

# 7TH SZEGED WORKSHOP ON EDUCATIONAL EVALUATION

27-28 APRIL 2015

## ABSTRACTS

Venue:

Szeged Committee of the Hungarian Academy of Sciences

7, Somogyi Street, Szeged

Institute of Education  
Doctoral School of Education  
Hungarian Academy of Sciences





# ABSTRACTS

27<sup>th</sup> April 2015

## Theoretical Foundations of the Assessment Based Improvement

### Session A

<i>Jarkko Hautamäki</i>	<i>Finnish Studies on Development and Advancement of Thinking at School</i>
<p>David Olson gives in his book title ‘Psychological Theory and Educational Reform’ an opening to the topic: the dilemma between psychology and schooling. The knowledge of the development of thinking is a different thing than to be able to use that knowledge in teaching for thinking. Also, we may be able to show in intervention studies an ideal good outcome – to show that an intervention had a true effect – but to scale that up to a fundamental and true feature of a national system of schooling is another thing. But, still, teaching and learning matters: Roger Titcombe is using the title ‘Learning Matters. The Truth about our schools’ and discusses of bad and good education. So, I will describe my (personal) understanding how to provide a good education, in Finland and elsewhere, as well as a kind of a truth of Finnish school education.</p> <p>The empirical story will be started by comparing two pieces of information. First, I will report our recent result from comparing learning-to-learn outcomes from 2001 and 2012 (9th graders, representative samples), which showed a decline of scores (effect size about .5) for competence and beliefs. Second, I will report an older study where we could with Kuusela to show that it is possible to increase permanently cognitive development at least .5 SD. In this study we applied, with a fully randomized design, the effects of Cognitive Acceleration on Science Education, CASE and Cognitive Acceleration in Math Education, CASE (Adey, Shayer). The surprising outcome was that also control students learned with lasting effects to think on higher Piagetian level. CASE and CAME and out measures were founded on Piagetian theory, i.e., Inhelder and Piaget cognitive schemata.</p> <p>This is the background why we in our CEA studies use such cognitive tasks which allow for matching between psychological theory and school subjects. There are reasons why we will argument for applicability of Piagetian formal operational thinking. This option is referred to as general plastic ability or general problem solving skills. But, in order to take also the scaling up seriously, we need to study development of thinking within schooling. This means that we have to provide knowledge of how municipalities, schools and classes take their parts in the meeting the variance of full age cohorts. Also other conceptual models are needed to build the conceptual bridge from psychological theory and education practice. Some of these will be given, as well as some interesting empirical results, to pave the way to future studies and co-operation.</p>	

<i>Benő Csapó</i>	From Content-Based to Assessment-Based Improvement of Reasoning Skills
<p>This paper summarizes the results of a research on improving reasoning skills carried out in the past decades, identifies the major constraints that limited the implementation of the promising outcomes of small-scale experiments in real educational settings and outlines the new directions of development made possible by the application of online technologies.</p> <p>The dissatisfaction with the traditional school curricula based on declarative knowledge (facts and figures of the particular scientific disciplines) became manifest in the 1970s in American educational research, followed by a number of studies and publication of books on the topic in the 1980s. In Europe, similar researches appeared in the 1990s. The first approaches intended to introduce new (direct, separated, stand-alone) school subjects and courses for the development of thinking skills, but they were not successful. The later works, which integrated the developing exercises with the teaching materials of the already existing school subjects (embedding, infusion, enrichment) and used the training for supporting the deeper understanding and better mastery of the subject matter knowledge itself as well, were more successful and resulted in measurable improvements. Despite the encouraging results a number of practical difficulties limited the spread of the suggested methods, including the lack of proper tests for monitoring students' intellectual development and time-consuming preparation of the modified teaching materials.</p> <p>Over the past four decades the intentions of transforming school education, making it more successful in improving students' general cognitive abilities have remained more or less unchanged; there are unvaried objectives behind the varying terminology (from general thinking skills, through key competencies to 21st century skills). On the other hand, a more elaborated knowledge-base (e.g. the theoretical consequences of the cognitive revolution and the empirical results of the large-scale international assessments, such as PISA) is available to re-conceptualize the old aims and the new technologies offer sophisticated low-cost tools for delivering training materials. The regular online assessments provide data for outlining the individual developmental trajectories of the students, the rich set of background data helps to identify the specific learning difficulties and these cognitive and contextual variables help to devise customizable personalized training. The training materials (playful exercises) can be delivered any time according to the students' needs and can be flexibly integrated into classroom activities, after-school programs or utilized even from home. However, realizing the potentials created by the already existing online assessment system (eDia) and the available technologies require further developmental work.</p>	

## Session B

<p><i>Gyöngyvér Molnár</i></p>	<p>The Power of Assessment: Efficient Testing for Personalized Learning</p>
<p>We need to re-consider the role of exams and the role of tests, in particular high stakes summative assessments that measure students' performance just in one instance and therefore in limited scope. Formative assessments carried out at school level are still norm-referenced, in most cases paper-based and focus on factual knowledge. They do not result in an effective education in the 21st century. This paper discusses how Information and Communication Technologies can support the transition from computer-based testing to efficient testing for personalized learning. It highlights the results of several large-scale empirical studies conducted among students, teachers and schools about the technological (infrastructure, students' ICT skills) and societal (students' and teachers' acceptance) conditions regarding introducing computer-based assessment in high stakes and low stakes testing in Hungary and gives a broad overview about the recent developments in this field. An online diagnostic assessment platform and system; consisting of 10.000 tasks in the field of mathematics, reading and science, called eDia; is being developed by the Center for Research on Learning and Instruction at the University of Szeged, which aims at developing the efficiency, quality and effectiveness of the Hungarian education system. In the last year more than 100.000 tests have been filled in within the system in the three main literacy domains.</p>	
<p><i>Samuel Greiff, Sascha Wüstenberg, Benő Csapó, Andreas Demetriou, Jarkko Hautamäki, Arthur C. Graesser &amp; Romain Martin</i></p>	<p>Educating Students Towards Becoming Proficient Problem Solvers: Initiating an Important Discussion Around 21<sup>st</sup> Century Education</p>
<p>This presentation is aimed at discussing the relevance of domain-general problem solving skills for a comprehensive approach to contemporary education. We argue that education in the 21st century needs to be comprehensive in the sense that it should equip students with domain-general problem solving skills in addition to domain-specific expertise. We argue that in their current state contemporary educational systems fall short of addressing the societal and individual needs to teach and to foster domain-general problem solving and that education needs to be extended. To this end, researchers face three main challenges: (1) to increase the relevant stakeholders' awareness of the existence and the importance of domain-general problem solving skills, (2) to optimize the ways in which such skills can be assessed, and (3) to explore ways to foster students in developing and maintaining these skills. This presentation reflects the contents of a paper with the same title that was published by the authors in Educational Research Review.</p>	

## Progress in the Partner Institutions

## Session C

<i>Sascha Wüstenberg, Katinka Hardt &amp; Samuel Greiff</i>	The Merits of Log File Analyses- Exemplary Findings Based on the PISA 2012 Problem Solving Data
<p>Since the advent of computers to psychological assessment, the potential of process characteristics capturing participants' test-taking behaviour stored in computer-generated log files has been praised. Analyzing log files enables researchers to not only measure the outcome of the solution, but also to investigate the preceding steps that result in the outcome. In this presentation, we shortly summarize log file analyses using own studies that investigated the relation between participants' strategic behavior and performance in Complex Problem Solving (CPS) tasks. Further, we present results based on analyses of the PISA 2012 problem solving log file data (N=16,000) of one problem solving task (i.e., Climate Control; <a href="http://cbasq.acer.edu.au">http://cbasq.acer.edu.au</a>). More specifically, we analyzed whether (1) the vary-one-thing-at-a-time (VOTAT) strategy, an optimal strategy when exploring Climate Control, is related to performance in the task and to overall performance in PISA 2012 problem solving, both on the individual and on the country level. Further, we (2) analyzed students' behavioural patterns on a fine-grained level and identified different groups of students with qualitatively different levels of strategic mastery. Results revealed that (1) students who applied the VOTAT strategy showed higher overall success in the Climate Control task and also had a higher level of overall problem solving proficiency. Further, (2) we could separate students with varying levels of strategic non-mastery, ranging from no systematic application of strategic behaviour to actually applying VOTAT but without translating the generated information into a workable solution of the problem. Implications for assessment and teaching are discussed on the backdrop of these results.</p>	
<i>Katarina Krkovic, Maida Mustafic &amp; Samuel Greiff</i>	The Role of Visual-Spatial Thinking in Complex Problem Solving Simulations
<p>Computer-based complex problem solving assessment simulations commonly rely on visual-spatial representations of problems (cf. Bühner, 2008). Thereby, during the interaction between problem-solver and problem space these visual representations may dynamically change. For instance, in some simulations, sliding a controller causes changes in a graph, and in another simulation, pressing a specific button can cause the change in how a specific problem object looks like (for instance, in the "Aquarium task" in the MicroFIN simulation). Very often such simulations require the representation of acquired knowledge in a graphical way (e.g., in MicroDYN, drawing arrows between input and output variables). For this reason it may be hypothesized that the visual-spatial thinking is a predictor of the performance in complex problem solving simulations beyond fluid intelligence. In order to investigate this research question we administered two complex problem solving simulations – MicroDYN and MicroFIN together with fluid reasoning assessment and visual-spatial thinking test to N&gt;500 seventh-graders in Germany. In</p>	

order to test for unique contribution of visual-spatial thinking to the explanation of variance in complex problem solving performance, we conducted a stepwise latent regression analysis. Furthermore we investigated if this contribution is equal for both simulations – MicroDYN and MicroFIN. The first results confirm the hypothesis and indicate an important role of visual-spatial thinking for the success in both complex problem solving simulations. The results of this study suggest that the design of a simulation can require specific competencies for the performance success, and emphasize the importance of investigating other simulation characteristics, for instance the use of numerical or verbal content and skills required to master such simulations.

<i>André Kretzschmar, Neubert, Sascha Wüstenberg &amp; Samuel Greiff</i>	Complex Problem Solving Within the Nomological Network of Intelligence
--	--

Complex Problem Solving (CPS) is a promising 21st transversal skill, empirically connected to several outcomes and recently included in educational large-scale assessments such as PISA. New assessment instruments of CPS, for example MicroDYN and MicroFIN, show convincing psychometric properties and evidence of convergent, discriminant, and criterion-related validity. However, most of previous studies only used single operationalizations of CPS and related cognitive constructs to examine the relation between them. Therefore, we investigated the relation between CPS and intelligence with broad operationalizations of both constructs on the basis of N=227 students. Results demonstrated that a substantial amount of CPS variance was not covered by several facets of intelligence (i.e. reasoning, mental speed, memory, creativity, general knowledge). Further analysis with regard to the relations to external criteria (i.e. school grades) led to differentiate findings. Implications for a potential CPS construct within the nomological network of intelligence as well as for future studies are discussed.

<i>Julia Rudolph, Christoph Niepel &amp; Samuel Greiff</i>	How Motivational and Cognitive Correlates and Their Interaction Influence Complex Problem Solving Processes
--	---

The present study contributes to a detailed understanding of Complex Problem Solving (CPS) processes by investigating how motivational and cognitive correlates are linked to CPS. CPS is the skill to handle dynamic and complex challenges by actively interacting with the problem (e.g. using an unknown computer program). To be able to understand how CPS can be fostered best, potential predictors and influences on CPS, like Need for Cognition (NFC), need to be understood. NFC is the intrinsic motivation to think and engage in effortful cognitive tasks and is known to be related to cognitive skills such as reasoning and CPS. However, as reasoning is a substantial facet of CPS, too, it is unclear if the correlation of CPS and NFC is merely an artefact based on their common link to reasoning. Furthermore, no empirical evidence shows how NFC actually taps CPS. One possible explanation is that NFC influences behavioral aspects of CPS, like time on task, resulting in a partial mediation of the link between NFC and CPS by time on task. In particular, it is unknown whether the extent of the link between NFC and CPS is depending on reasoning ability. That is, whether NFC has a larger link to CPS in students with greater cognitive ability. Such an interaction of NFC and reasoning with regard to their link to CPS (and time on task in CPS) would support theoretical considerations that NFC would help to activate existing cognitive resources, but only to the extent that they

are developed. Acknowledging those research gaps, the present study will theoretically and empirically answer the questions: (1) Are NFC and CPS linked beyond their common correlate reasoning? (2) Is the link between NFC and CPS mediated by time on task in CPS? (3) Is a potential effect of NFC on CPS (and CPS time on task) interacting with reasoning ability?

## Session D

<i>Christoph Niepel &amp; Samuel Greiff</i>	Complex Problem Solving, Reasoning, Working Style, and Extraversion: Examining the Link Between Extraversion and Cognitive Ability
<p>We explored the behavioral mechanisms underlying the link between extraversion and cognitive ability. We focused on reasoning and complex problem solving as cognitive abilities, and response latencies and action orientation during test taking as behavioral mechanisms indicating working style. Results obtained in a sample of 326 German adults generally indicated that specific working styles such as slower response latencies and higher action orientation mediated the link between extraversion and cognitive ability. We discuss the results in light of Eysenck's cortical arousal theory and suggest directions for future research.</p>	
<i>Maida Mustafic &amp; Samuel Greiff</i>	Exploiting the potential of complex problem solving in personnel selection: Overcoming weaknesses and maintaining strengths of current assessment approaches
<p>When it comes to education and educational policy, potential gender differences and the processes determining these differences are of crucial importance (e.g., OECD, 2009). Whereas differences between males and females in domain-specific abilities (e.g., reading, mathematical performance) are already well established, findings on gender differences in domain-general abilities are scarce. The current study therefore aimed to investigate whether males or females perform better in complex problem solving (CPS). Additionally, self-concept measures (problem solving, mathematical, verbal) were hypothesized to be associated with gender and with CPS, indicating a possible impact of self-concept in gender differences. Structural equation models (accounting for the hierarchical data structure) revealed on a sample of 555 German secondary school students (<math>M = 13.57</math> years, <math>SD = 0.66</math>) small effects of gender on knowledge acquisition and application in that girls performed (significantly) better than boys in knowledge acquisition and boys performed (marginally significantly) better than females in knowledge application, while controlling for intelligence. Additionally, the problem-solving self-concept was significantly associated with knowledge acquisition and application. This pattern of results remained the same after controlling for mathematical and verbal self-concept.</p>	
<i>István Thékes</i>	Results of a Complex English as a Foreign Language Vocabulary Test Among Hungarian 6th Graders
<p>The purpose of my study was to explore the functioning of a newly developed integrated vocabulary test that will be used online in the future. Assessment of vocabulary can be done either context-dependently or context-independently (Read, 2000). A few researchers (Nation, 1990; Henriksen, 1999; Qian, 2002) did a modelling of word knowledge. A major agreement among them is evident. That is, there is breadth and</p>	



depth of word knowledge and there is productive and receptive word knowledge. Vocabulary tests generally measure one of these aspects. There have been attempts to create integrated vocabulary tests. Laufer et al. (2004) developed CATSS, a computer adaptive test that involves passive recognition and active recall of words. Ishii and Schmitt (2009) developed such diagnostic measurement tools that can be implemented in the assessment of both breadth and depth of vocabulary.

An obvious tendency can be noticed in research in terms of foreign language vocabulary. Researchers tend to make clear distinction between breadth and depth of vocabulary. Meara (2009) interprets vocabulary breadth as the number of words learners know. Depth, on the other hand, means how well learners know these words. In this study we examine the breadth of vocabulary of 6th grade students. As far as types of foreign language vocabulary are concerned, the most applicable theoretical modality in my assessment is that of active recall, active recognition, passive recall and passive recognition (Laufer et al., 2004). Four important assertions must be underlined before tapping into the construct of word knowledge: 1) Students know only a limited number of words (Laufer, 2000); 2) students have limited knowledge of secondary meaning senses (Schmitt, 1998); 3) students have limited awareness of the different derivative forms of a word (e.g., silly, silliness) (Schmitt and Zimmerman, 2002); 4) students use L1 translations when understanding the meaning of L2 words (Jiang, 2004). Using visual stimulus is another important trait of my research. It is not so common that foreign language vocabulary tests use pictures as part of their data gathering instrument. The Peabody Picture Vocabulary Test (Dunn and Dunn, 1997) is a type of such instrument. Tonzar et al. (2009) found that picture-based methods of eliciting vocabulary is more efficient than the word-based method and learning lexis aided by visual stimulus is more motivating as well. Thus, the application of pictures in the tests has firm grounds. In my study 103 Hungarian 6th graders participated who had been studying English as a foreign language for two years. The test was administered in November 2013. On the basis of the BNC corpus and Nikolov's (2011) recommendations items from the first 3000 most frequent were selected. Upon developing the integrated vocabulary test, all of the major tests were considered with special regard to the VKS (Paribakht and Wesche, 1999) and the VLT (Nation, 1990). The test consisted of 7 elements with the following tasks: 1) audio-recorded words had to be matched with pictures, 2) audio-recorded words had to be matched with the definitions, 3) words had to be selected out of 6 that matched 3 pictures, 4) written words had to be matched with pictures, 5) written words had to be matched with definitions, 6) the meanings of meals had to be written by the learners next pictures, 7) on a 4-level scale learner had to prove if they knew the word or could even write sentence with it. In all tasks learners could reach maximum 9 points. I wanted to find the answer to the following research questions: How does each item function?; Which task proves to be the easiest?; How do the different test elements correlate? The reliability of the test proved to be satisfying ( $\alpha=.82$ ). In my presentation I go into details on the functioning of specific items and the item analysis. The first task was the easiest based on the mean of the results ( $m=86\%$ ) and as it was expected, the second one was the most difficult ( $m=38\%$ ) Strong and significant correlation was found between task 1 and task 4 ( $r=.91$ ) and task 1 and task 3 ( $r=.88$ ). Looking at the results with a focus on word class, it can also be concluded that the identification of nouns was the most simple for the students and nouns at A1 level were identified by even the lowest scoring

students. Classroom implications, limitations and future research are discussed. No integrative English vocabulary test had been developed for young learners yet, however, this new instrument will make it possible to map the word knowledge of students diagnostically.

## Session E

<i>Mari-Pauliina Vainikainen</i>	Educational Equity and Large-Scale Assessments
<p>One of the most important goals of large-scale educational assessments is to get information about educational equity and equality. Equity refers to access to services and resources regardless of the background of the student, including support structures for compensating the initial differences and disadvantages. Equality can more directly be measured looking at the learning outcomes of subpopulations, e.g. girls and boys or students from different socioeconomic background. School segregation is often seen as a mechanism producing inequity and inequality, and in this sense Finland has – at least until recently – been different from many other countries with its strong public basic education system and small differences between schools. However, even in Finland there are structures within schools that seem to produce systematic differences over time, and in many of the biggest cities also school-level segregation is increasing. Therefore, several large-scale multidisciplinary assessment studies have been launched to understand the mechanisms of differentiation and to have a critical look at educational equity in Finland of 2015. In this presentation I give an overview of the ongoing studies and their recent results, and present future visions of the course of the studies with a special emphasis on the functioning of the support structures for compensating the disadvantages. Possibilities for collaboration with other participants of the workshop are discussed.</p>	
<i>Risto Hotulainen, Mari-Pauliina Vainikainen, Ninja Hienonen &amp; Jarkko Hautamäki</i>	Log-Data and Person-Centered Approach: added value or just nice to know
<p>In earlier studies we have seen that students Time-on-task predicts performance in computer-based-assessment in Finnish Learning-to-learn tasks covering wide range of cognitive abilities (Kupiainen et al. 2014). However, we anticipate that there might be sub-groups having differences in their abilities and interests, and accordingly they might invest their CBA Time-on-task differently. Consequently, we expected to reveal several Time-on-task profiles evidencing either more general approach towards CBA tasks or alternatively more ability and interest based grouping. To study existence of sub-groups we used Latent profile analysis (LPA). LPA analysis is a probabilistic or model-based variant of traditional cluster analysis, and it is often used to identify the smallest number of latent classes (groups) that describe the associations among observed continuous variables. Participants (N = 3970) of the study were 9th graders who participated to national students' learning to learn assessment in 2012. Our analysis confirmed existence of five Time-on task groups. Each group differed statistically significantly from each other groups in time use in all used measures (n = 8) without a few exceptions. Two groups (longest time use and shortest time use) provided evidence for general approach towards CBA tasks and two other groups provided evidence for ability and interest based grouping. One group fell in</p>	

<p>between of these two groups. Discussion is needed to have arguments both for and against of LPA's use for studying CBA log data.</p>	
<p><i>Carolin Hahnel, Frank Goldhammer, Ulf Kröhne &amp; Johannes Naumann</i></p>	<p>Insights into digital text processing: On the dependency of the updating effect on reading purpose, hypertext structure and readers' behavior</p>
<p>Amongst other requirements, reading non-linearly structured digital text on the Web demands for the continuously monitoring and updating of one's mental text representation across several web pages. The ability to monitor and integrate new relevant information in working memory is described by updating – an executive working memory function. In this study, we investigated the updating effect in digital reading in dependence of the reading purpose, the hypertext structure and readers' actual behavior in a hypertext. We hypothesized that, after accounting for print reading skills, the digital reading association with updating (1) becomes stronger in tasks requiring the extraction and combination of multiple information, (2) increases when more information is available in a hypertext, and (3) depends on the extent students actually invest in processing task-relevant information. To test these hypotheses, score data in digital reading, updating, and print reading was collected from a subsample of 293 PISA students. The four PISA reading aspects of items (access and retrieve, integrate and interpret, reflect and evaluate, complex) were used as an indicator of the reading purpose; the number of web pages served as characteristic of the hypertext structure. Reading behavior was extracted from log data and operationalized as counts of students' visits on pages with relevant information and time they spend on these pages. Using generalized linear mixed models (GLMM), first results confirmed an association between digital reading and updating besides print reading, and showed further differential effects for reading aspects focusing on integration of and inferring on information compared to evaluating information according to one's own knowledge and experiences. However, no effect variation was found across the number of hypertext pages. Further analyses concerning the interaction between updating and students' reading behavior will be presented and implications on students cognitive processing of hypertexts will be discussed.</p>	
<p><i>Ingo Barkow</i></p>	<p>First Results From the Survey on Metadata Management in the Educational Sciences</p>
<p>The long-time preservation of items used in computer-based assessment is difficult as the metadata standards do not seem to be sufficient for storing complex content. Though metadata standards like DDI Lifecycle 3.2 and QTI 2.1 have improvements in regards to response domains and workflow currently only items which are close to questionnaires (meaning the use of simple stimuli) can be represented. To identify the white spaces in the different metadata standard a survey on metadata management in the educational sciences has been performed by the author as part of doctoral thesis in November and December 2014. This talk will present first results from this survey which might have an impact on the further development of existing metadata standards like DDI Lifecycle towards the domain of the educational sciences or give best practices on how to use sets of existing standards as an interim solution.</p>	

<i>Ulf Kröhne, Heiko Rölke, Michel Dorochevsky &amp; Frank Goldhammer</i>	Technology-Based Testing in the (German) National Educational Panel Study: Using CBA ItemBuilder for Innovative Items and Adaptive Assessments
<p>The National Educational Panel Study (NEPS) is an ongoing longitudinal study combining context assessments, questionnaires and competence tests in six starting cohorts, ranging from early childhood to adulthood. Assessments are conducted in various settings following a design that allows to describe development over time and comparisons between cohorts. The implementation of technology-based testing (TBT) has been started in NEPS for test-settings, in which computer-based test delivery is possible without affecting panel stability and participation rates of individuals or institutions. Psychometric preconditions for introducing TBT, investigated in a series of mode-effect studies are, amongst others, construct equivalence to allow comparability of test scores between cohorts and derivation of change scores. In addition, operational preconditions include an item authoring tool that supports a multi-site, distributed item development process within the NEPS consortium. Important purposes for TBT are not only to reduce interviewer effects and to use innovative response formats such as highlighting or drag-and-drop, but also to increase measurement efficiency by means of adaptive test assembly. Due to the panel design, the multiple dimensions and the unit structure, specific developments are required to fit the needs of adaptive test assembly for competence assessment in NEPS. Hence, instead of restricting the delivery system to a particular set of adaptive algorithms, we designed an interface between the delivery system and the psychometric algorithms, accompanied by a collection of meta-data that can be used to describe item pools and test-takers with selected background information such as gender or previous assessment results. The technological platform as well as an exemplifying application of adaptive testing with the CBA ItemBuilder using a specific implementation of this interface will be presented. As an outlook we will discuss different options for further implementations of such an interface, for instance, by integrating algorithms implemented in specific statistical software environments.</p>	

## Session F

<i>Attila Pásztor</i>	Computer-Based Assessment and Development of Inductive Reasoning Strategies
<p>Technology offers new opportunities to assess and develop thinking skills with the possibility of providing instant feedback, personalized instructional support and motivating learning environment (Csapó, Lőrincz &amp; Molnár, 2012; Wouters &amp; Oostendorp, 2013). The aim of this study is to explore these possibilities through the analyses of the effectiveness of an online training program which develops inductive reasoning strategies.</p> <p>The sample was taken from third and fourth grade students (N=314, age M=9.73, SD=0.67). The experimental group (N=88) received a five-week-long computer-based training (20- to 40-min. sessions) in their school's ICT room. The control group was matched based on pre-test scores, year and gender.</p> <p>The structure of the program was based on Klauer's paper-based inductive reasoning training program (Klauer, 1989). It consisted of 120 instructional tasks for developing six inductive reasoning strategies such generalization, discrimination, recognition of relations,</p>	

differentiating relationships, cross-classification and system construction. In contrast with Klauer's content general training program we embedded the learning tasks in mathematical content. Immediate feedback was given after every task and instructional support was provided in case of failure to guide the learning process. In order to test the effectiveness of the training program we developed an inductive reasoning test which consisted of 43 non-verbal items and assessed the same thinking processes as described above. Instant feedback was provided after test completion. We also measured children's attitudes towards the program with five-point-scale questionnaire items. Data collection and the online program were delivered via the eDia platform (Electronic Diagnostic Assessment System).

The reliability coefficient of the test was Cronbach's  $\alpha=.86$ . The experimental group significantly outperformed the control group on the post-test ( $t(174)=-2.288, p=.02$ ). The effect size of the training program was  $d=.33$ . However, not every child improved during the training and two of the inductive strategies did not develop significantly on the group level, children reported positive attitudes towards the program.

Our study demonstrates the advantages of technology in both assessment and development of reasoning skills in a computer-based environment. Further studies should investigate the transfer effects of the training and the influence of game elements (e.g., game story) on learning achievement and motivation.

<i>Kata Asztalos</i>	Measurement of Musical Perception with an Online Diagnostic Tool
----------------------	--

Rapidly developing technology-based assessment techniques are providing new opportunities for music education too. Diagnostic testing of musical abilities is necessary to consciously compensate deficits, establish developmental processes and to find talented children. High class sizes and short time usually limit the possibility of regular evaluation in music classrooms, thus innovative tools can substantially enhance the effectiveness of pedagogical work.

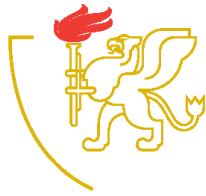
In the present study, an online diagnostic measurement tool was developed for examining basic musical abilities among 7-17 year old students. Participants were primary (N=1133) and secondary school students (N=658). The test was administered through the eDia online assessment platform. Students solved the tasks in classrooms equipped with computers in their own school using the operation system and browser installed originally on the computers. We collected information about the following musical hearing abilities: melody, rhythm, tempo, dynamic, harmony, pitch and timbre discrimination and visual connection. The test was followed by a background questionnaire asking students about the testing itself, their music related habits, school achievement, socio-cultural background variables and free time activities.

Distribution of the results shows that developmental tendencies of musical abilities are more dynamic during the first part of primary school, after the fifth grade focused musical training is required to reach better performance. For analyzing the structure of the examined musical abilities, we applied exploratory factor analysis with MPlus. Results are in accordance with the cognitive neuropsychological models of musical perception.

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.

<i>Renáta Kiss &amp; Ágnes Hódi</i>	Online Assessment of Kindergarten Aged Children's Speech Sound Perception and Production
<p>Several national studies examined children's metalinguistic awareness and its components in terms of the order of emergence in language, reading and writing acquisition. The spontaneous development of metalinguistic awareness in early childhood provides us with an opportunity to assess children in this domain from the age of 4. Phonemic awareness as a form of metalinguistic awareness is one of the key predictors of reading. Thus, its early development may contribute to an improved performance in reading (Adams, 1990; Castles &amp; Coltheart, 2004; Ziegler &amp; Goswami, 2005, Wagner et al., 1997; de Jong &amp; Van der Leijon, 1999; Hulme et al., 2002; Muter et al., 2004).</p> <p>Early diagnosis and intervention is substantial, as recognition of sounds at the beginning and at the end of the words emerges well before school start (Ziegler &amp; Goswami, 2005). In children with a typical developmental curve some components of phonemic awareness emerges spontaneously (Hulme et al., 2005; Castles et al., 2009), although other aspects develop as a results of learning to read.</p> <p>The fact that speech-sound discrimination forms a core part of the DIFER, the Hungarian school-readiness test battery (Nagy, Csapó &amp; Molnár, 2014) also underpins the importance of this domain in developing one's potential. The speech-sound discrimination subtest focuses on speech sound perception. However, another but similarly important component of later reading success is the ability to produce different sounds and words. Therefore, we developed an online assessment instrument for measuring both children's speech sound perception (27 items) and production (100 items). The perception subtest contains phonological awareness tasks (phoneme identification 18 items, syllable identification 9 items) and the production subtest consists of a rapid automatized naming (RAN) test. The tests were administered by means of the eDia online assessment system. The sample consisted of kindergarten aged children (N&gt;393) between the ages of 4-6.</p> <p>The online assessment provided us with ample opportunities and benefits. Nevertheless, several questions arose regarding the implementation of the assessment, children's and kindergarten teacher's attitudes towards testing, their behaviour and the perceived pressure during the actual test-taking process, ICT familiarity, validity and infrastructural issues. The presentation on the study will contain two major components. First, the assessment instruments and preliminary results will be introduced. Second, personal experiences and videos will be used to pave the way for the discussion of the above mentioned issues.</p>	

<i>Andrea Magyar</i>	The Development of a Criterion Oriented Online Measurement System for the Measurement of Word Reading Skills by the eDia Platform
<p>An appropriate reading level is an essential prerequisite for all further learning and it is important to monitor its improvement among young students and indicate who lags behind. József Nagy (2006), recognizing the crucial role of the vocabulary and word-reading skills during reading skills development, created a criterion-oriented paper-based test system which is suitable for the measurement of the knowledge of the most frequent 5000 Hungarian colloquial words.</p> <p>With the spread of computers the conditions for utilizing the opportunities of computer-based tests offered and made it possible to develop an online adaptive version of the previously paper and pencil based test. During the construction of the adaptive test system the main aim was to retain the original structure of the test, so the measurement of the word-reading skill has been established according to four different perspectives (headwords, inflected words, synonyms and word-meaning reading). Among the many possible adaptive structures the four-stage adaptive model has been chosen with five different difficulty levels in each stage.</p> <p>The development of the electronic test versions was conducted by the Online Diagnostic Assessment System (eDia) developed by the Center for Research on Learning and Instruction at the University of Szeged (Molnár &amp; Csapó, 2013). The platform is internet based, the testing process is carried out in the students' own schools. At the end of the test the pupils get immediate feedback about their performances. The test system is suitable for measuring the number of words the student is able to read and also can determine which words are not recognized, even if those words were not included in their test version.</p>	



**UNIVERSITY OF SZEGED**  
*Doctoral School of Education*

The workshop is financed by the TÁMOP 3.1.9-11/1-2012-0001 project entitled *Developing Diagnostic Assessments*.

The project is funded by the European Union and co-financed by the European Social Fund.

**SZÉCHENYI** 2020



HUNGARIAN  
GOVERNMENT

European Union  
European Social  
Fund



INVESTING IN YOUR FUTURE