning of higher education studies. *Frontiers in Psychology, 8*:2022.

Feyerherm, A. E., & Rice, C. L. (2002). Emotional intelligence and team performance: The good, the bad and the ugly. *The International Journal of Organizational Analysis*, *10*(4), 343-362.

Finnegan, P., & O'Mahoney, L. (1996). Group problem solving and decision making: An investigation of the process and supporting technology. *Journal of Information Technology, 11*(3), 211–221.

Gardner, H. (1983). *Frames of mind: the theory of multiple intelligences.* New York: Basic Book.

Goleman, D. (1995). *Emotional intelligence*. New York: Bantam Books.

Goleman, D. (1998). *Working with emotional intelligence*. New York: Bantam.

Graesser, A. C., Forsyth, C. M., & Foltz, P. (2017). Assessing conversation quality, reasoning, and problem solving with computer agents. In B. Csapó & J. Funke (Eds.), *The* *nature of problem solving: Using research to inspire 21st century learning* (pp. 245–261). Paris: OECD Publishing.

Greiff, S. (2012). From interactive to collaborative problem solving: Current issues in the Programme for International Student Assessment. *Review of Psychology*, *19*(2), 111-121.

Greiff, S., Wüstenberg, S., & Funke, J. (2012). Dynamic problem solving: A new assessment perspective. *Applied Psychological Measurement, 36*(3), 189–213.

Greiff, S., Wüstenberg, S., Molnár, G., Fischer, A., Funke, J., & Csapó, B. (2013). Complex problem solving in educational contexts – Something beyond g: Concept, assessment, measurement invariance, and construct validity. *Journal of Educational Psychology, 105*(2), 364–379.

Griffin, P., & Care, E. (Eds.). (2015). *Assessment & teaching of 21st century skills*. *Methods and approach.* Dordrecht: Springer.

Hao, J., Liu, L., von Davier, A. A., & Kyllonen, P. C. (2017). Initial steps towards a standardized assessment for collaborative problem solving (CPS): Practical challenges and strategies. In A. von Davier, M. Zhu, & P. Kyllonen (Eds.), *Innovative Assessment of Collaboration* (pp. 135–156). Springer International Publishing.

Hesse, F., Care, E., Buder, J., Sassenberg, K. & Griffin, P. (2015). A framework for teachable collaborative problem solving skills. In P. Griffin, & E. Care (Eds.), *Assessment & Teaching of 21st Century Skills*. *Methods and Approach* (pp. 37–56). Dordrecht: Springer.

Hsieh, I.-L., & O'Neil, H. F., Jr. (2002). Types of feedback in a computer-based collaborative problem solving group task. *Computers in Human Behavior, 18*(6), 699–715.

Jordan, P. J., & Troth, A. C. (2004). Managing emotions during team problem solving: Emotional intelligence and conflict resolution. *Human Performance*, *17*(2), 195-218.

Kasik, L. (2013). *Együttműködés kérdőív.* SZTE BTK Neveléstudományi Intézet.

Krkovic, K., Mustafic, M., Wüstenberg, S., & Greiff, S. (2018). Shifts in the assessment of problem solving. In E. Care, P. Griffin, & M. Wilson (Eds.), *Assessment and Teaching of 21st Century Skills* (pp. 55-73). Cham: Springer.

Krkovic, K., Wüstenberg, S., & Greiff, S. (2016). Assessing collaborative behavior in Students. *European Journal of Psychological Assessment, 32*(1), 52–60.

Liu, L., Hao, J., von Davier, A., Kyllonen, P., & Zapata-Rivera, D. (2016). A tough nut to crack: Measuring collaborative problem solving. In Y. Rosen, S. Ferrara, & M. Mosharraf (Eds.), *Handbook of research on technology tools for real-life skill development* (pp. 344–359). Hershey, PA: IGI Global.

McGrath, J. E. (1997). Small group research, that once and future field: An interpretation of the past with an eye to the future. *Group dynamics: Theory, research, and practice, 1*(1), 7–27.

Molnár, G. (2006). A tudáskoncepció változása és annak megjelenése a PISA 2003 vizsgálat komplex problémamegoldás moduljában. *Új Pedagógiai Szemle, 56*(1), 75–86.

Molnár, G. (2016a). A dinamikus problémamegoldó képesség mint a tudás elsajátításának és alkalmazásának képessége. *Iskolakultúra, 26*(5), 3–16.

Molnár, G. (2016b). Interaktív problémamegoldó környezetben alkalmazott felfedező stratégiák hatékonysága és azok változása: logfájlelemzések. *Magyar Pedagógia, 116*(4), 427–453.

Molnár, G. (2017). A problémamegoldó és tanulási stratégiák változása 11 és 19 éves kor között: logfile elemzések. *Magyar Pedagógia, 117*(2), 221–238.

Molnár, G., & Csapó, B. (2018b). The efficacy and development of students' problem-solving strategies during compulsory schooling: Logfile analyses. *Frontiers in Psychology,* 9:302.

Molnár, G., Greiff, S., & Csapó, B. (2013). Inductive reasoning, domain specific and complex problem solving: relations and development. *Thinking skills and Creativity*, *9*(8), 35–45.

National Research Council (2011). *Assessing 21st century skills.* Washington, DC: National Academies Press.

Neubert, J., Mainert, J., Kretzschmar, A., & Greiff, S. (2015). The assessment of 21st century skills in industrial and organizational psychology: Complex and collaborative problem solving. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 8*(2), 238–268.

OECD (2013). *PISA 2015 draft collaborative problem solving assessment framework.* Retrieved from http://www.oecd.org/pisa/pisaproducts/Draft%20PISA%202015%20Collaborative%20Problem%20Solving%20Framework%20.pdf

OECD (2017). *PISA 2015 results (Volume V): Collaborative problem solving*. Paris: OECD.

O'Neil Jr, H. F., Chuang, S., & Chung, G. K. W. K. (2003). Issues in the computer- based assessment of collaborative problem solving. *Assessment in Education, 10*(3), 361–373.

O'Neil Jr, H. F., Chung, G. K. W. K., & Brown, R. (1997). Use of networked simulations as a context to measure team competencies. In H. F. O'Neil Jr (Ed.), *Workforce readiness: Competencies and assessment* (pp. 411–452). Mahwah, NJ: Erlbaum.

Pásztor, A. (2016). *Az induktív gondolkodás technológia alapú mérése és fejlesztése* (Unpublished Doctoral Dissertation). SZTE BTK Neveléstudományi Doktori Iskola, Szeged.

Prati, L., Douglas, C., Ferris, G. R., Ammeter, A. P., & Buckley, M. R. (2003). Emotional intelligence, leadership effectiveness, and team outcomes. *The International Journal of Organizational Analysis*, *11*(1), 21-40.

Rosen, Y. (2017). Assessing students in human-to-agent settings to inform collaborative problem-solving learning. *Journal of Educational Measurement, 54*(1), 36–53.

Rosen, Y., & Foltz, P. (2014). Assessing collaborative problem solving through automated technologies. *Research and Practice in Technology Enhanced Learning, 9*(3), 389–410.

Salovey, P., & Mayer, J. D. (1990). Emotional intelligence. *Imagination, cognition and personality*, *9*(3), 185-211.

Scoular, C., Care, E., & Hesse, F. W. (2017). Designs for operationalizing collaborative problem solving for automated assessment. *Journal of Educational Measurement, 54*(1), 12–35.

Stadler, M., Becker, N., Gödker, M., Leutner, D., & Greiff, S. (2015). Complex problem solving and intelligence: A meta-analysis. *Intelligence*, *53*, 92-101.

Wüstenberg, S., Greiff, S., & Funke, J. (2012). Complex problem solving. More than reasoning? *Intelligence, 40*, 1–14.

Zsolnai, A., & Kasik, L. (2015). Az együttműködő viselkedés és az alapérzelem-felismerés online vizsgálata. In B. Csapó, & A. Zsolnai (Eds.), *Online diagnosztikus mérések az iskola kezdő szakaszában* (pp. 71–95). Budapest: Oktatáskutató és Fejlesztő Intézet.

THE CITED, DISSERTATION-RELATED PUBLICATIONS OF THE DOCTORAL CANDIDATE

Krkovic, K., Pásztor-Kovács, A., Molnár, G., & Greiff, S. (2014). New technologies in psychological assessment: The example of computer-based collaborative problem solving assessment. *International Journal of e-Assessment, 1*(1), online.

Magyar, A., Pásztor, A., Pásztor-Kovács, A., Pluhár, Zs., & Molnár, G. (2015). A 21. században elvárt képességek számítógép alapú mérésének lehetőségei. In Z. Tóth (Ed.). *Új Kutatások a Neveléstudományokban. Oktatás és nevelés - gyakorlat és tudomány (pp.* 230-243). Debreceni Egyetem: MTA Pedagógiai Tudományos Bizottság.

Molnár, G., & Pásztor-Kovács, A. (2015a). A problémamegoldó képesség mérése online tesztkörnyezetben. In A. Zsolnai, & B. Csapó (Eds.), *Online diagnosztikus mérések az iskola kezdő szakaszában* (pp. 341-366). Budapest: Oktatáskutató és Fejlesztő Intézet.

Pásztor-Kovács, A. (2017b, November). Innovatív kommunikációs lehetőségeket és automatikus kiértékelést biztosító kollaboratív problémamegoldó teszt kipróbálásának eredményei. Paper presented at the XVII. Országos Neveléstudományi Konferencia, Nyíregyháza.

Pásztor-Kovács, A. (2017a, April). The effect of the application of pre-defined messages on Collaborative Problem Solving behaviour in computer-based test environment. Paper presented at the 15th Conference on Educational Assessment, Szeged.

Pásztor-Kovács, A. (2016a). A kollaboratív problémamegoldó képesség mérésének elméleti és módszertani megfontolásai: egy pilot kutatás eredményei. *Magyar Pedagógia, 116*(1), 51-72.

Pásztor-Kovács, A. (2016b, November). Az előre definiált üzenetek alkalmazásának hatása a kollaboratív problémamegoldó teljesítményre online tesztkörnyezetben. Paper presented at the XVI. Országos Neveléstudományi Konferencia, Szeged.

Pásztor-Kovács, A. (2015). Kollaboratív problémamegoldó képesség: egy új, integratív elméleti keret. *Iskolakultúra, 25*(2), 3-16.

Pásztor-Kovács, A. (2014b, November). Kollaboratív problémamegoldó képességet vizsgáló harmadikgenerációs mérőeszköz fejlesztése és pilot tesztelése. Paper presented at the XIV. Országos Neveléstudományi Konferencia, Debrecen.

Pásztor-Kovács, A. (2014a, May). Assessing collaborative problem solving online – experiences of a pilot study. Paper presented at the 12th Conference on Educational Assessment, Szeged.

Pásztor-Kovács, A. (2013d, November). A kollaboratív problémamegoldó képesség online vizsgálata – egy innovatív mérőeszköz kidolgozásának lépései. Paper presented at the XIII. Országos Neveléstudományi Konferencia, Eger.

Pásztor-Kovács, A. (2013b, August). Creation of an online assessment tool for collaborative problem solving competence. Poster presented at the EARLI JURE pre-conference. Munich, Germany.

Pásztor-Kovács, A. (2013c, August). Methodological challenges in the assessment of collaborative problem solving. Paper presented at the 15th European Conference for the Research on Learning and Instruction. Munich, Germany.

Pásztor-Kovács, A. (2013a, April). A kollaboratív problémamegoldás mérése. Paper presented at the XI. Pedagógiai Értékelési Konferencia, Szeged.

Pásztor-Kovács, A., Magyar, A., Hülber, L., Pásztor, A., & Tongori, Á. (2013). Áttérés online tesztelésre - a mérés-értékelés új dimenziói. *Iskolakultúra, 23*(11), 86-100.

Pásztor-Kovács, A. & Pásztor, A. (2017, August 29th – September 2nd). Development of the Collaborative Skills Questionnaire (ColSQ) – first results of a pilot study. Paper to be presented at the 17th European Conference for the Research on Learning and Instruction, Tampere, Finland.

Pásztor-Kovács, A., Pásztor, A., & Molnár, G. (2018a). Development of an Online Collaborative Problem Solving Instrument: The Human–Human Version. *Educational Technology Research & Development.* Manuscript submitted for publication.

Pásztor-Kovács, A., Pásztor, A., & Molnár, G. (2018b, April). Innovative restricted communicational options for automated coding in an online Human-Human Collaborative Problem Solving instrument. Paper presented at the 16th Conference on Educational Assessment, Szeged.

Pásztor-Kovács, A., Pásztor, A., & Molnár, G. (2018c). Kollaboratív problémamegoldó képességet vizsgáló dinamikus teszt fejlesztése. *Magyar Pedagógia.* In press.