MOVING TOWARDS THE ASSESSMENT OF COLLABORATIVE PROBLEM SOLVING SKILLS WITH A TANGIBLE USER INTERFACE*

Eric Ras
Public Research Centre Henri Tudor, av. J.F. Kennedy 29, L1855 Luxembourg-Kirchberg, Luxembourg
eric.ras@tudor.lu

Katarina Krkovic
University of Luxembourg, Coudenhoven Kalergi, L-1855 Luxembourg-Kirchberg, Luxembourg
katarina.krkovic@uni.lu

Samuel Greiff
University of Luxembourg, Coudenhoven Kalergi, L-1855 Luxembourg-Kirchberg, Luxembourg
samuel.greiff@uni.lu

Eric Tobias
Public Research Centre Henri Tudor, av. J.F. Kennedy 29, L1855 Luxembourg-Kirchberg, Luxembourg
eric.tobias@tudor.lu

Valérie Maquil
Public Research Centre Henri Tudor, av. J.F. Kennedy 29, L1855 Luxembourg-Kirchberg, Luxembourg
valerie.maquil@tudor.lu

ABSTRACT
The research on the assessment of collaborative problem solving (ColPS), as one crucial 21st Century Skill, is still in its beginnings. Using Tangible User Interfaces (TUI) for this purpose has only been marginally investigated in technology-based assessment. Our first empirical studies focused on light-weight performance measurements, usability, user experience, and gesture analysis to increase our understanding of how people interact with TUI in an assessment context. In this paper we propose a research agenda for assessing ColPS of individuals using the Microworlds methodology implemented on TUIs. In a first example item, we use so-called MicroDYN items, which are independent microworld scenarios that rely on structural linear equations as underlying model. As the MicroDYN approach has been thoroughly empirically investigated for the assessment of complex problem solving of individuals, it offers a good basis for a reliable and valid assessment. We describe how this approach was applied to create an assessment item for a collaborative setting. This item described in this paper implements a simplified model of a MicroDYN item related to climate change using knowledge of previous studies. Therefore, the focus of the item's construction lies on meeting the requirements for a standardised high quality assessment. Finally, a research agenda is proposed to sketch the main research issues.

INTRODUCTION
In the last few years the term 21st Century Skill gained substantial visibility in scientific literature, for instance in the latest reports on technology-based assessment (OECD, 2012) as well as in the Digital Agenda published by the European Commission (Ferrari, 2012). These so-called 21st Century Skills refer to skills such as complex problem solving, collaborative problem solving, creativity, critical thinking, learning to learn, decision making, etc. (Binkley et al., 2012). Several researchers have stated that the acquisition of these 21st Century Skills and their development are only marginally investigated (Bennett & Gitomer, 2009) despite their unquestioned importance and relevance. This paper focuses on a particular 21st Century Skill, collaborative problem solving (ColPS), which encompasses the ability to successfully deal with untransparent and dynamically changing problems in a collaborative setting. ColPS is considered an important component for the success in life and for this reason will be included as transversal domain in one of the most prestigious large-scale assessments worldwide, the Programme for International Student Assessment (PISA) in its 2015 cycle (OECD, 2014). As emphasised by educators and policy makers, ColPS plays an important role in different life contexts - at schools, at home, or at work (Rummel & Spada, 2005). Especially on the labour market, the organisational researchers (e.g. (Cannon-Bowers & Salas, 1997; Hellinghausen & Myers, 1998)) emphasise more and more the importance of teamwork for organisational success and continuously look for new ways to assess collaborative skills in a recruitment process and in the course of personnel development.

* This article is a substantially extended version of an already published conference paper that was published in the conference proceedings of the CAA 2014 Conference [Maquil, V., Tobias, E., Greiff, S., & Ras, E. (2014). *Assessment of collaborative problem solving using linear equations on a tangible tabletop*. Paper presented at the 2014 International Computer Assisted Assessment (CAA) Conference, Zeist, The Netherlands.]. Substantial changes have been made for this extended version, but some parts have remained unchanged.

Despite the apparent importance of ColPS for success in different areas of life, up until now there has not been much progress in its assessment. However, recently, fast development of new technologies for gathering data in real settings allows us to assess specific 21st Century Skills such as ColPS. Due to practical limitations, conventional paper-pencil tests or even the use of desktop applications known from computer-based assessment do not entirely allow for highly dynamic and interactive assessment of 21st Century Skills. More natural forms of interaction such as touch, speech, gestures, and handwriting support assessment researchers to explore new approaches to skill assessment. Therefore, the availability of affordable devices that can collect such natural forms of interaction motivated us to develop several so-called tangible user interfaces (TUI) for assessment. ColPS assessment.

The use of TUI in assessment is new and, therefore, knowledge about how to design those systems and how to develop test items for TUI is limited. For this reason, we present the state of the art in related research topics and propose a study that will help us shed light on the use of TUIs for ColPS assessment purposes.

The aim of this article is to summarise open research issues in the domain of complex collaborative problem solving and to motivate the use of tangible user interfaces to assess this 21st Century Skill.

The upcoming section summarises research on relevant topics followed by the description of the overall research methodology. Further, we elaborate how a test item related to climate change has been developed for the TUI to assess collaborative problem solving of individuals. Finally we derive a first research agenda to assess complex collaborative problem solving with TUIs.

BACKGROUND

Collaborative Problem Solving

ColPS is one of the 21st Century Skills that has attracted the most attention from researches in recent years. This comes from the fact that collaborative work is becoming ever more significant in education and in everyday life. For instance, in schools students are often required to work in teams on science tasks; at work people collaborate on common projects; and at home families take household decisions collaboratively. An everyday example of ColPS is students organising a school trip together. They need to consider different factors such as distance, transport, or price, whereby some of factors can change in the course of planning. Students take over specific roles in the process (e.g. a person who finds bus schedules, the one taking care of hotel prices, the one who makes an overview of everything), exchange their knowledge and apply different strategies to find the nearest destination, with the optimal transport and the cheapest price. In theory, ColPS incorporates two dimensions – complex problem solving as the cognitive dimension and collaboration as the interpersonal dimension. However, ColPS is not just the sum of those two dimensions, but represents the interaction of complex problem solving skills and collaboration skills (OECD, 2014).

The first dimension, complex problem solving, is one of the 21st Century Skills that has been extensively investigated. It may be defined as searching for the path from the problem state to the goal state. Its main characteristics are that the problem the person is confronted with is untransparent and the problem situation is dynamic, meaning that it changes, for instance, in the course of time or under the influence of the problem solver (Fischer, Greiff, & Funke, 2011; Lai, 2011).

By adding the collaborative aspects to the complex problem solving construct leads us to the definition by O`Neil et al. (2003), who define ColPS as searching for the path from the initial state to the goal state while interacting with others working on a shared goal. Further, one of the most recognised large-scale assessments worldwide, the Programme for International Student Assessment (PISA) offers a more detailed definition of ColPS describing it 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).*

Whereas the dimensions of ColPS, problem solving and collaboration, are well examined as single constructs, the ColPS as a construct that incorporates both of those skills is still insufficiently investigated. Although recently there are initiatives researching the assessment of ColPS, especially in large scale testing context, the progress is still scarce at best. As outlined by Krkovic et al. (Krkovic, Pásztor-Kovács, Molnár, & Greiff, 2014) following aspects need to be considered when constructing the assessment for ColPS - how to create a standardised testing environment, to use computer agents or humans as collaborators, how to assess communication, and how to analyse the collected data. Up until now, there have been two approaches on how to assess ColPS – human-to-agent and human-to-human. The first approach incorporates human-to-agent collaboration, the partici-

pant collaborates with a computer-simulated agent to solve a problem collaboratively. The OECD will include the assessment of ColPS in its PISA 2015 cycle uses this approach. The application of human-to-agent approach can be justified by the fact that it offers standardised stimuli to each participant ensuring a standardized testing environment and reasonable scoring procedures (Graesser, Jeon, & Dufty, 2008). Another approach to the assessment of ColPS is to use human-to-human interaction as investigated in the ATC21s project (Griffin, Care, & McGaw, 2011) which in comparison to the human-to-agent approach offers more face valid situation of two or more individuals working together on a problem, providing more in-depth information about the collaboration process. However, the human-to-human approach has its limitations too, such as unstandardized assessment settings, or large log files coming from open conversations that can hardly be transferred into scoring (O'Neil, et al., 2003).
Considering benefits and limitations of both approaches, researchers need to search for an optimum solution for ColPS assessment that will on one hand offer scenarios that are scoreable and on the other hand be realistic and close to real life situations.

Programs such as PISA aim at large scale assessment, hence, the technology they use must be available in schools and can be implemented in classes. For this reason PISA uses computer-based assessment which allows for the administration in school to use their existing IT infrastructure. Due to the novelty and low deployment rate of TUI, their use for large scale assessment seems unpractical in the near future. As of now, assessment on TUI could be used in scenarios such as personnel selection in human resources management. This scenario is quite probable considering that, at the moment, there is no reliable methodology to assess ColPS skills in the course of the selection process, where skills such as team work play an important role. Moreover , by using innovative user interfaces instead of computer-based testing richer data sets can be collected.

TUI-based assessment

The aim of tangible user interfaces (TUI) is to make computer bits tangible and to allow users to grasp and manipulate them with their hands (Ishii, 2008). This concept has a number of advantages for social and contextual interactions, such as collaboration (Hornecker & Buur, 2006). To date, literature provides first exploratory results on the learning benefits of TUIs, however, in the field of assessment research is still in its infancy. In particular, the additional haptic dimension, the better accessibility (for instance for children), and the shared space that can be used in collaborative situations (Marshall, 2007) are claimed to be beneficial in learning situations. According to Klemmer et al. (2006), the physical objects and actions of TUIs allow using multiple senses of our human bodies and have an essential impact on our understanding of the world. They encourage rapid epistemic actions (i.e. try and error actions)and thus lower cognitive load by simplifying thinking processes (Esteves, Van den Hoven, & Oakley, 2013). Learning and assessment can be supported by representing problems in a new way, using physical and digital elements. Further, the way the users are solving tasks can be detected and feedback can be directly given. Nevertheless, no TUI has been systematically used and evaluated in the context of technology-based assessment (TBA).

Since research on using TUIs for assessment purposes is still at the very beginning, Ras et al. (2013) have conducted a series of empirical studies in order to identify harmonies and tensions concerning the use of TUIs for assessing skills of complex and collaborative problem solving in a low-stake assessment context. The aims were manifold: at a first stage it was interesting to observe how users interact with the table and the tangibles in general and which spaces are used for interaction, communication, and collaboration. Furthermore, each study also assessed the user experience and the usability of the device to ensure a continuous improvement cycle. The figures below show already two different devices.

Several items have been developed, two of which are depicted in Fig. 1 respectively Fig 2. The aim of the matching item was to assign the labels of the planets (put on small tangibles) to the images. The goal of the windmill item was to explore the impact of different variables such as wind speed, number of blades, and height of the windmill on the energy produced by the windmill. The output changed in real-time in response to the manipulations of the input.

Fig. 1. Matching item

Fig. 2. Simulation item

A description of the outcomes of these studies can be found in Ras et al. (2013). In summary, it was interesting to observe that different spaces where used to intensively communicate using gestures, talk, and physically manipulating objects. Besides the interactive surface also the space between participants or even the non-reactive border of the table was used to solve the item (e.g., to sort, group, or even exclude tangibles). The large space of the TUI allows for more than one person to get actively involved. But the position at the table also reveals that a person might feel responsible for a specific space (or even tangibles) or cannot access other spaces because of the distance or the closeness to another participant. We observed that all activities across space support the participants to acquire knowledge about the problem situation and apply the acquired knowledge, which are two important dimensions of ColPS (Funke, 2001; Greiff et al., 2013).

Another requirement for an assessment context is to provide an authentic and intuitive environment. Both the windmill item as well as the climate change item described in the next section represent real-world scenarios. Further, the physical tangibles allow the participants to manipulate the parameters in a natural and simple way. They enable more efficient interactions, such as turning a parameter without looking at it, while another parameter can be observed on the TUI surface. These activities and many others we observed help the participants to acquire knowledge about the problem situation. In addition, during the next step where the participants apply knowledge to solve the problem other patterns have been detected such as: users often hold them in their hands, use them to make gestures and suggestions, or they produced noise with the object to increase awareness in the group. Well-designed tangibles may implement a specific metaphor to increase its understanding but we cannot ensure that users immediately understand how to us or manipulate it. Holding objects in their hand can harm other users to contribute or even impact the complete solving strategy.

During the first step of the problem solving process, knowledge on the problem situation needs to be acquired by all participants. We have observed that, in particular, during a first phase several uncoordinated trial and error manipulations were done, even in parallel by several participants at once, followed by participants suggesting hypotheses of how parameters impact others without any deeper reflection. Hence, the challenge is certainly to distinguish these chaotic, unplanned activities from those that are more systematic and serve to understand the problem or even to solve the problem. For each study we interviewed assessment experts with respect to the item and the usefulness of the table for assessment. One interesting outcome was that they claim that a TUI supports the users in recognizing and understanding the perspectives of others – an important skill of collaborative problem solving. Gestures and spoken communication support them to easily explain hypotheses. Different perspectives of individuals (e.g., a misunderstanding) can be recognized through realising them on the table.

In assessment is it clear that the technology used for assessment can impact test performance. Aspects such as motivation, familiarity with a technology, even anxiety to use a specific technology need to be measured as disturbing variables. User experience is a typical measure to evaluate software and its measurement serves to improve both pragmatic and hedonic goals of the user. Unfortunately, all user experience and usability scales have been developed for classical software but not for applications running on tangible devices. They do not explicitly distinguish between physical and digital characteristics of TUI-based applications. Therefore, a research objective of these studies was also to investigate weaknesses of standardised measurement instruments tailored to evaluate usability and user experience. A first study revealed weaknesses of the Simple Usability scale (Ras & Maquil, 2011) and a new measurement instrument, which combines measures of several existing models, was developed and evaluated with the Windmill example (Ras, et al., 2013).

The windmill item is similar to a simulator: parameters can be manipulated and impacts on other output parameter can be observed. In the following section so-called MicroDYN items have been chosen as a task type for the collaborative problem solving scenario to be implemented on a TUI.

MICRODYN AS TASK FOR COLPS – A CONCRETE EXAMPLE

MicroDYN

This study is the first to suggest implementing so-called MicroDYN items on a TUI and to employ them in a group (i.e., a collaborative) setting in order to assess ColPS. Specifically, MicroDYN tasks are independent microworld items, which are based on structural equations. Until recently, MicroDYN items have only been used as a computer-based assessment approach towards assessing complex problem solving skills in individuals. Such an approach enables very well to control the difficulty and complexity of items, as well as the scoring procedures. In MicroDYN, the problem solver can manipulate a set of input variables, which reflects on specific output variables. Thereby, the problem solver needs to acquire knowledge about input and output variables and detect causal relations between them by representing it in a graphic model (dimension of *knowledge acquisition*, (Mayer & Wittrock, 2006)). In a subsequent step, the problem solver needs to apply the acquired knowledge in order to reach predefined goals (dimension of *knowledge application*, (Novick & Bassok, 2005)). Several studies have shown the validity of MicroDYN with regard to the assessment of complex problem solving (e.g. (Greiff, et al., 2013)).

Recently, MicroDYN items have been implemented into the human-to-agent computer-based ColPS assessment. In such tasks, the participant has limited controls over the input variables and must collaborate with a computer-simulated agent in order to acquire knowledge and apply it in the further step. Thereby, the participant can communicate with the agent by using a chat with predefined messages. In this study we expand this approach by implementing MicroDYN in the collaborative setting by using human-to-human approach, whereby the participants work together on tangible table interface. Moreover, the use of tangible table interface for the first time for ColPS assessment will incorporate triads instead of dyads that are used in computer-based version of collaborative MicroDYN tasks, which will enrich the collaboration aspect.

A tangible climate change item

Climate change as a complex problem

Today, scientists agree that the Earth's climate system is warming up and that a significant cause of the change is due to human activities (IPCC Working Group I, 2013). This is often referred to as *Global Warming*. Global Warming is mainly evidenced by increases in global average air and ocean temperatures resulting in the widespread melting of snow and ice leading to rising global average sea level. The main reason of this warming is caused by increases in concentrations of greenhouse gases, to which the largest contributor is carbon dioxide (CO_2). Certain waste management and agricultural practices aggravate the problem by releasing chemical compounds such as methane and nitrous oxide, which further fuel the warming, trapping ever more heat in the atmosphere.

To allow children to understand the relations between the different variables, as well as to follow the MicroDYN methodology of Greiff et al. (2013), a simplified model of climate change was defined. The model's principle is as follows. The system is configured with eight variables, each with three sub-variables (see Fig. 3) and their respective simplified impact on the total amount of CO_2 (i.e., approximated on the basis of (klimAktiv gemeinnützige Gesellschaft zur Förderung des Klimaschutzes mbH, 2014)). The variables allow children to explore the impact on global CO_2 emissions using examples they are familiar with such as meat consumption, transportation means to school respectively to their vacation destination, or simply the amount of clothing they own. By selecting one of the sub-variables which implement, in various forms, the high, medium, and low choices for the given variable, they express their choice. The system shows the effect on the different outputs as given by the MicroDYN methodology. The feedback is visual and based on a known scenario the children were confronted with prior to the experiment as part of their curriculum on climate change (such as the dire situation of polar bears). Children are then able to freely change their assignment of variables to explore the magnitude of variables on CO_2 emissions.

Collaborative widgets for manipulating variables

To allow for a maximum of simultaneous accessibility and shareability of controls we designed a novel type of widgets: the *Tangible Pie Widget*. Questions are visualized as immutable pie charts (Ø 15cm) with physical borders where each of the sections corresponds to one potential answer. Prior to the solving activity, each of the children gets one token per question. The children can place their token into one of the sections in order to give their personal answer to the provided question. This allows defining the weights of the different sections of the pie widget that can then be translated to the world population (see Fig. 3).

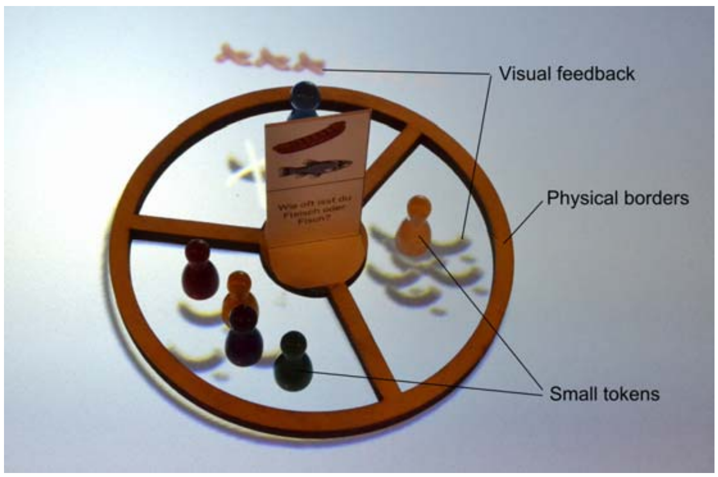

Fig 3. The tangible pie widget allows multiple users to change the weights of the sections by moving small tokens

For each parameter, the system provides feedback by depicting the value (see Figure XY) . Further it calculates and visualises the output of the total amount of CO_2 emissions and global temperature as defined by all pie widgets (see Figure 4).

The pie widgets can be used in two different ways. At the beginning, they are empty. Children place their token into the slices of the pie widget to cast their vote for each question. The system calculates the CO_2 output and the impact on global temperature in real time as votes are cast. This allows them to become familiar with the tool and to define the initial state. Furthermore, children can explore WHAT-IF scenarios, that is, they can change the value of variables by placing the tokens in a different section.

This approach incites all children to actively participate in setting the values and, in the second phase, allows them to collaborate in small groups to simultaneously manipulate the widgets. Exploring different scenarios will allow children to identify what behaviour has the most impact on CO_2 output and hence acquire knowledge about the underlying model.

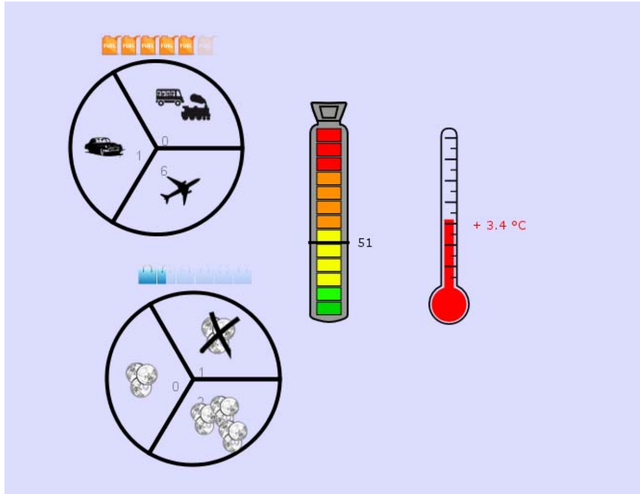

Fig 4. Feedback provided by the system

Fig 5. The system in use by a group of children

System architecture and design

The system features two main components: the TUI library (i.e., TULIP) that allows defining and instantiate all kind of tangibles; and a library that allows for defining complex problems. The TUI library is used to define all feedback and interactive tangibles as shown in Fig. 3 provided to the children. The complex problem solving library provides mathematical models such as linear equations which can be instantiated, for example, as seen in Equation 1 and 2. Each equation of the system is fed by the variables of one question while an instance of Equation 2 is used to provide the global CO_2 output.

$$\begin{aligned} Q_1(t+1) &= w_{X_1} * X_1(t) + w_{Y_1} * Y_1(t) + w_{Z_1} * Z_1(t) \\ Q_2(t+1) &= w_{X_2} * X_2(t) + w_{Y_2} * Y_2(t) + w_{Z_2} * Z_2(t) \\ &\vdots \\ Q_n(t+1) &= w_{X_n} * X_n(t) + w_{Y_n} * Y_n(t) + w_{Z_n} * Z_n(t) \end{aligned} \quad (1)$$

$$CO_2(t+1) = w_{Q_1} * Q_1(t+1) + w_{Q_2} * Q_2(t+1) + \cdots w_{Q_n} * Q_n(t+1) \quad (2)$$

With Q, the question dependent output, t the discrete time steps, w_i the weight of variable i such that $w_i > 0$, and X_i, Y_i, Z_i the input variables for a question i.

This setup allows us to weigh each variable according to their impact on the CO_2 output. While these are all approximations, we believe they are necessary to provide children with interesting questions they can answer based on their experience and their environment and see the effect of their actions even if it would normally be too small to see in a more sophisticated and real life simulation.

Cycle of use and data gathering

The cycle of learning and assessment includes several steps. In a first phase, children are asked to answer questions defined by four widgets lying next to the interactive surface. For each question, the children place one of their tokens into the related area of the widget. The table then provides immediate feedback about the total value of the parameter, as well as the CO_2 emissions they are creating per person, and the effect on temperature.

In a second phase, children can modify the different parameters of the model, in order to explore and understand the impact on CO_2 emissions. I.e., they can freely move the tokens to see how much the CO_2 value is increased or decreased related to each parameter. This phase has unlimited time and ends when the children have specified the model: they are provided with paper clouds of different sizes. They need to identify which parameter generates the high, low, and no CO_2. They place the cloud of the corresponding size next to the parameter. This phase aims at measuring the *knowledge acquisition*, i.e. the amount and correctness of explicit knowledge gathered during the exploration.

In the third phase, all tokens are first replaced according to the As-Is situation. Children are now asked to solve the new task, using the knowledge acquired in the previous phase. This phase aims at measuring *knowledge application*, i.e., the children's capacity of generating and acting out a solution. They are asked to reach a provided level of CO_2 emissions by moving as less tokens as possible.

Data is gathered from 20 groups of children aged 8 to 10. They will explore and solve the presented issues of climate change on the table. Audio and video logs will be recorded of the sessions and children will be asked to provide information by questionnaire to measure their gain in understanding on climate change.

RESEARCH AGENDA AND CONCLUSION

The construct for individual complex problem is well defined. This counts for both, cognitive and collaborative dimension. Nevertheless, assessing complex problem solving in an collaborative setting is new and needs further investigations. In addition, the use of tangible user interfaces for assessment is highly innovative but has only been marginally investigated so far. These research gaps lead to a research agenda with three main axes:

1. *Derive the construct for complex collaborative problem solving*: The interaction between complex problem solving skills and collaboration skills is unknown. Understanding the interplay of both skills requires additional studies. Task, such as those based on Microworlds (e.g. MicroDYN) which are proven to be valid and reliable for individual assessment, need to be adapted and redesigned for collaborative settings.
2. *Develop and evaluate new assessment instrument*: Another challenge is that on the one hand, we need standardised assessment instruments but on the other hand we need to provide enough flexibility to allow human-to-human collaboration processes. Tangible user interfaces have shown that collaborative knowledge acquisition and knowledge application activities are supported, but additional work is necessary to use them as a quality assessment instruments. It is apparent, that TUI cannot be used for large scale studies. Further, studies on usability and user experience are necessary to continuously improve the instrument.
3. *Validate the instrument*: Many dimensions such as gestures, speech, (tangible) interactions and localisation data can be gathered, but, as stated in the assessment literature, the challenge is to relating it to assessing ColPS? Solving strategy patterns need to be extracted to gain understanding about the different phases during knowledge acquisition and application (e.g., experimental units conducted in the simulation environment). Which indicators (e.g., pointing gestures, speech patterns, physical manipulations)are useful? Which qualitative or quantitative analysis should be applied?

By following this research agenda, our research will provide a deeper understanding on the ColPS construct and how TUIs can serve as assessment instrument for assessing 21st Century Skills that are inherently hard to assess using traditional approaches. Besides research on collaboration skills and on individual complex problem solving, the fields of inquiry-based learning and simulation in learning play a major role (Van Joolingen, De Jong, & Dimitrakopoulou, 2007). Outcomes from this research domain can be used to understand the experimentation process in a simulation environment or, for example, to prevent students from engaging in extensive, aimless play sessions on the TUI system. In collaborative testing, factors such as decreased anxiety, rich discussions, supported cognitive processes (e.g. retrieving information, thinking through the information better, etc.) have led to an improvement in test performance (Kapitanoff, 2009). Such indicators will also be of interest in a TUI context in the future.

The definition of a valid and reliable construct for ColPS is new, as is the use of TUI for assessing this skill. Their combination might prove a useful tool for assessing other 21st Century Skills (e.g., creativity, team work etc.) as well due to the novel perspective the system will offer on the domain.

REFERENCES

Bennett, R. E., & Gitomer, D. H. (2009). Transforming K-12 assessment. In C. Wyatt-Smith & J. Cumming (Eds.), *Assessment issues of the 21st Century*. New York: Springer Publishing Company.

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

Cannon-Bowers, J. A., & Salas, E. (1997). Teamwork Competencies: The Interaction of Team Member Knowledge, Skills, and Attitudes. In H. F. O. Neil & J. Mahwah (Eds.), *Workforce Readiness: Competencies and Assessment* NY: Lawrence Erlbaum.

Esteves, A., Van den Hoven, E., & Oakley, I. (2013, February 10-13). *Physical games or digital games?: Comparing support for mental projection in tangible and virtual representations of a problem-solving task.* Paper presented at the 7th International Conference on Tangible, Embedded and Embodied Interaction, Barcelona, Spain.

Ferrari, A. (2012). Digital Competence in Practice *JRC Technical Reports*. Seville: European Commission, Joint Research Centre, Institute for Prospective Technological Studies

Fischer, A., Greiff, S., & Funke, J. (2011). The process of solving complex problems. *Journal of Problem Solving, 4*(1), 19-42.

Funke, J. (2001). Dynamic systems as tools for analysing human judgement. *Thinking and Reasoning, 7*(69-89).

Graesser, A. C., Jeon, M., & Dufty, D. (2008). Agent technologies designed to facilitate interactive knowledge construction. *Discourse Processes*, *45*, 298-322.

Greiff, S., Wüstenberg, S., Molnar, G., Fischer, A., Funke, J., & 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., Care, E., & McGaw, B. (2011). The changing role of education and schools. In P. Griffin, B. McGaw & E. Care (Eds.), *Assessment and Teaching 21st Century Skills*. Heidelberg: Springer.

Hellinghausen, M. A., & Myers, J. (1998). Empowered Employees: A New Team Concept. *Industrial Management 40*(5), 21-23.

Hornecker, E., & Buur, J. (2006). *Getting a grip on tangible interaction: a framework on physical space and social interaction*. Paper presented at the SIGCHI Conference on Human Factors in Computing Systems 2006 (CHI 2006), Montreal, Quebec, Canada.

IPCC Working Group I. (2013). Climate Change 2013: The Physical Science Basis.

Ishii, H. (2008). *Tangible bits: beyond pixels*. Paper presented at the 2nd international conference on Tangible and Embedded Interaction, Bonn, Germany.

Kapitanoff, S. H. (2009). Collaborative testing: Cognitive and interpersonal processes related to the enhanced test performance. *Active Learning in Higher Education*, *10*(56). doi: 10.1177/1469787408100195

Klemmer, S. R., Hartmann, B., & Takayama, L. (2006). *How bodies matter: five themes for interaction design*. Paper presented at the 6th Conference on Designing Interactive Systems (DIS 2006), University Park, PA, USA.

klimAktiv gemeinnützige Gesellschaft zur Förderung des Klimaschutzes mbH. (2014) Retrieved June 27, 2014, from http://uba.klimaktiv-co2-rechner.de/de_DE/page/impressum/?ref=living-pt

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

Lai, E. R. (2011). *Collaboration: A Literature Review*. Boston: Pearson.

Mayer, R. E., & Wittrock, M. C. (2006). Problem Solving. In P. A. Alexander & P. H. Winne (Eds.), *Handbook of Educational Psychology* (pp. 287-303). New York: Lawrence Erlbaum.

Novick, L. R., & Bassok, M. (2005). Problem solving. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 321-349). Cambridge, NY: University Press.

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

OECD. (2012). Better skills, better jobs, better lives. A strategic approach to skills policies. Paris: OECD Publishing.

OECD. (2013). PISA 2012 assessment and analytical framework mathematics, reading, science, problem solving and financial literacy. Paris: OECD Publishing.

OECD. (2014). PISA 2015 Draft Collaborative Problem Solving Framework Retrieved June 27, 2014, from http://www.oecd.org/pisa/pisaproducts/Draft%20PISA%202015%20Collaborative%20Problem%20Solving%20Framework%20.pdf

Ras, E., & Maquil, V. (2011). *Preliminary results of a usability study in the domain of technology-based assessment using a tangible tabletop*. Paper presented at the Workshop Usability in Education (U-Ed) at the IHM 2011 conference, Sophia Antipolis, France.

Ras, E., Maquil, V., Foulonneau, M., & Latour, T. (2013). Empirical Studies on a Tangible User Interface for Technology-based Assessment - Insights and Emerging Challenges. *International Journal of e-Assessment (IJEA), CAA 2012 Issue: Pedagogy and Technology: Harmony and Tensions, 3*(1).
Rummel, N., & Spada, H. (2005). Learning to collaborate: An instructional approach to promoting collaborative problem solving in computer-mediated settings. *Journal of the Learning Sciences, 14*(2), 201-241.
Van Joolingen, W. R., De Jong, T., & Dimitrakopoulou, A. (2007). Issues in computer supported inquiry learning in science. *Journal of Computer Assisted Learning, 23*(2), 111-119. doi: 10.1111/j.1365-2729.2006.00216.x