TÁMOP 3.1.9-11/1-2012-0001
„Developing Diagnostic Assessments” project

5TH SZEGED WORKSHOP ON EDUCATIONAL EVALUATION

ABSTRACTS

15-16 April 2013

Venue:

Szeged Committee of the Hungarian Academy of Sciences

7, Somogyi Street, Szeged

Institute of Education
Graduate School of Educational Sciences
Hungarian Academy of Sciences

ABSTRACTS
15th April 2013

Session A

Samuel Greiff and Sascha Wüstenberg	Problem Solving: Current State of the Art and What There is to Come

This presentation gives a brief introduction on problem solving. It reiterates definitions of different types of problem solving (i.e., Complex Problem Solving, Collaborative Problem Solving, and Domain-Specific Problem Solving), reports recent research findings, and shortly discusses the relevance of problem solving for research, education, and policy. It serves as an introduction and a preview of the talks from the Luxembourg group and connects the different presentations to each other.

Christian Wolff, Samuel Greiff, Jonas Müller and André Kretzschmar	Beyond Reasoning? Searching for the Construct of Complex Problem Solving

Tests that assess Complex Problem Solving (CPS) differ strongly from reasoning tests. The most recent type of CPS tests – so called multiple complex systems (MCS) – has shown good psychometric properties. Moreover, the criterion validity of these tests goes beyond reasoning or working memory capacity in predicting real-life criteria. However, not much is known about the relationship between CPS and reasoning. Objective: To distinguish the construct of CPS conceptually and empirically from reasoning. Method: N = 339 university students participated in the study. They completed three CPS tests which follow the new MCS approach (MicroDYN, MicroFIN and Genetics Lab) and another CPS test representing the classical microworld approach (Tailorshop). We included the matrices subtest from the Intelligence Structure Test-Revised to measure reasoning. To test our hypothesis, we used structural equation modeling methodology to compare two models. In the restricted model, we specified the correlations within CPS tests and the correlations between CPS and reasoning to be equal. In the less restricted model, the correlations within CPS measures could be higher than the correlations between CPS and reasoning. Results: We found that the construct of CPS can be distinguished from reasoning. The correlations within CPS tests were higher than the correlations between CPS and the reasoning test. Indeed, this result only applies to CPS measures that are constructed following the MCS approach, but not to the classical microworld Tailorshop. Discussion: Our results suggest that CPS can be understood as a discrete construct of its own that can be distinguished from reasoning, given that the employed CPS tests follow the MCS approach. In the future, we need to find out how CPS can be related to existing theories of intelligence and cognition.

André Kretzschmar, Jonas Müller, Christian Wolff and Samuel Greiff	Complex Problem Solving Within the Lifelong Learning Project

In 21st century, non-routine skills become much more important whilst routine skills are decreasing. Thus, continuous knowledge acquisition and lifelong learning are essential to cope with new situations and demands especially in working life.

The LLLight'in'europe project (www.llllightineurope.com) seeks to clarify the framework conditions and determinants of lifelong learning. The corporate strategies, public policy environments and success as innovations, economic returns as well as skill acquisition of 60 companies and over 4000 individuals from different countries and different cultural backgrounds will be analysed. As a marker for general learning activities on an individual level the ability of complex problem solving (CPS) will be assessed.

We will present the actual progress of the project, demonstrate und discuss different operationalisations of CPS used in the project, and take a look at first empirical results concerning the CPS assessment.

Session B

Gidon Frischkorn, Samuel Greiff and Sascha Wüstenberg	The Development of Complex Problem Solving and its Relation to Reasoning, Age and Gender

Complex Problem Solving (CPS) has gained increasing attention in educational psychology in recent years as indicated by its implementation in international educational Large Scale Assessments such as the Program for International Student Assessment (OECD, 2010). However, research on the development of CPS is scarce, and the few existing studies are cross-sectional (Molnár, Greiff, & Csapó, in press). Therefore, we conducted a longitudinal assessment of CPS over two years with adolescent students to analyze CPS development more closely. CPS was measured with the MicroDYN approach (Greiff, Wüstenberg, & Funke, 2012) and CPS development was estimated with latent growth curve models. Further, reasoning, age, and sex served as predictors for CPS development. Results displayed positive linear growth patterns for CPS, which were positively related to reasoning. Additionally, older students showed better performance in initial CPS but exhibited smaller increases (i.e., less development) in CPS performance. No meaningful sex differences in initial CPS or CPS development were found. These results present an important step in investigating CPS development as CPS is highly relevant for education today.

Julia Rudoplh, Katarina Krkovic and Samuel Greiff	Domain Specific Problem Solving- Bringing Complex Problem Solving Research to School

The added value of domain-general complex problem solving skills (CPS) has been reported in a number of studies (e.g., Greiff & Fischer, 2013). Domain-general CPS is known to have an effect on scholastic achievement (Schweizer, Wüstenberg & Greiff, 2013). However, even though the development of competencies usually takes place within a domain (Friege & Lind, 2003), research on domain-specific CPS is scarce at the best. The effect of domain-specific CPS on scholastic achievement, its interaction with previous knowledge, and connection to motivational constructs are unsolved issues that need to be investigated to better understand scholastic achievement. The project PLaSA (Problem Solving, Learning and Scholastic Achievement) is aimed at closing those knowledge gaps with a focus on mathematical and physical CPS. The major research questions are: (1) How are static problem solving, domain-general CPS and domain-specific CPS connected? (2) Which role plays domain-specific CPS in scholastic achievement and grading? (3) To what extent does domain-specific CPS dependent on interest and vice versa? (4) How strong is the impact of elaboration strategies on domain-specific CPS? The presentation introduces the theoretical embedding of the hypotheses, the conception of domain-specific CPS, and a longitudinal research design within PLaSA. Anticipated obstacles and strategies to overcome them will be discussed.

Katarina Krkovic, Julia Rudolph and Samuel Greiff	Collaborative Problem Solving: Current Assessment Possibilities and Issues

Over the last decades the world has experienced rapid changes and globalization, causing a shift in working conditions and requirements: increasing complexity of tasks at work and in everyday life, expansion of tasks towards non-routine problems, and increasing need for collaboration (Autor, Levy & Murnane, 2003). However, our schools still do not teach us the 21st century skills necessary to tackle this shift – how to be creative, how to solve complex problems, or even more important, how to collaboratively solve problems. The first and the most important step towards implementing these skills into the educational system is to understand their theoretical background and to be able to appropriately assess them in individuals`. In our research we aim at

developing an optimal solution on how to understand and assess Collaborative Problem Solving (ColPS) as one essential skill in 21st century. In order to provide a sound measurement device for ColPS, we first need to make major decisions on how we intend to approach ColPS and to plan how to overcome possible issues of each approach. Major questions arising are: 1) Should the assessment relay on human-to-human or human-to-agent interaction? 2) How should the group composition look like – is a group of two enough? 3) How can the interaction be designed – open chat, limited chat, or video chat? 4) What test material can adequately depict both Collaboration and Problem Solving in a same task? These and other possible issues of ColPS assessment will be addressed and discussed.

Session C

Benő Csapó	Progress in the Developing Diagnostic Assessment Project

The second phase of the Developing Diagnostic Assessment project is devoted to make the online platform fully functional, to build large item banks, pilot them and to prepare the entire system to be used in the everyday school practice. This presentation reviews the main advancements that have been taken in the last year towards these goals at the different areas of the project. A new online platform has been installed allowing innovative item formats and being capable of serving more students. The framework development continued by revising and further elaborating the description of the content of assessment at the three main domains (reading, mathematics and science), with more examples, taking into account the possibilities of the platform. Exploring the possibilities of online assessment at fourteen further (minor) domains are in progress, online tests or test batteries have been prepared and piloted. The major part of the work is writing and calibrating several thousands of items for the main domains. In order to do this, a large number of item-writers (more than 300 so far) have been trained and a network of (with c.a. 300) partner schools has been set up, who participate in the pilot testing. Initial steps have been made towards developing training materials (e.g. computer games) for students who will be identified by the diagnostic system as lagging behind.

Gyöngyvér Molnár	*New Features of the eDia platform*

The main aims of the presentation are to give an overview about the (1) main features, (2) usage requirements and (3) developmental challenges regarding the eDia (electronic Diagnostic Assessment) platform. In the last year more than 30 thousand 1 to 6 Graders have been tested by the platform. A general browser and internet connection are sufficient for its use in classroom instruction and in school. The eDia supports a greater variety of item types. For example, children are not only able to read the instructions to the items, but also to listen to them, so the tests can also be used among learners who cannot read yet or who suffer from reading difficulties. In addition to sounds, images, animation, and video lend the items a richer and more real-world quality. Learners are able to give answers to the items in different forms; for example, they can mark, click, colour, move items, or rearrange them. The eDia system is therefore suited to mediating not only simple multiple-choice items, but also item types that make maximum use of the new options provided by technology. This opens doors to studying knowledge and skill domains researched previously in an innovative, new environment, which is more motivating for the learners (e.g. language abilities and musical literacy), as well as exploring features of skills not studied yet (e.g. dynamic problem-solving ability). By making use of the options in electronic testing, the eDia system is capable of supplying rapid feedback on learners' knowledge and skill levels. With its state-of-the-art technology and web applications the platform opens new perspectives for transforming classroom activities and effectively assisting teachers in their day-to-day instructional work and in carrying out assessment tasks.

Michel Dorochevsky	Project Introduction: Interactive Problem Solving and Lifelong Learning

The CBA Item Builder platform developed for the DIPF has recently been used in large scale assessment of complex problem solving items at the University of Helsinki with a sample of more than 6.000 pupils doing the tests over internet in a short timeframe of 4 weeks. We report about the challenges and lessons learned during the testing and the key-factors for a scalable testing system. Our conclusions clearly indicate, that moving to a cloud based delivery is within reach – making data collection more efficient and scalable in the near future.

Frank Goldhammer	Construct Validation of the ICT Literacy Scale TEO: Task Characteristics and Progress Measures

Technical devices and the Internet have become essential and ubiquitous parts of our daily life. ICT literacy enables to address information problems using information- and communication technology (ICT). As a general condition for solving information tasks in educational and workplace settings successfully, ICT literacy represents an important transversal and cross-curricular skill. This study focuses on the theoretical and empirical fundamentals of one major dimension of ICT literacy, namely evaluating online information as the cognitive skill to efficiently make judgements about the credibility of online-information taking various features of a website into account. For this study we implemented an innovative measure using simulation-based interactive tasks. Construct validity was investigated by the common approach to test the influence of those item characteristics on task success that were expected to influence targeted processes of evaluating online information. Second, based on logfile data process measures representing individual differences in evaluating processes directly were derived, and their influences on task success were investigated as well. A sample of N=400 students completed the scale Test for Evaluating Online Information (TEO). Using linear mixed models, effects of task characteristics that governed item development as well as effects of process measures could be shown.

Session D

Carolin Hahnel	Construct Validation of the ICT Literacy Scale TEO: Identifying the Nomothetic Span

Search engines became omnipresent at the World Wide Web. Often it's necessary to deal with them in order to get access to information in educational, occupational, and private settings. But they offer tons of different information of different quality and trustworthiness, thus, there is a need to identify and select relevant information in relative short time. As a special part of ICT literacy, evaluating online information as a cognitive skill focuses on judging efficiently about the credibility of online information in regard of structural and message based features of such search engine results. In this study construct validity of evaluating online information is investigated in terms of nomothetic span, means, a network of relationships of the used test score with other variables is constructed and tested in order to analyze the structure of evaluating online information. A total of 879 15-year-old students who participated in a computer based German add-on study in the context of PISA 2012 were assessed. Beside the Test for Evaluating Online Information (TEO) they completed scales on their reading skills (word, sentence, and text level), basic computer skills, and working memory capacity within a randomized design. First results show that especially reading literacy on a text level and basic computer skills are important influences on judging about credibility of search engine results. Furthermore, effects of working memory capacity and basic reading skills on word and sentence level seem to be mediated via reading comprehension. Structural differences are also investigated according to specific construct facets.

Jukka Marjanen, Sirkku Kupiainen	*The Difference Between Computer-Based vs. Paper-and Pencil Testing Now and 10 Years Ago- the Case of Finnish Learning to Learn Assessments*

There is an increasing endeavor on many fronts of assessment to move from the traditional paper-and-pencil tests to computer-based testing. In this process the main question is whether the results acquired via different means of testing are comparable for all test subjects. In this presentation the comparability of these two testing devices are examined using data from learning to learn assessments conducted in 2001 and 2010. The analysis indicates that the gap between paper-and-pencil vs. computer-based testing is now considerably smaller than 10 years ago. However on some tasks the difference is still too large to be left without consideration.

Mari-Pauliina Vainikainen, Jarkko Hautämaki	*Adaptations to Needs- Resonse-to-Tasks*

In this presentation possible future paths of learning to learn assessment and intervention are discussed by combining the two relatively separate research strands of the Centre of Educational Assessment at the University of Helsinki with the possibilities log-file analysis could provide for understanding the needs of students and differentiating teaching to tackle the growing between-student variation in school classes. The model for providing support for students in Finland has been changed lately in a partial educational reform, and the division to general education and special education has been replaced by a three-tiered model for providing intensifying support based on continuous assessment. In this model, computer-based assessment and intervention could be utilized to a much greater extent if more adaptive methods were available. Based on initial analyses of log data of a large-scale learning to learn assessment in 2012, ninth graders in need of support seem to differ from their classmates in task performance and not only in their test scores. These initial results indicate that process data analysis could be useful in planning computer-based interventions for learning to learn skills. In Finnish learning to learn tasks only a limited set of process data is collected - however, this spring more than 6000 students have also done MicroDyn-tasks as a part of learning to learn assessment, and combining the more rigorous process data analysis with the information about the students' special education needs could open new possibilities in understanding how learning to learn skills of these students could better be supported and enhanced.

Session E

Ágnes Hódi, Tímea Török, Renáta Kiss	New Opportunities in Technology-Based Assessment of Reading
Benő Csapó and Gyöngyvér Molnár	Comparing Face-to-Face and Technology-Based Assessment of School Readiness

There is a growing need for instruments assessing school-readiness that are easy to use. At an age when pupils still cannot read and write paper-based testing is not an option, and the objectivity and applicability of traditional individually administered face to face tests are limited. Reflecting to these needs we have begun to develop online instruments measuring school readiness and monitoring early development of several basic skills. This presentation demonstrates the potentials of using computer based online tests at a very early age when students still cannot read and write, as well as that technology not only provides additional features but essentially improves the applicability of assessment. In the first phase, five tests (speech sound discrimination, relational reasoning, counting and basic numeracy, inferential reasoning and deductive reasoning) of an already existing face to face administered school readiness test battery (DIFER) were transplanted into an online platform (eDia) with almost identical contents. Both the face to face and the online versions of the tests were administered to samples of about 400 first grade students. Results show that in the case of four tests the online versions show improved reliability. Only some items of the counting and basic numeracy test could be assessed via computer and the reliability of this shortened test decreased. A more detailed item analysis of the speech sound discrimination test using IRT scaling indicated improved objectivity of the computerized version.

Muriel Foulonneau, Eric Ras, Thibaud Latour and Christian Gütl	Generating High Quality Assessment Items From Multimedia Web Resources

The creation of assessment items is an expensive and heavy task. With the development of computerized adaptive testing as well as formative assessment and informal learning environments, the quantity of assessment items necessary is continuously growing. Manual approaches are not scalable and more and more attention is put on the semi-automatic generation of assessment items. We are developing algorithms to adapt these approaches to the Web environment with the generation of resources from Linked Open Data as well as from textual resources published on the Web, with mechanisms dedicated to item quality. We propose processes to generate items from both textual and semantic resources from the Web and approaches to the generation of distractors in multiple choice items and match items. When generating items from textual resources, distractors are created using natural language processing techniques such as latent semantic analysis and stylometry. With semantic resources, we use semantic similarity metrics in order to increase the quality of distractors. We propose a metric to represent item homogeneity and show its impact on item difficulty. We also propose a clustering method to create match items. We investigate the use of multimedia resources in the item generation process. The generated items are compatible with the IMS-QTI standard. We show the validation of our models with users and raise issues for the future to combine approaches for both semantic and textual resources for the generation of high quality items, including multimedia resources.

Session F

Eric Ras, Valérie Maquil and Thibaud Latour	Assessing 21st Century Skills on an Interactive Tabletop

Most of the research in technology-based assessment (TBA) dealt with the improvement of assessment of traditional skills, but since a few years so-called 21st Century skills challenge the field of TBA. Further, new technologies influence the way people learn, work and live today also hence let research and practitioners adapt the definition of related skills (e.g., digital literacy). During the last years, topics such as measuring solving strategies (i.e., measurement of dynamics in a test) and assessing collaborative skills have got more attention.

One the one hand, new technologies, such as natural user interfaces provide new opportunities to assess skills but on the other hand, they are bring new risks biasing assessment results, which we need to identify and understand first. Therefore, we have conducted several empirical studies to understand how users interact with so called tangible user interface (TUI) during an assessment task.

A simple matching item, suitable for measuring recall of factual knowledge, as well as a simulation item with the potential to assess higher order thinking skills were used for these studies.

Besides the evaluation of the usability, respectively, user experience (UX), we interviewed experts to derive a list of tensions and also advantages for using such a technology for assessment. Some tensions mentioned by the experts were related to the first phase of getting used to the system: When should the system start to assess the observed solving strategies? How could the system identify single atomic contributions (i.e. single experiments with the simulation parameters)? How could we include activities outside the interactive surface of the table? The table supports the user to 'recognise the perspective of others', an important sub-skill of collaborative problem solving – how can the table track these activities? Based on the outcomes of the studies we identified eight topics for which we discuss related tensions and advantages between the new technology and technology-based assessment. They will impact the future development of using TUI for assessment as well as the design of assessment models and methods.

Ingo Barkow, Hewiko Rölke, Krisztina Tóth	Log File Extractor – a Tool to Handle Output From Diagnostic Assessment Systems

Current assessment systems do not only produce result files as tables, but also record the interactions of the students (e.g. clicks, changes of answers, visit of certain pages) as log files including time stamps. These log files can be used to do further analyses like data mining or process mining to get more fine grained results or as basis for predictions. Unfortunately there is no standard format for those log files and therefore converting them into data formats which can be handled by statistical packages is a tedious and time consuming process. To support projects like Prokom where huge amounts of log files from the PIAAC study are analysed the German Institute for International Educational Research (DIPF) in co-operation with OPIT Consulting Kft. is currently developing a tool - the Log File Extractor - to handle different log file inputs and convert them into meaningful data for post processing. The tool also supports configurable data cleaning and conformity checking, selectable items for customizable extraction, data extraction with .csv conversion, reporting, pre-analysis on test taking process via tables and diagrams and visualization. The tool is scheduled to be released in December 2013.

Detlev Leutner	Educational Knowledge of Young Teachers at the Very Beginning of their Work at School

In Northrhine-Westfalia, the largest federal state of Germany, a complete cohort of 3.258 young and beginning teachers was tested on their educational knowledge acquired in their university studies. In the presentation I will describe how the test instrument was developed, and I will report some first results on its validity.

Jean-Paul Reeff	*National Occupational Skills Standards and Related Assessment Challenges. The Saudi-Arabia Case*

The Kingdom of Saudi-Arabia (KSA) is facing a number of societal and related educational challenges. Consequently and for several years already, KSA invests at all levels of the education system, and aims at creating "world-class" education and training conditions in the country. The focus of this presentation will be on recent and future-oriented efforts in vocational education and training (VET).

Enrolment in vocational education and training in the Kingdom is at about 8% of the relevant Saudi cohort, as compared to more than 40% in many developed countries. KSA authorities aim at raising this percentage substantially over the next 15 years by increasing both the quantity of vocational education and training offers and by improving the quality of these training measures.

One central element in the VET strategy of KSA is an improved fit between VET and labor market requirements. As a first step, KSA VET authorities - in close cooperation with Saudi employers - have started a major effort to analyze these requirements, to describe them precisely at an occupation level and to use them as a binding reference for subsequent vocational education and training. In a first step, they developed National Occupational Skills Standards (NOSS) for 50 occupations and will develop NOSS for another 200 occupations from 2013 on.

These NOSS are the basis of a substantial number of new vocational training colleges, to be implemented by foreign training providers. Using NOSS as an outcome reference, training providers will have far-reaching freedom with regard to curriculum implementation and training innovation. However, a newly created national agency will evaluate education and training outcome quality with a newly to be developed NOSS-directed, technology-based assessment.

The presentation will outline the current VET efforts in KSA; describe the NOSS development paradigm and discuss the challenges of NOSS-directed, technology-based assessment in authentic work-place environments.

Based on the KSA case, options for international cooperation in VET and related assessment will be presented and discussed.

Session G

Kata Asztalos	Online Diagnostic Testing of Musical Hearing Abilities in Primary School

Technology based assessment is a rapidly developing area, which can be extended to musical abilities as well. To use objective tests in music education is basically important because the current form of assessment is highly subjective and informal. The enrichment of the evaluating culture could be a solution for this problem.

In this presentation, we demonstrate the advantages of technology based assessment at the area of musical abilities and summarize the first results of a pilot study. Participants of the study were first (N=219), third (N=207) and fifth grader (N=227) primary school students selected from schools of Csongrád county. The distribution of the results indicated that the difficulty of the test fitted well to the ability level of students, the means for the entire test were 43%, 58% and 63% for the three age groups. The sample contained students from general elementary schools and from a music elementary school. The instrument of the assessment was an online test administered through the eDia platform. The items of the test had several innovative features impossible to realize on paper. Students received the instructions and musical stimuli through headphones. The musical stimuli cover a wide range of abilities from the successive and simultaneous dimensions of music. We collected information about melody, rhythm, tempo, dynamic, harmony, pitch and timbre discrimination, tonality hearing and visual connection. Short melodic excerpts were selected from among classical music and Hungarian folk music pieces. The musical test was accompanied by a brief questionnaire. Students took the tests in their own schools. Results demonstrate that students from music elementary school achieved significantly better than their peers in normal schools, and their development is faster.

The online technology-based musical ability test could be a useful supplement of the high-quality music pedagogical work and diagnostic information can facilitate the focused improvement process. Further research is required for examining the predictive and diagnostic value of the assessment of musical abilities.

László Hülber	*Transition to Online Testing*

The introduction of technology-based testing methods bears various development potentials, resulting in a paradigm shift in measurement-assessment processes. However, before paper-and-pencil (PP) procedures are replaced through computer-based (CB) ones, a number of comparative analyses must be conducted to establish in what ways and to what extent results are influenced by the introduction of the new medium. No learner groups may enjoy positive or suffer negative discrimination due to the migration other (Molnár, 2010; Lent, 2009). Therefore, in order to assess all media-related impacts, I carried out a large-sample comparative analysis of paper-and-pencil and computer-based testing procedures in the field of mathematics within the less intensively researched focus group of primary school students (grade 1 to 6), relying on a wide spectrum of item types. The analysed parameters were mainly focussed around the processing of task-related information, the psychological structures activated and the type of approach applied while solving the individual problems. The reliability indices (Cronbach-α) show acceptable values with both paper-and-pencil (≥ 0.81) and computer-based (≥ 0.84) procedures. There appears to be no significant difference between the students' performances via the two media ($t=2.14$, $p=0.45$), whereas the correlation coefficient of the item difficulty values has been found to increase with each grade ($r1=0.7..r6=0.9$). No statistical difference was shown between the results achieved in the case of open-ended and closed question types (ANOVA, $F=1.16$, $p=0.43$). The only significant media impact, associated with various parameters, has been identified in the first two grades. Based on the findings, online testing procedures appear to be reliable and applicable with primary school students. In order to minimize eventual differences, both the forms of task representation and the technological parameters must undergo comprehensive regulation. The resulting standards shall highly reduce all differences related to differing test structures. Differences arising from varying practical experience with information-technological tools can be reduced by means of a tutorial program aimed at the basics of mouse and keyboard handling. A well-aimed tutorial program, however, including an application to test the practical use of the equipment, could provide the learners with basic practical knowledge and reduce the influence of preliminary computer use on the results. For a more comprehensive mapping of the online testing environment further comparative analyses are required with a focus on text production as well as other fields and subjects.

István Thékes	The Correlations Between the Processing of English as a Second Language

The purpose of the present study was to explore the correlations between English as a second language (ESL) idiom knowledge, native language idiom knowledge, inductive reasoning skills, metacognitive skills and effectance motivation. Wray (2000) gives more than twenty terminologies for 'idioms'. I am using the following words interchangeably: 'idioms', 'chunks, 'collocations'.

In the past twenty years ESL vocabulary research has turned its focus on learners processing chunks (Wray, 2001). Several researchers (Nattinger, deCarrico, 1992; Lewis, 1993) argued for the teaching of idioms as single morphemes. Wray and Fitzpatrick (2009) assert that idioms are such linguistic components that the context-dependent use of theirs makes it evident whether the speaker is native or non-native. Siyanova-Chanturia et.al (2011) studied the processing of collocations by natives and non-natives with an eye-tracking apparatus and found that natives are significantly faster than their non-native counterparts in comprehension. Csapo and Nikolov (2009) studied the correlations between ESL proficiency and inductive reasoning; however no research had studies the metacognitive, motivational and inductive factor of idiom processing. I hypothesized that all these above-mentioned factors have significant impact on ESL idiom

	processing.
Attila Pásztor	Developing an Online Assessment Tool for Creativity- Challenges and Possibilities

The significant role of creativity in modern society is undisputed. The ever more rapid economic, social, and technical development we see today requires new and original ideas and solutions, while creativity is indispensable for success in a wide range of jobs in the 21th century (Florida, 2004).

One way of preparing for the unpredictable and new challenges is the improvement of creativity, in which schools might play a central role (Binkley, Erstad, Herman, Raizen, Ripley & Rumble, 2012; Newton, 2012). However, in order to increase creativity diagnostic assessment tools are necessary for setting up precise diagnosis.

Research conducted in the field has proved that creativity is an extremely complex phenomenon and there are many approaches to studying creativity (Sternberg, 1999; Runco, 2007). Therefore numerous tools exist in order for different dimensions of creativity to be assessed (Lemons, 2011); however only few focused on the possibilities of the online assessment of creativity (Lau & Cheung, 2010; Villalba, 2009).

The aim of our research is to develop an online assessment tool for creativity which can also be used in large scale assessments.

The test consists of 54 items and is based on the instruments of Torrance (1966), Wallach and Kogan (1965) and Mednick and Mednick (1967) aiming to measure divergent thinking (Guilford, 1959). The pilot study takes place among 10-12 year old students. The online data collection is carried out with the eDIA (Electronic Diagnostic Assessment) platform via Internet. Answers are scored by the generally used scales in the measurement of creativity: fluency, flexibility and originality. Background variables are also examined such as different thinking skills (e.g. inductive reasoning), academic achievement and social background.

As a result of our pilot study the possibilities of online assessment of creativity are explored. On the basis of the database generated by the pilot study we also investigate the possibilities to develop an online evaluating system which is able to score students' answers automatically in order to reduce time and cost of the testing process and make the assessment tool suitable for large scale assessments.

Andrea Magyar	Comparative Study on Computerized Adaptive Testing and Fixed Item Testing

Computerized adaptive tests (CAT) are one of the most sophisticated computerized methods for assessing the level of abilities of examinees. During the testing process the test is adapted to the examinees' ability level by administering only items of appropriate difficulty for the examinees. An often used type of CAT is the multistage adaptive test (MST), in which modules are administered instead of items. Several researches have compared the measurement precision of the MST and fixed item tests, but only some of them concentrated on testing young students in computer-based environment.

The aim of this study is to present the results of an empirical study compared the functioning of fixed and multistage adaptive tests among young children. The research was conducted among 10-14 year-old primary-school students (N=158). A fixed inductive reasoning test was administered to half of the students; the other part solved a four-stage adaptive test (1-3-3-3 structure). According to the results, the reliabilities of the two tests were very similar, and there was no significant difference regarding the mean and the standard deviation. The measurement precision was estimated by the test information functions (TIF). Regarding the ability levels, on average ability level there was no difference between the amounts of information of the two tests, but the MST solved more information at the lower and upper ability levels.