UNIVERSITY OF SZEGED
CENTER FOR RESEARH ON LEARNING AND INSTRUCTION
www.edu.u-szeged.hu/ok

TÁMOP 3.1.9-08/1-2009-0001
„Developing Diagnostic Assessments” project

3TH SZEGED WORKSHOP ON EDUCATIONAL EVALUATION

ABSTRACTS

27-28 April 2011

Venue:

Szeged Committee of the Hungarian Academy of Sciences

H–6720 Szeged, Somogyi utca 7.

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

SZEGEDI AKADÉMIAI BIZOTTSÁG
SZAB•1961

ABSTRACTS

Opening Session

Benő Csapó	Developing an Online Assessment System: Results of the First Phase and Future Plans

In the framework of a major research and development project at the University of Szeged, an Online Diagnostic Assessment System (ODAS) is being constructed. In the first phase of the project the broader theoretical and technical preparations took place, such as developing frameworks for diagnostic assessment for reading, mathematics and science, exploring the possibilities of diagnostic assessments at other domains and writing test items and establishing item banks. TAO (see: www.tao.lu) was chosen as the platform for the online system. After implementing the system and translating the language modules, several pilot assessments have been carried out to test its capabilities. Moreover, the ICT infrastructure of the partner schools has been mapped out.

In the second phase of the project, to be started in September 2011, ODAS needs to be extended and prepared to assess several hundreds of thousand of students, and to store and maintain several thousands of items. Furthermore, it has to communicate with the teachers of the assessed students and prepare aggregated information for other stakeholders. The challenge of this development is to build a system which is able to serve a much larger population than TAO was originally designed for while preserving its advantages.

This presentation outlines the plan for the next phase of developing ODAS, identifies some problems and bottlenecks of the current system and invites the audience to discuss the alternatives for the extension of the system. As the next phase of the project will also be carried out in an international cooperation the experiences of the partner research groups will be a substantial input in the course of finalizing the project plan.

Session A

Joachim Funke	Interactive Problem solving - From a Static to a Dynamic Perspective

Problem solving competency is a central objective within the educational programs of many countries. The acquisition of increased levels of problem solving competency provides a basis for future learning, for effective participation in society and for conducting personal activities. Students need to be able to apply what they have learned to new situations. Problem solving requires the transformation of a given state to a desired goal state. Traditionally, problem solving was measured via the solution activities in the context of static tasks, i.e. tasks that did not change over time. In the PISA 2012 Framework, a shift from these static tasks to more interactive problems has occurred. Authentic, relatively complex problems will be used within the PISA 2012 problem solving assessment. These tasks require direct interaction by the solver to uncover and discover relevant information. Examples are the problems commonly faced when using unfamiliar everyday devices such as remote controls, personal digital devices (e.g., mobile phones), home appliances, and vending machines. Other examples arise in situations such as physical conditioning, feeding animals, growing plants, and social interactions. Problem solving skills are necessary to achieve more than a basic level of skill in dealing with such situations. Interactive problem solving is a step towards the interplay between a person striving for a goal and an unknown environment that responds dynamically to the activities of the problem solver. At the same time, the focus on the interaction gives a more distinct weight to the process of problem solving. The talk presents the rationale for this paradigm change and explains the theoretical background for the development of interactive problems.

Samuel Greiff	Recent Developments from MicroDYN and MicroFIN

In educational large-scale assessments such as PISA and PIAAC an increasing interest in measuring cross-curricular competencies can be observed only recently. These are now considered valuable aspects of school achievement. Dynamic problem solving (DPS) is an interesting construct for diagnosing domain-general competencies. However, DPS requires to be measured computer-based and only the recent emerge of computers and the shift to computer-based assessments in large-scale contexts within platforms as TAO allows the measurement of DPS to be considered seriously. New developments in MicroFIN and MicroDYN, two innovative approaches to capture DPS, are introduced. While MicroDYN is based on linear structural equation systems, MicroFIN rests upon the formal framework of finite state automata. Within MicroFIN, a broad range of different automata frequently used in daily life can be designed (e.g. ticket vending machines, washing machines, etc.). Newly developed MicroFIN items as well as advances in DPS scoring are presented and first empirical results including findings on the relationship of both approaches are described. Finally, milestones to be taken when applying dynamic systems in a large-scale context to 15-year old students are discussed. Necessary theoretical and software developments to measure DPS as individual competence in the 21st century are specified.

Session B

Sascha Wüstenberg	Dynamic Problem Solving and Intelligence

Dynamic problem solving is commonly seen as a key qualification for success in life and therefore receives interest from international large-scale assessments like PISA. This growing interest increases the need for efficient assessment procedures. As a possible measurement device, I present MicroDYN, a new approach bringing together a formalized item pool and Dörner's Theory of Operational Intelligence (1986). Three facets to measure Dynamic Problem Solving competency are theoretically derived and empirically evaluated: (1) information retrieval, i.e., use of appropriate strategies to identify connections between variables, (2) model building, i.e., the correctness of acquired knowledge, and (3) forecasting, i.e., the ability to control a given system. First results indicate that a 3-dimensional model fits the data well. Regarding construct validity, model building predicts final school grade as good as classical IQ (measured by APM) and explains in a multiple regression variance beyond it. Additionally, the relationship between DPS and other potentially influential constructs like computer anxiety or attitudes towards computers as working tool are presented. The results emphasize the importance of the construct in the assessment and prediction of cognitive performance from a theoretical and empirical point of view. By this, MicroDYN is the first psychometrically sound measurement device in the history of problem solving research.

Sonnleitner, P., Brunner, M., Keller, U., Greiff, S., Funke, J., Martin, R., Latour, T. Hazotte, C., Mayer, H.	The Genetics Lab – On the Usability and Psychometric Properties of a newly developed Microworld to assess Complex Problem Solving in Student Samples

At the beginning of the 21st century new educational goals have been defined that emphasize higher order thinking skills, cross curricular competencies, or the (domain-general) ability to solve complex problems (e.g. Kuhn, 2009; Ridgway & McCusker, 2003; OECD, 2003). Thus, the psychometric sound assessment of these competencies is of vital importance to enable evaluations of related educational interventions. Although microworlds - developed in the field of Complex Problem Solving research - are probably the most fruitful approach to satisfy this need, the vast majority of studies investigating their psychometric properties only included university students.

To fill this gap, we developed the Genetics Lab (GL), consequently focusing on a younger target population, starting with class level 9. We included standardized, automated instructions and game-like characteristics such as immediate feedback to increase acceptance and make the handling of the GL more comprehensible. To fulfill the high expectancies of today's students, also described as "digital natives", we laid special emphasis on the design and usability of the assessment.

The paper discusses two studies, both evaluating the usability and psychometric properties of the GL. First results indicate that the GL shows high usability and provides psychometric sound performance indicators for our target population. In addition, we demonstrate the benefits of capturing students' problem solving strategies by means of behavioral data which provides key information for educational interventions.

Andreas Fischer	Facets of Problem Solving: Ideas for Multidimensional Assessment

Complex or Dynamic Problem Solving (DPS) has a tradition of experimental research lasting almost three decades. Based on his extensive works on complex scenarios and problem solving Dörner (1986) developed a theory of operational intelligence, claiming that complex problems feature a set of charactersitic demands: (a) the polytely of the task at hand, (b) the complexity of its structure, (c) the interconnectivity of the involved variables, (d) the intransparency of these connections, and (e) the dynamics of the system. Each of these features aims at a corresponding facet of an individuals DPS competency: (a') Evaluation processes and priority setting in case of conflicting goals (= evaluation). (b') sparing of irrelevant information (= reduction of information). (c') successive generation of an adequate internal representation of the system structure (= model building). (d') exploration of important information (= information retrieval). (e') prognosis of future developments in light of the actual situation and past decisions (= forecasting). Greiff and Funke (2010) proposed a computer-based testing environment (MicroDYN) for assessing three of the five facets of operative intelligence (information retrieval, model buidling and forecasting) using minimal complex systems as items. First results support the three-dimensional internal structure. Additionally, construct validity regarding school grades, intelligence and occupational problem solving showed promising results (DPS proved to be an important predictor for school grades with incremental validity to intelligence). Currently, the two missing facets (reduction of information and evaluation) are to be integrated in MicroDYN, and the internal structure of the construct has to be further explored, to develop a sufficiently validated five-dimensional facet diagnosticum for assessing DPS.

Session C

David Tobinski & Prof. Dr. Annemarie Fritz	Complex Problem Solving Within Systems with Eigendynamics in the Adolescence

Complex problem solving within systems with eigendynamics (CPSSED) is a challenge for cognitive psychology. Previous research focused on adulthood so there is a gap in measuring competencies in CPSSED in childhood and adolescence. Therefore the paradigm EcoSphere (ESP) has been developed. ESP offers two designs to be explored time-discrete in steps or time-continuous. The simulation uses a system of differential equations. The actual semantic embedment describes an ecosystem with an animal- and a plant-population. The system can start in stable or instable states, so the participants only would have to observe the system or explore it with a following goal-directed intervention. Current analysis inquires the control performance and the correctness of mental models. In ESP mental models are captured in a standardized and digital profile to guarantee highest objectivity.

Gyöngyvér Molnár	Assessing Developmental Aspects of Dynamic Problem Solving

According to the recent tendencies in educational assessment problem solving belongs to the key competencies in the 21st century; it has already been assessed in the framework of large-scale national and international projects as well. However, the previous measurements focused on one measurement point, one type, and one dimension of problem solving. However, knowledge and competencies can also be studied from a developmental perspective (*Molnár* and *Csapó*, 2007) and from different dimensions (*Csaó*, 2004). The paper discusses and compares three different types, three different dimensions of problem solving: discipline, literacy and reasoning; and presents the research design of an evaluation study focusing of different dimensions of problem solving in a developmental aspect. In our three-dimensional system the discipline dimension views problem solving as a means of applying knowledge in a new context (*Molnár*, 2003); the literacy dimension of problem solving is measured by dynamic problem solving (*Greiff* and *Funke*, 2010); and the reasoning dimension of problem solving is measured by research results concerning learning and development of cognitive abilities (inductive reasoning and intelligence). Participants' age ranged from 9 (grade 3) to 17 (grade 11) years. The different tests in the same dimension included a large number of common anchor items that allows the transfer of all results to the same scales. Static paper-and-pencil tests (real-life complex problem solving and intelligence), static (inductive reasoning) and dynamic computer-based (dynamic problem solving) tests were used in the study. In addition to the test, questionnaires regarding socio-economic background, ICT familiarity and attitudes were administered to students in computer-based mode as well.

Session D

Ingo Barkow	Metadata standards in computer-based surveys and cognitive assessments - challenges and opportunities

More and more surveys and assessments drive towards computer-based platforms using design tools and frameworks the next step of development aims towards interoperability. As studies like PISA, NEPS or PIAAC have a design of repeatable waves without interoperability relying on a propriety platform leaves the organizer of a study sometimes stuck with a vendor as the transformation costs to re-design or re-implement those items would be too high for a potential bidder. Furthermore the use of a metadata standard provides higher re-usability and analysis possibilities in research data centers as results from items can be harmonized to be exported into file standards of the user's choice through ETL processes or to be used in a reporting engine. To accommodate those issues metadata standards in surveys (e.g. DDI3, SDMX, Dublin Core) or cognitive assessments (e.g. QTI or SCORM) are formed. This presentation will introduce some of those standards and show their advantages and limitations. Furthermore it tries to give an outlook to future directions of metadata standardization.

Hervé Cholez	Security in CBA: Where are we?

The use of Computer-Based Assessment (CBA) is increasing. Several challenges must still be considered to improve the quality and interests of CBA. One of these challenges is security. Moreover, to establish CBA systems as a trusted supporting medium an essential step is to improve the security of them. CBA security is particularly challenging due to the wide range of attacks these systems have to handle. In addition, the large diversity of contexts, stakes, scales, and processes makes it difficult to depict the complete situation. As for any IT system, “classical” security attacks can occur within the Information System (IS). However, particular forms of attacks pertaining to the specific processes and stakes of CBA can also be identified as brain dump, cheating, test-takers authentication, etc. A state of the art on the security in computer-based assessment (CBA) has been conducted and will be exposed in this presentation. Several risks and countermeasures frequently studied in papers will be presented. Finally, a statistical analysis of the literature will be exposed. This analysis will show the main concerns of the literature and the security topics which are still challenging.

Session E

Michel Dorochevsky	Advances in the CBA Item Builder Platform - Complex Problem Solving Assessment

The CBA Item Builder platform developed for the DIPF has recently been used in large and medium scale assessments at the University of Heidelberg/Germany and Szeged/Hungaria as well in PIAAC 2011. The types of items have focused on assessing complex problem solving incorporating dynamic systems (MicroDYN) and finite state machines (MicroFIN). The key element of the CBA Item Builder Platform is a graphical authoring tool which can be handled by test developers without any programming knowledge. We report about our experiences and latest developments in particular for automatic scoring of complex problem solving items.

Frank Goldhammer, Yvonne Pfaff, Heiko Rölke	Measuring Individual Differences in ICT Literacy: A CBA-Item Builder Builder Use Case

The computer and the Internet have become essential parts of our daily life. Thus, computer-related skills are considered as an important condition for successful problem-solving in educational and workplace settings. This study focuses on two major facets of ICT literacy, namely accessing information and evaluating information. The ICT literacy facet ‘access’ is defined as the skill and speed to use basic interaction functions of graphical user interfaces of computers to access, collect, and provide information. ‘Evaluate’ refers to the cognitive skill to efficiently make judgments about the quality of online-information using structural and message features of a website.

To assess the two ICT literacy facets in a performance-oriented way, tasks were developed using the simulation capabilities of the CBA-Item Builder. Tasks include simulated computer environments, e.g., web browser, text editor, the test taker is supposed to interact with. Based on item results as provided by the CBA- Item Builder, i.e., item score, timing information, and logged test taker events, we look at new approaches to measure ICT skills taking into account not only accuracy, but e.g., also efficiency of task engagement. First results on the psychometric properties of the tasks are presented.

Session F

Younes Djaghloul	When the Scoring meets the Complex Event Processing: The TESLA component of TAO platform

The scoring is one of the most important tasks in an e-assessment process. The output of the scoring is a set of values of variables that give an indication on the tested competencies. The common characteristics of scoring are: the definition of the scoring rules is done during the item building phase; a direct comparison between the given answer and the expected one; the value of a score variable is set just after passing the test. However, the current needs in the scoring phase are more complex than this simplistic method. Actually, we need to have: (a) the possibility to define scoring variables after the building item phase, (b) the possibility to calculate the score values in several levels of the process, (c) the possibility to manage the Problem Solving Assessment, in which the strategy is also important as the final answer. This last need constitute one of the most important challenges in e-assessment solution. Our work tends to address these needs by improving the existing solutions on three aspects: a flexible definition of score variables in pre and post test; an efficient mechanism to calculate score values and data captured based on collected events and the variables' definition; the support of scoring based on the analysis and processing of complex patterns met in the sequences of collected events. The main point of our approach is to use some ideas from the well known domain of CEP (Complex Event Processing) in order to establish a scoring solution. Since CEP is largely used in a lot of critical domains (as the security, attack detection, business profiling…) we claim that the existing works in CEP can be an excellent source of inspiration that will help us in our approach. Thus, our contribution consists of extending the Result Module of TAO platform by a specialized component named TESLA (TAO Event based Scoring using Log Analysis techniques). The TESLA component is based on four elements: 1) an efficient Event Log Model that allows tracing actions done by the test taker, 2) a variable description language based on XML schema to define the score variable, 3) the TAO Event Processing Engine provides a set of services to analyze the logs, the main one is the Event Complex Pattern Matching Service that will be used in the problem solving scoring, 4) a visual interface that simplifies the definition of these variables. Since the TESLA component is a part of TAO, the results can be stored in the Result Ontology to benefit of all the existing services of TAO platform.

Zsolt Lavicza	Engaging the GeoGebra Community in Teacher Training, Development, and Research

GeoGebra is rapidly gaining popularity in the teaching and learning of mathematics around the world. Currently, GeoGebra is translated to 56 languages, used in 190 countries, and downloaded by approximately 300,000 users in each month. This increased use compelled the establishment of the International GeoGebra Insitute (IGI) that serves as a virtual organization to support local GeoGebra initiatives and institutes. There are already 55 established institutes on every continent, which pursue training and support of teachers, develop teaching materials, and carry out research projects. In my talk, I will review examples of applications of GeoGebra at different levels of education, introduce various lines of research with GeoGebra, and outline the future of software and community developments.

Session G

Iván Devosa & Gergely Vízvári	Recent Developments of TAO CAPI

TAO CAPI is modified version of TAO. It was designed to work with small groups of users in a time, most of times one user in one time. Earlier - and even now - it was only used for data recording in international assessments like PISA I. The fill-out form is generated form an XML file, which is quite difficult to create: no easily useable GUI editor is attached to TAO CAPI, so our first goal was to create an easy access, user friendly questionnaire generator that can be used by not only programming experts and computer teachers but all other teacher. Our plan is to develop the generator to offer more types of items, and to make it able to reopen and edit the existing questionnaire. In the presentation this on-line generator will be introduced. The second problem has not been solved yet: TAO CAPI cannot handle more questionnaire parallel, and it cannot work if larger group of user try to access it the same time. According to our researches, the database connection could be responsible for the experienced phenomenon: if more than 10 users connect the TAO CAPI system becomes unstable, and it´s usually stops working. The possible solution could be the refactoring of the database connections.

Thibaud Latour & Patrick Plichart	A TAO roadmap for 2011 and beyond: addressing performance, scalability, security and new features while preserving flexibility

As a general purpose CBA platform, TAO architecture requires careful attention to guarantee its usability in very different contexts. In order to do so, TAO should propose sound and good foundations for the development of the required functionalities. However, in addition to specific functionalities characterizing the context of use, non-functional aspects of the platform should also be carefully addressed. The purpose of this presentation is to give an overview of the salient challenges that needs to be tackled in terms of scalability and security on the non-functional side, especially to suit large-scale assessment, and advanced result analysis on the functional side. First results pertaining to scalability at the level of data model and storage will be presented, alongside with next steps on this topic. Finally, the general roadmap towards the new versions of TAO in 2011 will be presented.