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

2ND SZEGED WORKSHOP ON EDUCATIONAL EVALUATION

ABSTRACTS

14-15 April 2010

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

Session A: Using the possibilities of ICT measuring problem solving

Michel Dorochevsky	*Building Rich Problem Solving Items with the CBA Item Builder*

The CBA Item Builder developed for the DIPF has been extended in the last 6 months with features supporting the development of a new class of problem solving items, in particular incorporating dynamic systems (MicroDYN) and finite state machines (MicroFIN). These types of items will also be used in the PISA 2012 study. We will demonstrate the new functionalities and report from our latest experiences including the PIAAC study. Although the power of the CBA Item Builder has increased, in particular the possibility to specify logic and behaviour in items, we still believe that our main goal has not been compromised: providing an authoring tool which can be handled without any programming knowledge. We will report from experiences of early advanced item developers.

Joachim Funke and Samuel Greiff	*Dynamic Problem Solving: A New Perspective for Large-scale Assessments*

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 context allows the measurement of DPS to be considered seriously.

We present MicroDYN and MicroFIN, two innovative approaches to capture DPS based on different formalisms that each provides an infinite item pool. Advantages of the formal frame of reference, first empirical results, theoretical considerations and software issues are thoroughly discussed. Additionally, we discuss milestones to be overtaken when applying dynamic systems in a large-scale context to 15-year old students and specify the steps necessary to take to measure DPS as individual competence in the 21st century.

Session B: Reading, Mathematics and Science Frameworks for Grades 1-6 in Hungary

Csaba Csíkos	*The Evaluation of the Application of Mathematical Knowledge in Classroom and Everyday Contexts: the Literacy Dimension of the Mathematical Framework in Grades 1-6*

This presentation aims at describing the theoretical basis for developing the Assessment Frameworks and Standards for Mathematics (hereafter: Framework) in grades 1-6. There are three main sources of the educational objectives intended to be covered by the Framework. One of these sources is the system labeled as the „literacy" dimension of the educational objectives: societal needs and expectations about the application of mathematical knowledge.

The presentation has three main parts: (1) historical and philosophical roots for both curricular issues and classroom practices of knowledge application. The New Math and the Realistic Mathematics Education movement have been two attempts to challenge the so-called "traditional" mathematics education approach. (2) The societal needs and curricular expectations in other school subjects point to the application of and the transfer gained from mathematical knowledge. (3) The evaluation of the application of mathematical knowledge has been extensively researched in the field of word problems. In the Framework, there are four word problem categories defined within the application dimension: (i) bare tasks containing mathematical symbols, (ii) archetype or pseudo-realistic tasks, (iii) realistic tasks enabling for horizontal mathematization, (iv) authentic tasks emerging from real-life situations and expectations. Sample tasks covering each categorie will be provided.

Erzsébet Korom	*Assessment Dimensions of the Science Framework for Grade 1-6*

The paper shows the theoretical foundations and assessment dimensions of the scientific evaluation standards. (1) The disciplinary dimension focuses on the content knowledge and aspects of science teaching and describes the features of expert knowledge and the processes of knowledge acquisition, conceptual development and conceptual change as well as. Content knowledge is manifested in two different ways in detailed standards: along scientific disciplines (Non-living systems, Living systems, Earth and the universe) and along the common basic concepts, facts and relationships of different scientific disciplines (matter; energy; structure and function; systems and interactions; continuity and change; scientific investigation; relationships between science, technology and society) reflecting the integrative approaches of science teaching. (2) The psychological or reasoning dimension emphasizes the contribution of science learning to the development of basic thinking skills and of general intellectual abilities, on the basis of research findings in the fields of learning and the development of cognitive abilities. Requirements of assessing basic thinking skills and the domain-specific elements of scientific reasoning will be developed in the detailed standards. (3) The literacy dimension reviews various interpretations of science literacy and describes the societal needs and expectations about the application of acquired scientific knowledge in different contexts (school and everyday life), emphasizing the importance of knowledge transfer and the understanding of interactions between science, technology, society and environment.

Edit Katalin Molnár	*Developing Standards for Reading: A Hungarian Case*

Developing standards is a novel enterprise in Hungarian education. The presentation outlines the main aspirations and some of the issues to be resolved. The objectives include what is expected: (1) providing a theoretical, research based framework; (2) developing detailed standards for reading in grades 1-6; (3) building a corresponding test bank. The issues to be resolved mostly concern the new aspects the objectives entail in international and Hungarian contexts. (1) The theoretical framework covers three areas: psychological components and processes; social and cultural expectations; and school based contents. These widely overlap however, they may yield different interpretations of what it means to read. At present there exists a gap between the needs for, and the traditions of, reading education, the bridging of which has been attempted and contested in different ways. (2) The detailed standards are to be developed with the inclusion of the perspective of special needs students. The standards aim to promote the continuous development of elementary skills and facilitate the emergence of more complex abilities necessary to process more complex texts. The inclusion of reading motivation in the standards seems both necessary and problematic. (3) There is not enough research evidence yet to provide a firm basis for developing items assessing the comprehension of Hungarian sentences. Computer-based testing presents further challenges.

Session C: Assessing and Training of Arithmetic Competencies

Annemarie Fritz-Stratmann	*A Developmental Stage Model for the Arithmetical Development in Age 4-8*

A common theme of conceptualizing competence and developmental models is to find hierarchical structures of abilities, which can be interpreted as developmental lines or sequences. Which abilities are basal, which are based on others? Such a developmental sequence based on when certain concepts are usually understood and how they are based on each other, would allow researchers and teachers to allocate every child to a certain developmental level. This allocation would allow age cohort comparison and would thus help to find out whether children are accelerated, normal or retarded. Furthermore one could predict the next developmental step children have to take to move on to progress to the next level.

In the following is an effort of doing such modeling and its empirical coverage concerning arithmetical learning for children aged four to 8 years. A main question of this modeling was to find out which mathematical concepts are basal for the understanding of mathematics and how those are based on each other. This first talk shall introduce the theoretical foundation of this model, while consequent talks of Dominique Arndt and Katleen Sahr will give empirical evidence.

Based on literature and empirical results concerning arithmetical learning the following five stage model was created (Fritz, Ricken, Gerlach, 2007):

Stage I: The ability to distinguish small sets and to count and enumerate them.

Stage II: The ability to name predecessor and successor of a given number on some kind of mental number line and to solve small addition tasks by counting or use of the number word sequence.

Stage III: Understanding the connection of number and set in a cardinal number concept.

Stage IV: Part-part-whole concept, organizing the knowledge of breaking and assembling sets.

Stage V: The understanding of congruent intervals between the numbers of the number line (relational numbers).

Dominique Arndt	*Assessment of Early Arithmetic Skills Based on a Developmental Stage Model*

The developmental stage model of Fritz, Ricken and Balzer (2009) allows teachers and researchers to describe and understand the arithmetical development of children aged four to eight years. Based on this model it should be possible to generate tasks, which cover the very core concepts of each stage of the model. Such an ex ante assignment would allow to verify the validity of the model, if proven correct by ex post analysis using Rasch models and task difficulty analysis. Therefore several tests were developed in the context of a research cooperation. Within this project a screening test for groups up to five children (with subtests for kindergarten children, first and second graders) and an extended version for individual testing and more precise diagnostics were developed and evaluated. Both tests fitted the Rasch model (Weighted Fit MNSQ < 1.2 for all items) with acceptable reliability (Screening: .864; individual testing: .945). Estimates for the variables were ordered in a way predicted ex ante by the model. The correlation between both tests was $r=.85$ ($p<.001$), although the group setting slightly underestimated performance of most children systematically. The tests developed are thus applicable to the developmental stage model and are therefore suitable to assess the arithmetical development of four to eight year old children.

Katleen Sahr	*Training of Basic Arithmetic Competencies with 5-to 8-year-old Children*

Early support of basic arithmetic competencies is important for the general development and academic career of children. For this purpose, we conducted a longitudinal study in which we screened N = 1298 children with an average age of 80.3 months (SD = 14.0) for their basic arithmetic skills according to a five-stage developmental model recently proposed by Fritz, Ricken and Gerlach (2007). In addition, cognitive abilities were assessed by means of Coloured Progressive Matrices. Two months later arithmetic skills and cognitive abilities as well as working memory and language scales were assessed in detail for 204 low achieving children who just reached stage I, II or III. Then these children were assigned to one of four training groups: (1) math training (MARKO-T), (2) working memory training, (3) social skills training, and (4) control without any training. Our aim was that each child reaches at least one stage above its starting position. Each training program consisted of up to 20 one-hour training sessions administered within eight weeks. The math training (MARKO-T) was based on the stage model by Fritz et al. and was conducted adaptively in individual settings: That means each child proceeded to training problems of the next stage if and only if it had mastered the previous stage. The working memory training was conducted in group sessions, whereby each child worked individually on a computer. The social skills training was also conducted in group sessions and included games and communication tasks. After the training, arithmetic skills and cognitive abilities as well as working memory and language were assessed again. ANOVA results revealed that the math training had a positive effect, $F(3,176)= 17.63$; $p<.001$; part. $\eta 2 =.23$. Thus, our study showed that MARKO-T, as a relatively short but distinctive and adaptive training of basic arithmetic competencies, can lead to substantial gains in arithmetic skills.

Session D: Online assessment

Benő Csapó	*The First Phase of Developing an Online Diagnostic Assessment System*

A research and development project at the Center for Research on Learning and Instruction, University of Szeged aims at devising an online diagnostic assessment system for the first six years of primary school. The project was launched in 2009. The activities of the first phase of the project are organized into seven work packages: (1) Devising assessment frameworks for reading, mathematics and science; (2) Exploring diagnostic assessments at further cognitive and affective domains; (3) Developing item banks in reading, mathematics and science; (4) Creating a platform for online testing; (5) In-service training of teachers to prepare them to use the system; (6) Devising diagnostic assessment instruments for SEN students; and (7) Meta-analysis of the data of national and international assessments. TAO is used as a delivery platform, and the system will be implemented in ca. 200 partner schools. This presentation outlines the general theoretical background and the structure of the activities of the project; and presents some of the first results of the pilot work.

Krisztina R. Tóth	*Comparing Paper- and Computer-Based Test-Taking Processes*

Technology-based testing makes it possible to consider information about the testing process. These pieces of background information are stored in automatically generated .log files and can be used for better understanding of students' performance, monitoring testing processes, calibrating online item banks.

In the past years, we accomplished several pilot studies to assess media-effect, to identify differences between paper-and-pencil (PP) and computer-based (CB) test results of the same tests or of parallel testversions. This paper presents the results of these assessments with emphasis on testing processes. Fifth (11-year-olds) and sixth graders (12-year-olds) were involved in the measurement in 2008 and 2009. Different tests were used to compare testing processes in CB and PP medium. Reading comprehension, inductive reasoning and mathematics were assessed in CB and PP format. The online data collection was carried out with the TAO platform.

This presentation will demonstrate whether there are any differences in solving multiple-choice or open-ended tasks in computer-based and paper-based environment.

Edit Tóth	*Teachers' Attitudes Towards Assessments and Accountability: First Results of a Questionnaire Administered to Primary School Teachers*

In the last twenty years the assessment of effectiveness and efficiency, the notion of accountability have become an important question in the economically developed countries.
In OECD countries the evaluation of students by standardised tests plays a significant role in the system monitoring of public education. The increased use of achievement assessments may have a lot of possible positive and negative consequences on schools and classrooms.
In our presentation we aim to show elementary school teachers' beliefs about the effects of efficiency assessment on teachers and the teaching process. We also present the teachers' views about testing in a computerised environment. Data collection was carried out electronically by means of the TAO system with a partly adapted and partly self-designed questionnaire (Nteacher=890, Cronbach-α=.81–.90).

Teachers think that it is important that School Achievement tests, which are important for schoolwork, should regularly be done but at the same time they think that they do not provide an objective basis to evaluate schools. There are many forms of evaluations however, it seems that the National Assessment of Basic Competencies programme puts the biggest pressure on teachers. Moreover, teachers also experience anxiety when it comes to Standardized international tests and centralised final exams. Data reveals that teachers are most likely to be pressurised by their own and the school management's expectation and they consider students' and parents' requirements to be the smallest burden when it comes to increasing their students' performance on the National Assessment of Basic Competencies programme. Thanks to the results, teachers' work has changed in many ways. Although two thirds of the teachers never use computers to assess their students' knowledge, almost 90% of them would use it, depending on the subject, if they had the possibility.

Session E: Tools and Methods of Educational Data Mining

Karl Devooght	*Educational Data Mining: A Further Step Toward Student Modeling*

Educational Data Mining is an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings. The using of those methods aims at better understanding students' behavior and at proposing more and more Educational IT Supports which meet needs of both students and teachers. We propose an overview of prominent results based on actual techniques in this field. Among existing topics, we will focus on two promising subjects: i) Log Analysis for Student Modeling; ii) Predictive Student Assessment.

Barkow Ingo	*Metadata Editor- A Documentation Tool Developed for the German National Educational Panel Study*

The National Educational Panel Study (NEPS) - financed by the German Federal Ministry of Education and Research is an educational longitudinal study with a planned running time of decades and a very complex design. This means for the German Institute for International Educational Research (DIPF) who is responsible for the programming side of this project a challenge in data processing, data archiving and general database structure. The demands resulted in developing a upcoming Data Warehouse structure for analysis and reporting purposes as well as the decision to use DDI3-compliant metadata storage.
The session will show the progress of database developments in NEPS since the last SWEE and introduce the Metadata Editor as work in progress. Furthermore there will be information about planned DDI-compliant extensions to this tool and the Data Warehouse structure in general. Also included is a description of the DDI export (and reduced import) functionality via web services and the database layout of the project describing the connection to DDI. Furthermore there is a possible outlook on the connection between the existing TAO Transfer XML-based CAPI instrument for PIAAC and the NEPS Metadata Editor including complex routing and dynamic text capabilities. Also mentioned will be the preparation for accessing the stored information about items and instruments for later analysis and reporting purposes.

Session F: Formative and summative assessment

Eric Ras	*How Can Formative Assessment Contribute to the Effectiveness of Personal and Adaptive Learning Environments of Today?*

Since a few years, a lot of research effort has been spent on context-aware (informal) learning approaches that adapt the learning content, instructional design, and the learning activities to the needs, preferences, and situation of the user. Personal learning environments are created by learner themselves and intelligent recommender systems provide any kind of learning resources to the learner. These developments have also had a positive impact on the field of computer-based assessment (CBA) as well: New application scenarios have been created, and new skills and competencies need to be measured. This talk will emphasize on how formative assessment can be a mean to enhance learning in general and how it can increase the quality of evidence for intelligent adaptations in technology-enhanced learning settings.

Oliver Wilhelm	*Measuring Progress in Student Achievement*

Educational assessment is usually restricted to measures collected at a single time point. When it comes to gaining estimates of learning and achievement changes this is clearly insufficient. For many applied purposes such estimates are warranted. However, gaining such estimates is laborious and challenging. Nevertheless, change in achievement is the parameter of pivotal applied interest. In the present study we investigate antecedents, covariates and consequences of changes in student achievement.

Our research questions are focused on understanding change in student achievement in German (mother tongue) and Mathematics. About 150 9th graders were recruited. Apart from two pre- and two post-test sessions they complete 40 test sessions of two hours – approximately every 2 weeks. In each testing session participants complete nationally calibrated measures of student achievement in German and Mathematics. Additionally, they complete two blocks of automatically generated parallel tests of three working memory tasks (N-back, Alpha-span, Memory updating). Finally, participants complete a variety of questionnaires on school-related behaviour and achievement related personality traits repeatedly. A control group only completed the pre- and post-test sessions.

In preliminary analysis we found growth in all domains. Change factors for mathematics and working memory from growth curve models were positively correlated. We will report results from 22 measurement points (the first year of the study) focusing on change in Mathematics achievement and its covariates.

We will discuss the structure of Mathematics achievement and its covariates by focusing on the status quo and changes in terms of inter- and intra-individual variation. We will suggest new applications in the monitoring of student progress and discuss hints on early indicators of school problems.

Sirkku Kupiainen	*Fifteen Years of Learning to Learn Assessment in Finland*

Learning to learn, one of the eight key competences in the EU framework for lifelong learning, was written into the evaluation strategy of the Finnish National Board of Education (NBE) in 1995 as one of the salient outcomes of basic education. As the concept was then new in the Finnish educational discussion, the NBE commissioned a group of researchers at the University of Helsinki to construct an indicator for its measuring. Already next year, the Centre for Educational Assessment (CEA) piloted an instrument based on a framework resting on earlier literature, loosely following Hirsh's notion of "tool conception of education" (Hirsh 1996) as a set of affective, conative and cognitive aptitudes formed through good teaching.

The cognitive dimension of the developed instrument rests on the developmental psychological understanding of the malleability of thinking (Adey, Bruner, Demetriou, Gardner, Klauer, Nuthall, Piaget, Resnick, Shayer, Sternberg) and is measured with tasks based on and resembling—but not equivalent to—those used in learning at school. Due to the practical requirements, the measuring of the affective and the conative dimension is done using self-report questionnaires, based on a socio-cultural understanding of the 'epistemic mentality' (Claxton, Cole, Galperin, Leontjevv, Lompsher, Vygotsky) on the one hand, and on motivational, volitional and self-concept traditions (Harter, Little, Pintrich, Rosenberg, Shavelson, Skinner) on the other, formulated to reflect the central role of the school as the societal setting for learning for the targeted age groups. In this, the Finnish research follows Snow's model for educational assessment (Snow 1990, 1994).

However, the understanding of learning to learn as an outcome of the application of the mental tools (thinking skills, emotional and affective dispositions) student s have at their disposal to the task at hand leads to strong co linearity of the measured dimensions, accentuated by the lack of a clear future criterion for "success". Also, due to the practical requirements of large scale assessment, the instrument does not cover all aspects of learning to learn with the same rigorousness.

In the present paper, we will present some key findings from the large body of both cross-sectional and longitudinal studies implemented during the past 15 years, covering some 80 000 students of different ages, highlighting some of the salient problems we are currently working on regarding the assessment of the complex phenomenon of learning to learn in classroom setting.

LIST OF PARTICIPANTS

Name	Institute
Adey, Philip	King's College, UK
Arndt, Dominique	University of Duisburg, Germany
Barkow, Ingo	DIPF, Germany
Binkley, Marilyn	University of Luxembourg, Luxembourg
Bogaerts, Jerome	University of Luxembourg, Luxembourg
Csapó, Benő	University of Szeged, Hungary
Csíkos, Csaba	University of Szeged, Hungary
Devooght, Karl	University of Luxembourg, Luxembourg
Dombi, Judit	University of Pécs, Hungary
Dorochevsky, Michel	SoftCon IT-Service GmbH, Germany
Fritz-Stratmann, Annemarie	University of Duisburg, Germany
Funke, Joachim	University of Heidelberg, Germany
Greiff, Samuel	University of Heidelberg, Germany
Habók, Anita	University of Szeged, Hungary
Horváth, Veronika	University of Pécs, Hungary
Józsa, Krisztián	University of Szeged, Hungary
Kinney, Mark	University of Toledo College of Education, USA
Kinyó, László	University of Szeged, Hungary
Korom, Erzsébet	University of Szeged, Hungary
Kupiainen, Sirkku	University of Helsinki, Finland
Latour, Thibaud	Public Research Centre *Henri Tudor*, Luxembourg
Menyhei, Zsófia	University of Pécs, Hungary

Name	Institute
Molnár, Edit Katalin	University of Szeged, Hungary
Molnár, Éva	University of Szeged, Hungary
Molnár, Gyöngyvér	University of Szeged, Hungary
Molnár, Pál	University of Loránd Eötvös, Budapest
Nagy, Lászlóné	University of Szeged, Hungary
Nikolov, Marianne	University of Pécs, Hungary
R. Tóth, Krisztina	University of Szeged, Hungary
Ras, Eric	University of Luxembourg, Luxembourg
Reeff, Jean-Paul	DIPF, Germany
Sahr, Katleen	University of Duisburg, Germany
Szenczi, Beáta	University of Szeged, Hungary
Tóth, Edit	University of Szeged, Hungary
Wilhelm, Oliver	Humboldt University of Berlin, Germany
Wüstenberg, Sascha	University of Heidelberg, Germany
Zsolnai, Anikó	University of Szeged, Hungary