New technologies in psychological assessment: The example of computer-based collaborative problem solving assessment

Katarina Krkovic and Samuel Greiff, University of Luxembourg and
Anita Pásztor-Kovács and Gyöngyvér Molnár, University of Szeged

Abstract
Computer-based assessment is a relatively new but exponentially growing field in psychological assessment. Its advantages are numerous – flexibility of application, the opportunity to use video and audio material, the construction of dynamic tasks, and the availability of log-file data that allow us to capture the entire process of solving one task. Since collaborative problem solving is a skill that is becoming increasingly important in the 21st century due to the shift in requirements on labour market and in education, constructing an appropriate assessment tool for it is a high priority in current research. The computer-based assessment of collaborative problem solving appears to be a logical solution for capturing the entire process of collaboration and problem solving. However, the construction of computer-based collaborative problem solving assessment comes with many questions that need to be discussed – How can we create the assessment environment for collaborative problem solving assessment? Can we use computer agents instead of real humans to simulate the collaboration and how does the perception of social presence of these agents influences test-takers' behaviour? How can we structure and assess communication in collaborative problem solving? How can we properly use contextual data obtained from log-files? These and other questions are addressed in this paper.

Introduction – the possibilities and constraints of computer-based assessment

Fascinating developments in informational technology, new software, the Internet, and social networks have been causing tremendous changes in many life sectors, from industry, entertainment, and transportation, to work and education. Many disciplines have changed significantly due to computerisation, and they benefit from it in different ways. For instance, in medicine, many medical devices such as pacemakers or prosthetics function with the help of a computer. Research in the natural sciences – physics, mathematics, chemistry, or astronomy – is nowadays not imaginable without computer assistance. In psychology and in the educational sciences, computerisation is still a work in progress. In some areas of psychology, computers are already showing exceptional efficiency. For instance, computer-generated virtual reality simulations already offer relatively successful therapy for different phobias such as the fear of flying or the fear of small animals (Rothbaum et al. 1997). Moreover, in clinical psychology, several available programs offer a cognitive training of attention, reactiveness, memory, planning behaviour, or the like for patients with neurological damage (Gontkovsky et al. 2002). Moreover, in the educational sciences, computers are used for learning purposes (Tamim, et al. 2011; Rienties, et al. 2012; Young, et al. 2012).

However, in the field of psychological and educational assessment, researchers are still discovering how to properly adapt paper-pencil versions of psychological assessments into computerised versions and how to properly construct new computer-based assessments. Although there are potentially major benefits that come along with computer-based assessment (e.g., cost savings, easier data saving, automated data scoring, automatic item generation, and the possibility of using different types of items (Bennett et al. 1999; Greiff, Holt, and Funke 2013), researchers are still looking for ways to exploit these potential benefits.

Converting paper-pencil tests into computer-based assessment
Regarding the adaption procedures, tests are currently being converted to computer-based formats as the result of a general trend (Parshal et al. 2002). As Parshal et al. further explain, these computer-based formats are automatically considered to be the first choice and better than traditional paper-pencil tests, which is not necessarily the case. Moreover, in the past, the importance of adaptation procedures and equivalence studies was often underestimated and has partly resulted in computer tests with poor psychometric quality. In fact, Huff and Sireci (2001) analysed the impact of computer-based administration of originally paper-pencil tests on their validity and emphasise the importance of validation studies. They further suggest that the issue at hand is not only the equity of paper-pencil and computerised versions of the same test, but also the question of whether computer-based administration introduces any additional construct-irrelevant variance into an assessment. Such irrelevant variance could be the result of computer proficiency, computer platform familiarity, user interface, or even the presence of computer anxiety of the test taker.

Furthermore, when adapting a paper-pencil version of a test into a computerised version, motivation and acceptance are other aspects that need to be considered. Motivation as a factor that can influence the results of any kind of test often depends on whether the test is a low-stakes test (e.g., when results do not have a significant effect on the future of the participant) or a high-stakes test (e.g., entrance exams). Many experimental studies have shown the potential of computer-based assessment to capture a person's motivational level. This can be done by collecting and objectively analysing contextual information (see e.g., Csapo, Lorincz, and Molnar 2012), for instance, measuring response time. Rushed answers may suggest guessing behaviour or a lack of motivation (Wise and Kong 2005). Acceptance is another important factor that can influence participants' behaviour in a test situation.

Various empirical studies have examined whether the use of computers in assessment changes participants' level of acceptance. The results of these studies have yielded the conclusion that the acceptance of a paper-pencil version and a computer-based version of the same test do not differ. However, the studies have mostly been carried out on a student sample, and in today's world, students have a lot of previous experience in handling computers. It should be considered that there might be an effect of computer experience and computer anxiety on acceptance and on test scores in computer-based assessments (McDonald 2002).

Construction of new computer-based assessment tools
Besides adapting existing paper-pencil assessment tools to computer-based versions, the expansion of informational technologies brings the opportunity to construct new innovative computer-based instruments. Moreover, new technologies are enabling us to assess some aspects of human behaviour and cognitive abilities that we were not able to assess before. For instance, computer-based assessment enables the implementation of various types of material – videos, audio material, drag-and-drop tasks, or dynamic items – into the assessment procedure (Parshall et al. 2010). The use of these kinds of items renders it possible to assess more complex skills such as problem solving in interactive environments, creativity, or collaborative problem solving skills – items that were not possible to construct for paper-pencil tests.

To conclude, besides the convenience and new possibilities that computer-based assessment promises, constructing a computer-based psychological assessment or adapting an existing paper-pencil test to a computerised version can be a very delicate procedure that requires very extensive validity studies of different kinds. In this paper, we illustrate the above-mentioned complex planning process of constructing a computer-based assessment tool with an actual example of a computer-based assessment of collaborative problem solving. Thereby, we introduce and discuss the main issues that need to be considered when constructing such an instrument.

Introduction to 21st century skills

The literature in the last few decades has emphasized the importance of "new" so-called 21st century skills on the labour market and consequently in education. For instance, Binkley et al. (2012) explain 21st century skills as the skills essential for a potential worker to be successful in today`s world. Furthermore, Autor, Levy, and Murnane (2003), who investigated the shift in requirements in the 21st century labour market, emphasise the importance of "new" 21st century skills that future workers will have to possess to be successful and identify collaborative problem solving as one of those skills. This comes from the fact that we are confronted by the need to solve problems in a group in various life areas. Whether we are aware of it or not, we solve problems collaboratively in many aspects of our lives – in our families, when we make decisions and solve household problems, and also in our work, where we collaborate on different levels with colleagues to exchange knowledge and expertise and to try to reach common goals.

The requirement of dealing with problems on a group level begins as early as childhood in the educational context. In schools, students are frequently asked to solve problems collaboratively – prepare a presentation together, perform science experiments collaboratively, and participate in various group activities. In all these situations, students need to communicate, exchange information, share ideas, negotiate, and solve conflicts in order to be successful. Besides collaborative problem solving, students are increasingly often asked to learn collaboratively. This comes from the fact that more and more teachers accept the idea that compared to frontal teaching methods, which are more convenient to apply and require less time for preparation, instructional methods based on collaboration offer some great advantages (Topping 2005). In the 80s and 90s, when collaborative teaching techniques started to spread with growing speed, many experiments were conducted

to provide some empirical evidence for these advantages, which practicing teachers already sensed (Slavin 1990; Johnson and Johnson 1996). Moreover, research showed the benefits of collaborative learning with respect to pupils' problem solving and critical thinking skills, creativity, metacognition, the amount of transferable acquired knowledge, and also motivation for learning. Considering personal and social aspects, collaborative learning techniques were also shown to be successful at decreasing students' anxiety levels and increasing their self-confidence. To that end, Kagan (1994) reported that the cohesion of children within their classes and their attitudes toward their classmates also improved in collaborative learning settings. Furthermore, the development of information and communication technology had an impact on the practice of collaborative learning, opening the gate to many new possibilities for instruction and also creating the necessity of the new strand of research of computer-supported collaborative learning, which, since then, has emerged tremendously (Stahl, Koschmann, and Suthers 2006).

Collaborative problem solving as a construct and the related skills
In the literature, different authors have provided inconsistent definitions of collaborative problem solving as a construct. For instance, O`Neil et al. (2003) describe collaborative problem solving as searching for the path from the initial state to the goal state while interacting with others working on a shared goal. One other attempt at an appropriate definition comes from one of the most recognised large-scale assessments worldwide, the Programme for International Student Assessment (PISA). There, collaborative problem solving is defined as "…the capacity of an individual to effectively engage in a process whereby two or more agents attempt to solve a problem by sharing the understanding and effort required to come to a solution and pooling their knowledge, skills and efforts to reach that solution" (OECD 2013). The inconsistencies in the available definitions are mainly comprised of the fact that authors have different perceptions of the dimensionality of the construct. Whereas O`Neil et al.'s (2003) definition provides a clear differentiation between collaboration and problem solving skills as two separable parts of collaborative problem solving, the definition provided by the OECD considers collaborative problem solving to be a mixture of the two (i.e., problem solving and collaboration), which again relies on different dimensions: establishing and maintaining a shared understanding, taking appropriate action to solve the problem, and establishing and maintaining team organization (OECD 2013).

Furthermore, in educational research, the terms cooperative learning, collaborative learning, and collaborative problem solving are used nearly synonymously. Underwood J. and Underwood G. (1999) define collaborative learning as learning environments in which small groups of students work together to achieve a common goal. Hence, the focus of the collaborative learning research strand lies on the learning processes and learning outcomes of the group. On the other hand, collaborative problem solving research refers to problem solving activities that involve interactions within a group of individuals (Zhang 1998). Thus, in order to draw a clear line between these two related research strands, it is important to consider the fact that collaborative learning and collaborative problem solving both happen in the group context but focus on the different processes - learning versus problem solving.

State of the art – the assessment of collaborative problem solving
Notwithstanding the fact that various authors define collaborative problem solving
differently, there is a certain level of agreement when it comes to the importance of
assessing this construct in individuals. For the last several years there has been a
strong initiative with the goal of creating a valid assessment instrument for
collaborative problem solving. However, there are still no reliable and valid
instruments, nor are there many experimental studies. Therefore, the collaborative
problem solving research strand needs to rely on the findings of related research
areas such as problem solving, collaboration, or collaborative learning. More
specifically, in recent decades, many researchers have aimed to assess the problem
solving skills of individuals, and these research strands have offered a solid basis for
constructing a problem solving environment for collaborative problem solving
assessment (Greiff et al. 2013; Molnár, Greiff, and Csapó 2013; OECD 2010).
Additionally, in the collaboration research strand, there are many findings on group
behaviour, individuals' behaviour in the group, and various outcomes of solving a
problem together (Pazos, Micari and Light 2010; Turel and Zhang 2010). Hence,
collaborative problem solving research can build on the findings of these related
research strands.

The major problems that can be detected in existing attempts to provide
assessments of collaborative problem solving consist of determining the structure of
the problem that collaborators should work on (e.g., if the problem is topic related, or
context free), deciding on the structure of the communication (e.g., open chat, or
predefined messages), and determining how to control for the influence of different
kinds of collaborators on the test takers' behaviour (e.g., using computer agents or
real human agents as collaborators).

Hsieh and O`Neil (2002) assessed the performance of individuals in a team. In their
empirical study, they applied a so-called knowledge mapping task through a
simulated Internet Web space (Schacter et al. 1999) to follow the information
searches of the team members who could communicate by exchanging predefined
messages. Despite the many new and creative solutions provided by this setup, it
still incorporated only one type of problem and was not able to control for different
types of collaborators in the situation. Moreover, using predefined messages
lowered the face-validity of the setting.

The ATC21S project (Griffin and Care 2012) is also one of the few projects that
aimed to assess collaborative problem solving. Here, students were instructed to
work on different types of problems in pairs and to communicate through chat
sessions. This assessment approach presented different types of problems to solve.
However, involving various dyads, the control of the team context variable as an
influencing factor seemed to become even weaker, not to mention that automatic
scoring was not possible because of the use of open chat format.

Rosen and Tager (2013) reported another attempt to assess collaborative problem
solving on the individual level. This method implemented certain elements of the two
above-mentioned assessment tools but also introduced one new element. Here,
students were able to exchange predefined messages through a chat box while
solving the interactive task in dyads. However, in this assessment tool, one member
of the dyad was actually a computer agent. The application of only one task left the

variable of the problem type uncontrolled again, but some progress was made in controlling the group context variable by creating a single ideal collaborator in the character of a computer agent. Nevertheless, as we will discuss later in this paper, even this promising assessment method provided some constraints to consider. For this reason, we will address these possible constraints in further sections.

The challenges of computer-based collaborative problem solving assessments
In this section, we will take a closer look at some of the problematic issues that arise when creating a computer-based assessment tool for collaborative problem solving: creating a standardised testing environment, using computer agents as collaborators, assessing communication, and analysing the collected data. We will then provide a deeper exploration of the thus-far-applied solutions.

Creating a standardised testing environment
One of the biggest challenges of creating a suitable assessment instrument is to achieve a standardised environment in which collaborative problem solving will occur. To realise this from a psychometric point of view, each individual should be placed in the exact same situation in order to obtain comparable data. However, when creating a test environment for collaborative problem solving, there is a very large set of variables that need to be controlled. Besides vigilance and a person's specific emotional state and character, which are the variables on a personal level, the characteristics of the situation in which a person solves a problem collaboratively can also be a factor that can affect the person's behaviour.

More specifically, working under time pressure, working in a situation ridden by conflict, or working in a stressful or in a peaceful environment can all be factors that moderate a person's behaviour. For instance, when a person works under time pressure, she/he can minimise time spent in conversation in order to have more time for the task itself. In the conflict situation, a person can decide not to collaborate and therefore not show any collaborative behaviour. The type of problem also influences people's motivation, and results may further depend on whether or not the person has prior experience with that type of problem (e.g., Is the person an expert on the topic or a complete novice?) (Chi, Feltovich, and Glaser 1981).

Furthermore, in reality, people collaborate with different kinds of people – friends or strangers, highly competent or incompetent people, assertive, submissive, or competitive individuals. To that, the results of a person`s work depend not only on him/her but also on the people he/she is working with, on the level of comfort the person experiences in the given group, and on the way the person complements the group with respect to his/her cognitive and social skills and prior knowledge. However, in assessing the collaborative problem solving skill of an individual, there is still no instrument that is able to include all these factors of influence and at the same time is able to fulfil the basic psychometric requirements of reliability and content and internal validity and ensure transparent and reasonable scoring. However, due to the emerging importance of collaborative problem solving in educational psychology, there is a strong initiative to finally provide a sound measurement of this construct.

For this reason, the authors of this paper propose that the testing situation be standardised so that all test takers will be put in exactly the same testing situation,

that they work on a context-free problem so that none of the test takers has prior experience with it, and that all test takers have the same collaborator as a partner, thus offering exactly the same stimuli to all participants. Because this is imaginable only if a computer agent is used as the collaborator, we will discuss this matter in the next section.

Using computer agents as collaborators
In order to assess the collaboration part of collaborative problem solving, we need to have someone in the assessment situation to collaborate with. If the collaborators are actual people whose behaviours are generally unpredictable, it becomes nearly impossible to construct a standardised setting for the assessment. Computers offer a sophisticated solution for controlling this variable: the use of computer agents as collaborators in the assessment instead of real humans (OECD 2013; Rosen and Tager 2013). The application of a computer agent capable of written or oral speech and even gesturing seems to be a very promising way to optimise the conditions for standardisation as the outcome of the collaboration would not depend on the partner's skills and behaviour (Graesser, Jeon, and Dufty 2008). Through the application of this solution, we can offer an optimal standardised partner to make it possible for the test taker to demonstrate his or her skills. As each participant would work in the same situation with the same collaborator, each participant's results would be comparable.

However, despite the great potential of achieving a relatively high level of standardisation, the degree to which this solution resembles real-life collaboration is arguable, and moreover, it even raises some ethical questions. First, the face validity of using a computer agent compared to a human agent is questionable. Despite developments in programming the behaviour of computers (i.e., communicating emotions, reacting adequately to people's emotions), generating and managing conflicts are still barely expected from a computer agent. Second, different experiments have verified the important role of social presence in human behaviour – people act differently when they interact with a computer than with a real human (Weinel et al. 2011). Miwa and Terai (2012) found that people's behaviour is influenced by instructions that indicate whether the partner is a computer or a human and not necessarily by the partner's behaviour itself. Therefore, it can be assumed that using a computer agent as a partner and providing transparent instructions about this fact will not trigger the same behaviour as collaboration with a real human will; thus, this assumption must be empirically investigated. Finally, the other question that arises is whether it is even ethical to construct a computer agent that is so realistic that the person working with it does not notice that she/he is working with a computer, and the argument could even be made that this practice deceives the test-taker.

All in all, although using computer agents as collaborators is not unproblematic, the authors of this paper highly recommend this approach because it represents the only method of achieving comparability across test takers` results.

Assessing communication in collaborative problem solving
Another challenge in the computer-based assessment of collaborative problem solving is to adequately structure and furthermore to assess communication, which is the essential component of collaborative problem solving. As the goal is to provide

an assessment that will be applicable in different settings, for example, in a large-scale assessment or in a school setting where special equipment (e.g., headphones, high screen resolution, or high internet speed) is not available, a video or audio chat session is not a good option. However, if we exclude image and sound, an important part of communication is lost: that is, nonverbal communication (i.e., gestures, touch, physical distance, facial expression, or eye contact). Nevertheless, nonverbal communication is a very important aspect of communication considering that two thirds of human conversation is nonverbal (Nistor 2012).

One way to structure communication in a collaborative problem solving task is to use chat boxes. Despite the lack of nonverbal information, the written form of information exchange can still be very beneficial. Specifically, one of the advantages of using written communication is that more or less automatic scoring of the obtained data can be applied. Moreover, considering the fact that collaborative problem solving in a written form is more and more common in our daily lives, the idea of assessing collaborative problem solving skills through chatting should not be viewed as a particular constraint (Hermann, Rummel, and Spada 2001).

In order to investigate communication patterns better, different types of content-analysis software enable relatively detailed and reliable content analyses of written communication in collaborative problem solving assessments. However, in this field, there are further constraints – such software is not available in all languages, its development is expensive, its application is time consuming, and therefore, its application not possible in every assessment setting, for instance, in a large-scale assessment. For the purposes of a large-scale assessment, automatic scoring appears to be necessary, and thus, the application of predefined chat messages could provide a solution. A chat with predefined messages mostly relies on pre-studies that investigate the most important and most common messages that are sent through open chat (Hsieh and O'Neil 2002; Rosen and Tager 2013). However, besides underlining the inflexibility of this solution, which can cause frustration in the participants, we also have to address the question of whether predefined messages are suggestive. That is, does offering a specific message at the same time implies that the message should be used. A possible research question in this area is, for instance, what effect the sequence of messages has (i.e., primacy/recency effects: the first or the last message a person sees in the list influences which message he/she is going to choose).

In summary, the written form of communication in collaborative problem solving assessment is probably the best possible choice. Moreover, research findings suggest that the use of predefined messages instead of open chat sessions is a good solution to enable automatic scoring of the obtained data and to further improve the psychometric quality of the assessment instrument.

Analysing collected data – the use of log-files
After the data are collected, further questions arise. For instance, there is the question of how qualitative data (e.g., the choice of approach used for collaboration – asking questions, or making demands) can be quantified, and if, for instance, the quantity of exchanged messages can be an indicator of participants' openness to collaboration. The quantity of exchanged messages can be obtained from so-called log-files. A log-file is a file in which a computer documents all its activities during a

session, for example, during one test (Andrews and Zhang 2000). These files include not only the numbers of clicks and mouse movements but also the time that passes between each movement, mouse click, and so forth. Such "meta-data" extracted from log-files can be a rich source of additional information about test takers` behaviour in a collaborative problem solving situation. Several studies have reported using log-files to examine think-aloud protocols, eye tracking, head movement, or facial expressions recorded by computers (Van Gog, Paas, and van Merriënboer 2005; Csapo, Lorinz, and Molnar 2012). Considering the results of aforementioned studies, log-file data may play a very useful role in assessments of collaborative problem solving. For instance, mouse movements may provide information about who is the first one to react or which collaborator provides the actual answer. This kind of information may let us draw conclusions regarding the dynamics of collaboration. However, in order to use log-file data, researchers must face the challenge of investigating which information is actually important for collaborative problem solving and how this information can be interpreted.

Discussion and outlook

Computers play an increasingly important role in psychological assessment. The level of exactness, savings of time and money by reducing the need for human effort, as well as the potential of automatic scoring systems are enormous advantages that would seem wasteful not to exploit. Moreover, they provide new opportunities in testing, such as using videos, drag and drop items, and many types of contextual behavioural data (Greiff, Holt, and Funke 2013; Parshal et al. 2002; Bennett et al. 1999). In this paper, we used the computerisation of collaborative problem solving assessments as an example to demonstrate how many significant constraints researchers need to overcome and how many crucial questions still remain unanswered. To mention some: Is it possible to create a standardised environment for computer-based collaborative problem solving? Can a computer agent be considered a face-valid collaborator? How can we overcome the constraints of using a computer agent as a collaborator? How can we structure communication, and moreover, how can we assess it properly? Finally, what do we do with the contextual data collected by log-files in computer-based assessment?

Generally, the most important question and challenge for researchers is how to make the best of the computer-based assessment – how to overcome the aforementioned constraints and exploit the advantages to the fullest. In the case of collaborative problem solving assessment tools and in order to outline the exact limitations, a great number of validation studies are waiting to be carried out. For instance, data obtained by the assessment tools that apply the human-to-agent collaboration condition need to be compared empirically with the data collected by giving the same problems to groups in a human-to-human condition.

Similar as for the collaborative learning it is to expect that collaborative problem solving plays an important role in classrooms and in learning processes in schools. Since the educational methods are increasingly turning from frontal methods to the pupil cantered and collaborative learning methods (Slavin 1990; Johnson and Johnson 1996), it is highly possible that good collaborators in problem solving situations benefit from these skills in their school results. For this reason it is very important to be able to assess collaborative problem solving in individuals` and further on, to facilitate it.

All in all, the necessity of assessing collaborative problem solving, which comes from its obvious importance in job arena and in education, is pushing researchers forward to search for new, innovative, and promising ways to assess this essential skill in the future.

References

Andrews, J. H., and Zhang, Y. 2000. Broad-spectrum studies of log file analysis. In *Proceedings of the 22nd international conference on Software engineering*, 105-114. ACM.

Autor, D. H., Levy, F., and Murnane, R. J. 2003. The skill content of recent technological change: An empirical exploration. *Quarterly Journal of Economics 118*: 1279-1333.

Bennett, R. E., Goodman, M., Hessinger, J., Kahn, H., Ligget, J., Marshall, G., and Zack, J. 1999. Using multimedia in large-scale computer-based testing programs. *Computers in Human Behavior 15:* 283-294.

Binkley, M., Erstad, O., Herman, J., Raizen, S., Ripley, M., and Rumble, M. 2011. Defining 21st Century Skills. In P. Griffin, B. McGaw, and E. Care (Eds.). Assessment and teaching 21st century skills. Heidelberg: Springer.

Chi, M. T. H., Feltovich, P. J., and Glaser, R. 1981. Categorization and representation of physics problems by experts and novices. *Cognitive Science* 5: 121-52.

Csapo, B., Lorincz, A., and Molnar, 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. 235-254. New York: Springer.

Daniel, R. C., and Embretson, S. E. 2010. Designing Cognitive Complexity in Mathematical Problem-Solving Items. Applied Psychological Measurement 34, 5: 348-364.

Dunbar, K., and Fugelsang, J. 2005. Scientific thinking and reasoning. In K. L. Holyoak and R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning*. 705-725. New York, NY: Cambridge University Press.

Gontkovsky, S. T., McDonald, N., B., Clark, P. G., and Ruwe, W. D. 2002. Current directions in computer-assisted cognitive rehabilitation. *NeuroRehabilitation 17*, 3: 195-199.

Graesser, A. C., Jeon, M., and Dufty, D. 2008. Agent technologies designed to facilitate interactive knowledge construction. *Discourse Processes 45:* 298–322.

Greiff, S., Holt, D., and Funke, J. 2013. Perspectives on problem solving in cognitive research and educational assessment: analytical, interactive, and collaborative problem solving. *The Journal of Problem Solving 5*: 71-91.

Greiff, S., Wüstenberg, S., Molnar, G., Fischer, A, Funke, J., and Csapo, B. 2013. Complex Problem Solving in educational settings – something beyond g: Concept, assessment, measurement invariance, and construct validity. *Journal of Educational Psychology 105*: 364-379.

Griffin, P. and Care, E. 2012. *Challenges in internet-based CPS Assessment.* Paper presented at the Conference of the International Testing Commission, July 2012, in Amsterdam, NE.

Hermann, F., Rummel, N. and Spada, H. 2001. Solving the case together: The challenge of net-based interdisciplinary collaboration. In P. Dillenbourg, A. Eurelings, and K. Hakkarainen (Eds.), *Proceedings of the first European conference on computer-supported collaborative learning*. 293-300. Maastricht, NL: McLuhan Institute.

Hsieh, I.-L. and O'Neil, H. F. Jr. 2002. Types of feedback in a computer-based collaborative problem solving group task. *Computers in Human Behavior 18*, 1: 699-715.

Huff, K. L., and Sireci, S. G. 2001. Validity issues in computer-based testing. *Educational Measurement: Issues and Practice 20, 3*: 16-25.

Kagan, S. 1994. Cooperative Learning. San Clemente, CA: Kagan Publishing.

Johnson, D. W., and Johnson, R. T. 1996. Cooperation and the use of technology. In Jonassen, D. H. (Eds.), *Handbook of Research for Educational Communications and Technology,* 1017-1044*.* New York, Macmillan Library Reference.

McDonald, A. S. 2002. The impact of individual differences on the equivalence of computer-based and paper-and-pencil educational assessments. *Computers & Education 39*, 3: 299-312.

Miwa, K., and Terai, H. 2012. Impact of two types of partner, perceived or actual, in human-human and human-agent interaction. *Computers in Human Behavior 28:* 1286-1297.

Molnar, G., Greiff, S., and Csapo, B. 2013. Inductive reasoning, domain specific and complex problem solving: relations and development. *Thinking skills and Creativity 9*, 8: 35-45.

Nistor, G. 2012. The Role of the Nonverbal Communication in Interpersonal Relations. *Procedia - Social and Behavioral Sciences 47*: 552-556.

OECD. 2010. *PISA 2012 Problem Solving Framework*. Paris: OECD.

OECD. 2013. PISA 2015 Draft Collaborative Problem Solving Framework. http://www.oecd.org/pisa/pisaproducts/Draft%20PISA%202015%20Collaborative%20Problem%20Solving%20Framework%20.pdf (accessed: October 17 2013)

Parshall, C. G., Harmes, C., Davey, T., and Pashley, P. J. 2010. *Innovative items for computerized testing.* In W. J. van der Linden and C. A. W. Glas (Eds.). Computerised adaptive testing: Theory and practice (2nd ed.). Norwell, MA: Kluwer Academic Publishers*.*

Parshall, C. G., Spray, J. A., Kalohn, J. C., and Davey, T. 2002. Practical considerations in computer-based testing. Springer: New York.

Pazos, P., Micari, M., and Light, G. 2010. Developing an instrument to characterise peer-led groups in collaborative learning environments: assessing problem-solving approach and group interaction. *Assessment & Evaluation in Higher Education 35*, 2: 191–208.

Rienties, B., Kaper, W., Struyven, K., Tempelaar, D., Van Gastel, L., Vrancken, S., and Virgailaite-Meckauskaite, E. 2012. A review of the role of Information Communication Technology and course design in transitional education practices. *Interactive Learning Environments 20*, 6: 563-581.

Rosen, Y., and Tager, M. 2013. Computer-based Assessment of Collaborative Problem Solving Skills: Human-to-Agent versus Human-to-Human Approach. http://researchnetwork.pearson.com/wp-content/uploads/CollaborativeProblemSolvingResearchReport.pdf (accessed May 23, 2013)

Rothbaum, B. O., Hodges, L., and Kooper, R. 1997. Virtual reality exposure therapy. *The Journal of Psychotherapy Practice and Research 6*, 3: 219-226.

Schacter, J., Herl, H. E., Chung, G., Dennis, R. A., and O'Neil, H. F. Jr. 1999. Computer-based performance assessments: a solution to the narrow measurement and reporting of problem-solving. *Computers in Human Behavior 15:* 403–418.

Slavin, R. E. 1990. Cooperative Learning: Theory, Research, and Practice. NJ: Prentice-Hall, Englewood Cliffs.

Stahl, G., Koschmann, T., and Suthers, D. 2006. *Computer-supported collaborative learning: An historical perspective.* In Sawyer, R. K. eds. Cambridge handbook of the learning sciences. 409-426. Cambridge, UK: Cambridge University Press.

Tamim, R. M., Bernard, R. M., Borokhovski, E., Abrami, P. C., and Schmid, R. F. 2011. What Forty Years of Research Says About the Impact of Technology on Learning A Second-Order Meta-Analysis and Validation Study. *Review of Educational Research 81*, 1: 4-28.

Topping, K. 2005. Trends in peer learning. *Educational Psychology 25*, 6: 631-645.

Turel, O., and Zhang, Y. 2010. Does virtual team composition matter? Trait and problem-solving configuration effects on team performance. *Behaviour & Information Technology 29*, 4: 363–375.

Underwood, J., and Underwood, G. 1999. *Task effect on cooperative and collaborative learning with computers*. In K. Littleton and P. Light. eds. Learning with computers: Analysing productive interaction. 10–23. London: Routledge.

Van Gog, T., Paas, F., and van Merriënboer, J. J. G. 2005. Uncovering expertise related differences in troubleshooting performance: combining eye movement and concurrent verbal protocol data. *Applied Cognitive Psychology 19:* 205-221.

Weinel, M., Bannert, M., Zumbach, J., Hoppe, H. U., and Malzahn, N. 2011. A closer look on social presence as a causing factor in computer-mediated collaboration. *Computers in Human Behavior 27*: 513-521.

Wise, S. L., and Kong, X. 2005. Response time effort: A new measurement of examinee motivation in computer-based tests. *Applied Measurement in Education 18*, 2: 163-183.

Young, M. F., Slota, S., Cutter, A. B., Jalette, G., Mullin, G., Lai, B., and Yukhymenko, M. 2012. Our Princess Is in Another Castle A Review of Trends in Serious Gaming for Education. *Review of Educational Research 82*, 1: 61-89.

Zhang, J. 1998. A distributed representation approach to group problem solving. *Journal of the American Society for Information Science 49, 9:* 801-809.