Simple Thinking using Complex Math vs. Complex Thinking using Simple Math – A study using Model Eliciting Activities to compare students' abilities in standardized tests to their modelling abilities.

Steffen M. Iversen, University of Southern Denmark
Christine J. Larson , Indiana University

Abstract: *Traditional mathematics assessments often fail to identify students who can powerfully and effectively apply mathematics to real-world problems, and many students who excel on traditional assessments often struggle to implement their mathematical knowledge in real-world settings (Lesh & Sriraman, 2005a). This study employs multi-tier design-based research methodologies to explore this phenomenon from a models and modeling perspective. At the researcher level, a Model Eliciting Activity (MEA) was developed as a means to measure student performance on a complex real-world task. Student performance data on this activity and on traditional pre- and post-tests were collected from approximately 200 students enrolled in a second semester calculus course in the Science and Engineering department of the University of Southern Denmark during the winter of 2005. The researchers then used the student solutions to the MEA to develop tools for capturing and assessing the strengths and weaknesses of the mathematical models present in these solutions. Performance on the MEA, pre- and post-test were then analyzed both quantitatively and qualitatively to identify trends in the subgroups corresponding to those described by Lesh and Sriraman.*

ZDM-Classification: M15, D65, C80

1. Introduction

This study was inspired by two questions formulated by Lesh & Sriraman (2005a, p. 7):

"Why do students who score well on traditional standardized tests often perform poorly in more complex "real life" situations where mathematical thinking is needed? Why do students who have poor records of performance in school often perform exceptionally well in relevant "real life" situations?"

Lack of transfer is one possible answer to the first question. According to Niss (1999), knowledge of mathematical theory does not guarantee that this is transferred to an ability to solve non-routine "real life" problems involving mathematics, but there is evidence that it is possible to design teaching settings that foster and solidify such abilities in students.

While this is a plausible answer to the first question, the second question remains unresolved. Furthermore, one might ask whether these two questions are indeed linked.

A teaching experiment was designed to try to shed some light over these issues. The questions addressed in this paper are:

1. What are the connections between students' abilities in standardized tests and their abilities working with messy "real life situations" involving mathematics (i.e. situations where mathematical modeling is emphasized)?

2. If there are examples of students who perform poorly in standardized tests and perform well on more situated real-life situation involving mathematics (as claimed by for example Lesh & Sriraman, 2005a), how can we explain this?

2.Theoretical framework

There is a general agreement among teachers and researchers that mathematical modeling is an important aspect of the mathematics education (e.g. Gravemeijer & Doorman, 1999; Lesh & Doerr, 2003; Burkhardt, 2006). Over the last few decades, applied mathematics has played an increasingly important role in the development of other scientific disciplines such as engineering, nanotechnology, economics, and biology. Many educators and researchers in mathematics and mathematics education believe that this fact should be reflected in classrooms through the implementation of student activities involving mathematical modeling. Such activities should (ideally) provide the students with tools usable beyond school, and give them a peek into the "real life" mathematics used outside the classroom.

2.1 Model Eliciting Activities

We agree with Wittmann (1998), Burkhardt (2006) and Lesh & Sriraman (2005b) that mathematics education can be viewed as a *design science*, and this study was conducted therefore as a *multi-tiered* teaching experiment (in the sense of Lesh and Kelly, 2000), centred around an open-ended modelling task which falls under the category of *Model Eliciting Activities* (MEAs) as described by Lesh & Doerr (2003). MEAs are operationally defined as problem solving activities that are constructed using six concrete principles of instructional design. Each of the principles invoke a related question as follows: 1) the Reality Principle – Does the situation appear to be meaningful to the students, and does it build on to their former experiences? 2) the Model Construction Principle – Does the situation create a need to develop significant mathematical constructs? 3) the Self-evaluation Principle – Does the situation demand that the students assess their own elicited models continuously? 4) the Construct Documentation Principle – Does the situation demand the students to reveal what they think while they work on solving the problem? 5) the Construct Generalization Principle – Can the elicited model be generalized to other similar situations? 6) the Simplicity Principle – Is the problem situation simple? (Lesh & Doerr, 2003).

2.2 Testing & assessment

According to Niss (1999) many of the assessment modes and instruments that are currently used in mathematics education do not provide proper insight into students' abilities - what they know, understand and are able to achieve. This is especially true in the case of higher order knowledge and insights. Watt (2005) also emphasizes that students' abilities to solve meaningfully contextualized problems using higher order skills is not considered properly in standardized tests.

Standardized tests tend to focus on a small subset of the ideas, understandings and abilities that are emphasized in more open-ended task-based problem solving situations (Resnick & Resnick, 1993), and according to Clarke & Lesh (2000) they actually *"tend to emphasize what is not important, while neglecting what is important."* (p. 117). This is because traditional tests often focus on a number of short-answer questions that primarily invoke the usage of low-level facts and skills while representative work samples (as for example MEAs) tend to focus on problem solving situations that involves a smaller number of "big ideas" that involve higher order understanding and abilities (Clarke & Lesh, 2000).

Leder et al (1999) points out that using only one type of assessment will reduce the chances of high achievement for the students that are able to show their mathematical abilities and knowledge through other types of assessment. Not being able to identify through standard tests the group of students that show great abilities in mathematical modeling also means that a wide range of abilities and knowledge belonging to an unidentified group of students are not appreciated and nurtured properly. Standard tests in this way assess a non-representative subset of goals and objectives in mathematics education that can not be extrapolated to the whole range of mathematical abilities and objectives (Watt 2005, Stephens, 1988). This also means that students performance in standard tests cannot in general be used to infer more general mathematical abilities (see also Goleman, 1997). Unfortunately research shows that alternative assessment forms in mathematics are often regarded as highly subjective among teachers (Watson, 2000), in spite of the fact that this need not be the case (Watt, 2005).

3. Presentation of the teaching experiment

The teaching experiment was conducted at the University of Southern Denmark[1] in the winter of 2005. The experiment took place during a 7 week calculus course and involved about 200 students (178 students took the pre-test, 201 did the MEA problem, and 199 students took the final exam). The students were all first year students at the University's *Faculty of Science and Engineering* and the calculus course was their first encounter with the subject of Mathematics at the tertiary level. Before

[1] Situated in Odense, Denmark.

entering the Faculty of Science and Engineering a student has to have completed the highest level of mathematics[2] in the Danish upper secondary school and it was therefore assumed that all the students that participated in the experiment had done this.

The experiment was carried out during the first class session of the semester. During this session, students worked individually on a standardized pre-test (described below in section 3.1). After completing this examination, students worked in small groups toward a solution to the Model Eliciting Activity entitled "The Penalty Throw Problem," which is described in more detail in section 3.2. Each group was required to submit in writing their final solution to the MEA. The pre-test and the MEA were observed by one of the researchers as well as a set of teaching assistants. After completing the whole of the calculus course the majority of the students took the final calculus exam and the students' results from this official examination was also incorporated into the study described in this paper (see section 3.3).

3.1 Designing the pre-test

Before the modelling activity was introduced to the students, they all took a standardized pre-test consisting of 11 problems. The problems varied in subject from simple vector calculus to reducing algebraic expressions to integrating simple functions. The test was based on the official mathematics syllabus from the upper secondary schools in Denmark. The official written exams from the Danish Ministry of Education from the last three years (2003-2005) were analysed, and from this a collection of problems was identified as typical for these examination sets[3]. A collection of 11 problems were then designed based on these standard problems from the official sets so that the problem posed in the teaching experiment in no had the same difficulty level as the problems posed in the official exams. Both when taking the mathematics exam at the upper secondary school and when doing the pre-test, no aids in the form of calculators, books or notes were allowed (although calculators were allowed during the MEA as well as during the final exam). Each student took the pre-test individually, and results were scored using a standardized grading system as is done when the official exams are graded.

3.2 The Penalty Throw Problem – an example of a Model Eliciting Activity

After the pre-test the students were presented to an MEA called *the Penalty Throw Problem*. This problem was designed especially for the teaching experiment according to the six specific principles of model eliciting activities described above. The modelling activity was inspired by *the Volleyball Problem* described by Lesh & Doerr (2003), in which students are asked to put together a number of volleyball teams based on selected individual statistics for the players. One distinguishing characteristic of the Volleyball Problem and the Penalty Throw Problem is the fact that the data available to the students is heterogeneous in nature, consisting of both quantitative and qualitative information. Relevant skills therefore include abilities to a) quantify qualitative information, b) aggregate a variety of different kinds of qualitative and quantitative information, c) accomplish the preceding goals using iterative processes that do not depend on a single formula, d) generate graphs to compare profiles of individual players or teams (Lesh & Sriraman, 2005a, p. 20). According to Gainsburg (2006) this task is essentially a design task, and the models elicited by the students are likely to be complex. Furthermore developing a good model for such a MEA depends not only on recognizing mathematical patterns in the data but also understanding the different aspects of the games of volleyball or handball respectively.

The game of handball was invented in Denmark about 1900 and has grown to be a sport played world wide, with its strongest following in Europe. In Denmark handball is a national sport, and the national teams are closely followed when participating in big tournaments such as the European Championship (EC), the World Cup (WC), and the Olympic Games (OG). The women's national team is very popular in Denmark, mainly because of their overwhelming international success in the last fifteen years.

[2] Referred to as level A or A+ in the Danish upper secondary schools.
[3] The official exams from the last (currently 8) years can be found at http://us.uvm.dk/gymnasie/almen/eksamen/opgaver/?menuid=150560 (in Danish).

Winning the last three possible Olympic gold medals (1996, 2000 & 2004) and also winning both the EC (1994, 1996, 2002) and the WC (1997) have made this national team extremely popular in Denmark. With the popularity of the national team and the subsequent extensive press exposure, it was reasonable to expect most of the students to have some previous knowledge about the game of handball, and the women's national team in particular. It was therefore decided to use this as the subject matter of the MEA. With the women's national team competing in the WC just one month after the teaching experiment was to take place, this particular event was chosen as the perfect frame for a relevant modeling task thereby adhering to the Reality Principle for MEA design mentioned above.

Choosing a very contemporary and realistic setting for the Penalty Throw Problem was no coincidence. Using real contexts enables the students to draw on personal extra-mathematical knowledge, and this may empower the students further (Carraher & Schliemann, 2002). In the study of Gainsburg (2006) it is also concluded that

" *In any field, much of an adult's modeling competence comes from years of experience in the occupational domain. Teachers and curriculum developers therefore, must be care-ful to choose problem situations that are familiar to students and for which they have a basis for developing and judging the appropriateness of their models*." (p. 33).

As in many other sports (e.g. basketball or soccer) scoring on penalties awarded is very important for success in the game of handball. In the Penalty Throw Problem the students were asked to work out a procedure to pick out the three handball players (of an already chosen squad) most suitable for throwing the penalties for the women's national handball team at big tournaments such as the WC. The development of such a procedure requires students to construct a model, thus satisfying the Model Construction Principle. The students were given different sets of data on the players, and it was emphasized that their model should be sharable and reusable in other situations (for example to future tournaments where other players were available) – thus respecting the Construct Generalization Principle. This data allowed students to evaluate their procedure (as is emphasized in the Self-Evaluation Principle). The description of the procedure required of the students comply with the Construct Documentation Principle and the overall simplicity and clarity of the task satisfies the Simplicity Principle. Thus we see that all six principles for MEAs are indeed considered in the development of the Penalty Throw Problem.

The data material handed out to the students included quantitative information, such as data regarding factors as pulse and shooting velocity, from a (fictitious) test of players but also qualitative data in the form of a (fictitious) statement about each player from the player's club coach. It was up to the students to sort the information and use what they found relevant for the development of their models.

3.3 Post-test

At the end of the 7 week calculus course the students had to complete an official standardized 2 hour written exam (performed individually and with the use of algebraic calculators allowed) and these exams were all scored using standard methods. The exam consisted of a number of basic calculus problems that were similar to problems that the students had worked on during the course (as is the case in most standard exams). One of the researchers was engaged in the official scoring process of these official exams and we were therefore able incorporate the students' official exam results into the study where this exam is regarded as a standardized post-test.

4. Analysis of the collected data

Before being able to compare the students' achievements in the pre- and post-test with their achievements in the MEA, a reasonable assessment procedure for the MEA solutions had to be developed. We wanted to use a publicly known and tested assessment tool, so that our evaluation of the students' MEAs was not regarded as arbitrary, and reasonable comparisons to the standardized tests could be made. We therefore chose *The Quality Assurance Guide* (QAG) as a basis.
This assessment tool was developed and tested extensively by Clarke & Lesh (2000) and is

designed to evaluate products that students develop when working with MEAs. The QAG makes it possible to evaluate MEAs in a standardized way, and divides the students MEA solutions into 5 different levels of performance. These levels are based on the fact that every MEA consists of a task where a group must create a generalizable and reusable tool for a clearly identified client. The QAG provides guidelines for determining how well the client's needs are met by the student solution, thus quantifying the quality of the solution.

This provided a quantification of student performance that allowed statistical comparisons to be made with the standardized test scores from the pre- and the post-test respectively.

Since we not only wanted to be able to assign simple scores to the students MEAs, but also wanted to analyse the solutions qualitatively, we needed another assessment tool that could provide us with more insight about the nature of the students work on the MEAs. Again taking the QAG as a starting point, a more detailed evaluation tool was developed. With this tool we were able to identify and evaluate key characteristics of the students MEA solutions. This allowed identification of characteristics common to many of the students' solutions and characteristics that were unique or uncommon.

4.1 Models done by the students

With over 70 group solutions to the MEA problem analysed in this study, it is obvious that not all of the elicited models can be reported in this paper. Still something general can be said about what kind of conceptual tools the students used to mathematize the Penalty Throw Problem. Fig. 1 shows a collection of the mathematical tools most frequently used by the students when solving the posed MEA, along with three typical ways these tools were combined into a mathematical model by the students.

Obviously the model sketched in Example 1 is very simplistic and hardly any mathematical tools were used by students who did this sort of model. The models sketched in Example 2 & 3 were (fortunately) more frequent among the students' solutions to the problem, and are typical examples of how most of the students mathematized the Penalty Throw Problem. Although these two typical procedures were followed by many of the groups, the quality of their solutions varied considerably. As figure 1 suggests, most of the conceptual tools the students used are quite simple mathematically. To infer rankings or scale numbers in relatively simple ways can hardly impress many math teachers at this educational level. Still some of the best solutions to the Penalty Throw Problem used almost only these simple mathematical tools.

If we first take a look at all the students' solutions we can identify certain characteristics that separated the groups that did well on the MEA (here referred to as High MEA Achievers) from the groups that performed poorly in the MEA (Low MEA Achievers). We have identified three aspects. First of all; the High MEA Achievers are characterized by the fact that they consider possible data relations properly. In this specific MEA, such considerations can involve drawing lines of tendency that show which parameters most strongly impact the final model. The Low MEA Achievers mostly considered the available data intuitively.

Secondly there was a big difference in the way the students used representational and presentational media to describe the situation and develop their models. Where the group of High MEA Achievers used different kinds of media (graphs, charts, tables etc.) to develop their models, the group of Low achievers either didn't use any medias at all or only used media that held very little explanation or descriptive power.

Finally almost all the High MEA Achieving groups used some kind of mathematical notation (either CamelCase or standard x & y-notation), whereas the group of Low MEA Achievers used almost no form of mathematical notation in their models at all. So the groups were not necessarily separated by which mathematical tools they used in their models, but also by the way they were able to use (or not use) different (often simple) mathematical tools to describe and interpret the extra-mathematical situation.

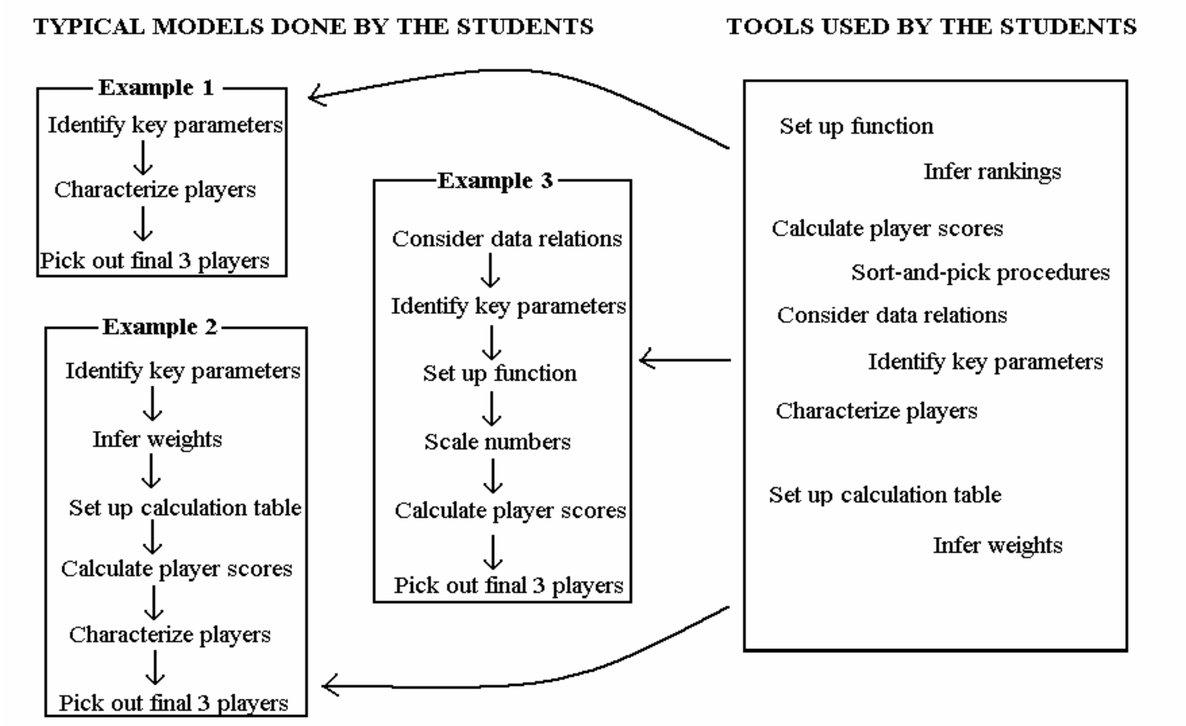

Fig. 1: Some typical mathematical tools used and models elicited by the students

4.2 Quantitative Analysis

One of the challenges in comparing MEA performance with performance on more traditional types of assessments arises from the fact that MEAs are traditionally done in groups. Models and Modeling research often investigates how understandings evolve between students that work in small groups when developing mathematical models. To deal with this issue, we have analysed the data by looking for correlation between individual MEA scores and individual pre-test scores and looking for correlation between group MEA scores and the sum of groups' pre-test scores thereby hoping to draw a more adequate picture of the relations between the two forms of mathematical assessment.

4.3 Correlating Pre-test Score with MEA Score – by Individual

There was no statistically significant correlation between an individual's pre-test score and his/her MEA score (fig. 2).

This lends credence to the claims of Lesh and others that (1) there are many students who are highly capable of applying mathematics to real-life situations in powerful ways – and that these students are often identified as low ability by traditional assessments, and (2) there are students who are identified by traditional assessments as highly capable, yet many of these students struggle to apply their mathematical knowledge to the real-world in a useful way.

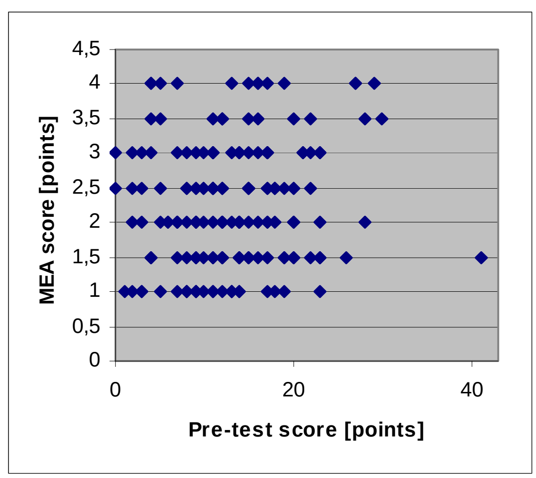

Fig. 2: Students MEA score with respect to their pre-test score [by individual].

To explore this claim, we are interested in the groups of students in both the upper left and lower right quadrants – those who either did very poorly on the pre-test and very well on the MEA or those who did very well on the pre-test and very poorly on the MEA, respectively. These cases may shed some light on why MEAs identify a different set of students as having useful mathematical abilities – both what those abilities are, and how they emerge in this new type of assessment. However, because it is perhaps not appropriate to only compare a score on a group assessment with one received on an individual assessment, we also consider the relationship between groups' MEA score and their aggregated pre-test score.

4.4 Correlating Pre-test Score with MEA Score – by Group

To be able to compare group scores we had to sort the groups by size. The standard group size was three students. For practical reasons this group size could not be preserved for all the students. Much as in the individual analysis, there was little, if any, relationship (correlation of 0.13) between a group's pre-test score (calculated as the sum of the group members' individual pre-test scores) and MEA score (fig. 3).

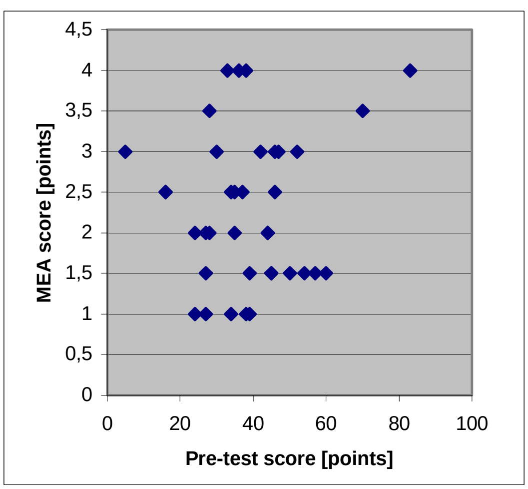

Fig. 3: MEA score with respect to pre-test score [groups with 3 persons].

For reasons discussed above, we will analyze the groups whose aggregated pre-test score was low and MEA score was high and vice versa. We will refer to these two subgroups as "Assessment Jumpers", because the two kind of tests placed these groups in different ends of the assessment scale.

4.5 Correlating pre-test, and MEA score with post-test score

Although far from perfectly correlated, the pre-test is much more correlated with the post-test, than the MEA score with either the pre- or the post-test (fig. 4 & 5).[4]

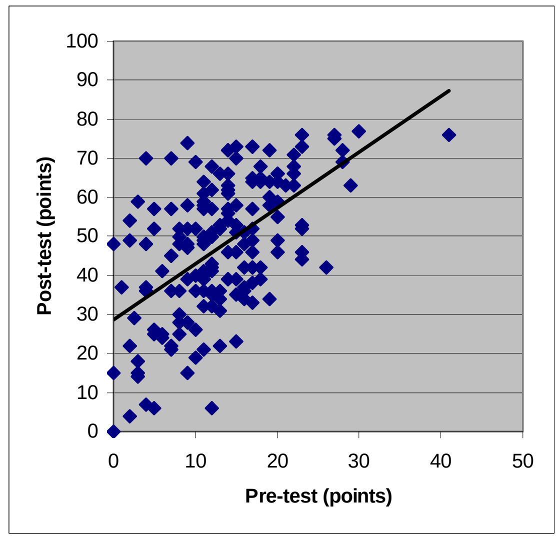

Fig. 4: Students post-test score with respect to their pre-test score [by individual].

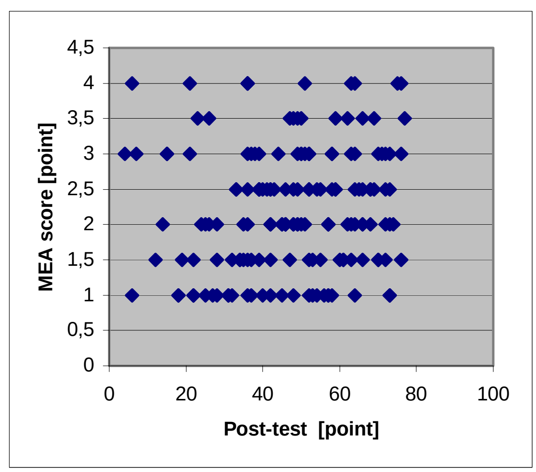

Fig. 5: Students MEA score with respect to their post-test score [by individual].

[4] And thus it seems that the fact calculators were not used in pre-test, but were allowed in the post-test didn't seem to hinder correspondence.

So neither the pre-test nor the post-test did a good job revealing which students were able to do well when working with the MEA problem. Again this supports Lesh's claim that there is no straightforward connection between students who do well in standardized math test and students who are good at working with real-life problems involving mathematical considerations. At least we can conclude that more factors than performance on standardized tests needs to be considered when trying to identify able modellers in the classroom.

4.6 Qualitative Analysis

After finding no significant statistical connection between students who did well on the pre- and post-test and those who did poorly on the MEA (and vice versa), we performed a qualitative analysis of the students' performances on the pre- and post-tests and the MEA. We especially concentrated on analysing the group of students that were referred to as "Assessment Jumpers" earlier on. The first subset consists of five groups who didn't do very well in the pre-test but still managed to give a good solution to the MEA problem. They belong to the upper left quadrant of Fig. 3. In the same way four groups were identified who belong to the lower right quadrant in Fig. 3; meaning that they as a group did well in the pre-test but failed to produce a useful model when working on the MEA problem. These two subgroups were analysed hoping we can say something about the lack of correlation between the MEA scores and the pre-test and post-test scores respectively.

4.7 Low Pre-test Score, High MEA Score Subgroup

The groups in this subgroup used relatively simple mathematics in sophisticated ways so that their solutions, for the most part, took into account all the data. It seemed that the mathematical complexity of the methods used in these groups' respective solutions were correlated with the groups' aggregated pre-test score. In other words, the higher the group's pre-test, the more complex the methods employed. This did not necessarily mean that their mathematical model as a whole was better.

The group with the lowest aggregated pre-test score (5 out of 126 possible points) used a multi-phase ranking system. The next two groups (with scores of 16 and 28) both employed a (linear) weighted scoring scheme that was supplemented with procedures to ensure a well-rounded selection of players. The two groups with the highest pre-test scores (of 33 and 36) both used lines of tendency to find relations between categories of data and identify the most important categories thereby being able to judge which of the available data was the most important to consider in their final models.

Overall, these groups tended to use ranking and weighting in their scoring mechanisms, and some used proportional reasoning and/or lines of tendency to explore relations among the data, and interpretations seemed to be very linear in nature. These groups showed that simple mathematics can be used to create sophisticated solutions to real-world problems.

4.8 High Pre-test Score, Low MEA Score Subgroup

These groups had a tendency to ignore entire categories of data without justification (besides their own intuition), and they also seemed to struggle with advancing their models from initial stages. Little data was used in their actual models and they generally focused on small areas of investigation at a time hoping to fit the data into neat prefabricated mathematical constructions as for example a linear function.

One group attempts to define a linear function of three variables, but uses the same letter (x) to represent all three variables obviously causing difficulties for the reader. Another group does consider relations between differing categories of data but fails to incorporate this into their final solution. Overall, these groups had a tendency to over-trivialize the problem, struggle to move beyond initial ways of thinking, and end up with models that largely involved primarily ranking and picking. They focus on the math – not the development of a model, and therefore often fail to mathematize the situation properly. Furthermore, these groups often failed to check their models against initial data, and therefore didn't go through iterative modeling cycle(s) as is common and desirable for students working on MEAs (Lesh & Doerr, 2003). One possible explanation for this would be that they are accustomed to being right most of the time in math class, so an over-confidence in their ability to solve what they may perceive as an easy problem may actually be hurting their performance. Another possible explanation is that the students with the lowest pre-test score in the group may be the most vocal, so the group solution may more heavily represent their view (a teaching assistant made this observation about one group in particular).

4.9 A pre-test retrospect

Continuing the focus on the so-called "Assessment Jumpers", we can break down their solutions to the pre-test to identify on which problems each of the two subgroups collected their points compared to the whole population of students. A statistical analysis reveals that most students earned the majority of their points on problems 1, 6, 7 and 11 in the pre-test.

The subgroup of "High Pre-Test – Low MEA Achievers" collected their pre-test points in a variety of the problems; the subgroup of "Low Pre-Test – High MEA Achievers" earned the vast majority of their pre-test points on problems 1,6,7 and 11, which were identified by looking at the total population of students as the most common ones for the average student to solve. This analysis refutes one hypothetical explanation: that this subgroup's (Low Pre-test-High MEA) good abilities in the MEA could be attributed to the fact that they were better to solve the more challenging mathematical problems than the average students, but failed to collect the easy points in the pre-test. On the contrary, the subgroup of "Low Pre-Test – High MEA achievers" *does* actually collect the easy points, but fail to solve the harder mathematical problems in the standardized pre-test.

4.10 The importance of students' attitudes, beliefs and motivation

Contrary to the results presented in this paper Maass (2006, p. 136) in the last issue of ZDM through a thorough analysis shows results that indicate that there *do* exist significant relations between students mathematical abilities and their modelling competencies. Of course the results found in the different studies are somewhat dependent on the specific settings for the empirical investigations. Another possible explanation of why no significant correspondence is detected in our study can perhaps be found in the importance of including factors as the students' attitudes towards modeling and mathematics in the analysis of the study.

Maass identifies four ideal types of modellers regarding attitude towards modelling and mathematics. According to our observations during the actual teaching experiment and the results of the study we can probably characterise the two subgroups of "Assessment jumpers" as belonging to Maass' type I – the reality-distant modeller (corresponding to High pre-test – Low MEA) and type II – the mathematical-distant modeller (corresponding to Low pre-test – High MEA).

Maass (2006, p. 138) states that

"Especially a negative attitude towards the modelling tasks basically appeared to hinder the development of modelling performances."

Also Burkhardt (2006, p. 191) concludes that when working with modelling in the classroom

"The performance across mathematics of some low-achieving students shows really substantial improvement – many have been 'turned off' mathematics by its perceived irrelevance to anything that interests them"

Hence one factor that may contribute to the discrepancy between the students' performance on the standardised tests and the MEA is the influence of students' attitudes towards the different kind of tasks which were not considered in detail in our study

4.11 Simple mathematics in complex situations

The qualitative analyses of all the students' elicited models shows that by far most of the models consisted of quite simple mathematical elements such as ranking, calculating averages, setting up simple functions (linear combinations mostly) etc. This was also the case of the two identified groups of "Assessment Jumpers", although some characteristic differences also can be identified between these two groups. The models done by the “High Pre-Test – Low MEA” groups generally didn’t incorporate much of the available data into their models, and focused on small areas of investigation at a time. Furthermore, these groups had a tendency to try to “force” more advanced math into their models, without moving beyond initial ways of describing the problem. Although the pre-test demonstrates that the groups are capable at working with quite advanced mathematical objects in the pre-test, they fail to use this when trying to mathematize the real-life problem about handball. Almost the opposite is the case for the “Low Pre-Test – High MEA” group. Here relatively simple mathematical constructs are used in quite sophisticated ways, so that the groups for the most part succeed in incorporating most of the available data into their model. Although the quantitative analysis of their pre-test shows that they collect most of their points in the problems most students were able to solve they succeed to construct mathematical models that involve quite complex thinking and structuring.

These observations offer another explanation regarding the discrepancy between performance on standardized tests and on modelling tasks. An explanation we coin in the following statement

Modeling Eliciting Activities (and perhaps most modeling at this educational level) often involve Complex thinking using Simple Mathematics, whereas standardized test (as the ones considered here) often involves Simple thinking using Complex Mathematics.

The mathematical tools used by the students in the MEA problem are mostly quite simple mathematically, but the context in which these tools are applied are complex. On the other hand, the mathematical tools needed to perform well in the pre-test (and the post-test) were quite advanced, but the contexts to which they should be applied were in fact simple.

Looking again at the entire group of students, one sees that the students who do well when working with the MEA generally use their thinking processes to interpret, organize, and manage their thinking -- rather than just focusing on the mathematical tools involved. They are able to put information together in new and unique ways, and they balance their reasoning with their emotions in a sensible way when constructing their conceptual tools or models. Using creative, critical problem-solving, decision-making and innovative thinking processes, and being able to evaluate, judge and predict consequences this group of students shows characteristics that can best be described as holistic complex thinking. Realizing that among this group of students is in fact students who perform poorly in the standardized tests in the teaching experiment, the study reported in this paper supports the claim of Lesh and others that it is in fact possible for students that are not normally recognized as high-achievers to develop powerful conceptual tools (models) for describing complex systems that use almost only elementary mathematical concepts accessible to students at a broad range of educational levels (Lesh & English, 2005).

5. Conclusions

Our study indicates that it is not necessarily the students who perform well in the traditional testing environments who also do well in more complex problem solving situations such as MEAs. Two possible explanations for this are (1) the influence of students’ task-specific attitudes, beliefs and motivation on performance, and (2) complex thinking using simple math versus simple thinking using complex math. From this we conclude that standardized tests don’t necessarily provide us with a satisfactory picture of students’ mathematical abilities. We are not the first to make this claim, and according to Niss (1999) there is actually an increasing mismatch between the goals of the mathematics education community and their assessments modes. Our study indicates that the consequence of this is that important mathematical abilities held by the students are never tested or even identified. In addition, much teaching in mathematics is designed to prepare students for traditional assessments – failure to test modelling abilities therefore often means not working with mathematical modeling in math class at all!

The modelling activities described in the paper challenged the students to produce valid interpretations of a non-structured environment, which involved significant forms of learning. The students’ interpretations required an intertwining of their existing (mathematical as well as extra-mathematical knowledge) with new knowledge into a complex conceptual system (a model), that in order to function well needed to be shaped and integrated to fit this new situation. The mathematics used for this was often simple – but the thinking processes involved were certainly complex.

In this way our study fits the findings of recent research of the Models & Modeling approach to mathematics – findings that have shown that it is in fact possible for students regarded as average to describe and mathematize complex situations using quite simple mathematics, and thereby develop useful and powerful mathematical models. We therefore agree with Lesh & English (2005) when they claim that attention should be shifted from asking what kind of computations students can perform correctly to asking what kind of situations they can describe productively.

We do not believe that all teaching in mathematics should consist of modeling, but in the light of the study described in this paper we propose that a shift of paradigm is needed when teaching mathematics. Instead of focusing primarily on teaching students to do Complex Mathematics using Simple Thinking, focus should be shifted towards also teaching students Complex Thinking using Simple Mathematics.

References

Burkhardt, H. (2006). Modelling in Mathematics Classrooms: reflections on past developments and the future. *Zentralblatt für Didaktik der Mathematik*, 38 (2), 178-195.

Carraher, D. W., & Schliemann, A. D. (2002). Is everyday mathematics truly relevant to mathematics education? In M. E. Brenner & J. N. Moschkovich (Eds.), *Everyday and academic mathematics in the classroom* (pp. 131–153). Reston, VA: National Council of Teachers of Mathematics.

Clarke, D. & Lesh, R. (2000). Formulating Operational definitions of Desired Outcomes of Instruction in Mathematics and Science Education. In Kelly, A. E. & Lesh, R. (Eds.). *Handbook of Research Design in Mathematics and Science Education* (pp. 113-?). New Jersey: Lawrence Erlbaum & Associates.

Gainsburg, J. (2006). The Mathematical modelling of Structural Engineers. *Mathematical Thinking & Learning*, 8 (1), 3-36.

Gravemeijer, K., & Doorman, M. (1999). Context problems in realistic mathematics education: A calculus course as an example. *Educational Studies in Mathematics*, 39, 111–129.

Leder, G. C., Brew, C., & Rowley, G. (1999). Gender differences in mathematics achievement – Here today and gone tomorrow? In G. Kaiser, E. Luna & I. Huntley (Eds.). *International Comparisons in Mathematics Education* (pp. 213-224). London: Falmer Press.

Lesh, R. & Kelly, A.E. (2000). Multi-tiered teaching experiments. In A.E. Kelly & Lesh R. (Eds.). *Handbook of Research Design in Mathematics and Science Education* (pp. 197-230). Mahwah, NJ: Lawrence Erlbaum Associates.

Lesh, R. & Doerr, H. M. (Eds.) (2003). *Beyond Constructivism: Models and modeling perspectives on mathematics problem solving, learning, and teaching*. Mahwah, NJ: Lawrence Erlbaum Associates.

Lesh, R. & Sriraman, B. (2005a). John Dewey Revisited – Pragmatism and the models-modelling perspective on mathematical learning. In A. Beckmann, C. Michelsen, & B. Sriraman (Eds). *Proceedings of the 1st International Symposium of Mathematics and its Connections to the Arts and Sciences*. May 18-21, 2005, University of Schwaebisch Gmuend: Germany. Hildesheim, Berlin: Verlag Franzbecker, pp. 7-31.

Lesh, R. & Sriraman, B. (2005b). Mathematics Education as a design science, *Zentralblatt für Didaktik der Mathematik*, 37 (6), 490-505.

Lesh, R. & English, L. D. (2005). Trends in the evolution of models & modeling perspectives on mathematical learning and problem solving. *Zentralblatt für Didaktik der Mathematik*, 37 (6), 487-489.

Maas, K. (2006). What are modelling competencies? *Zentralblatt für Didaktik der Mathematik*, 38(2), 113-142.

Niss, M (1999). Aspects of the nature and state of research in Mathematics Education, *Educational Studies in Mathematics*, 40**,** 1–24.

Watson, A. (2000). Mathematics teachers acting as informal assessors: Practices, problems and recommendations, *Educational Studies in Mathematics*, 41, 69-91.

Watt, H. M. G. (2005). Attitudes to the Use of Alternative Assessment Methods in Mathematics: A Study with Secondary Mathematics Teachers in Sydney, Australia. *Educational Studies in Mathematics*, 58, 21-44.

Wittmann, E. (1998). Mathematics Education as a ´Design Science´ . In A. Sierpinska & J. Kilpatrick (Eds.). *Mathematics Education as a Research Domain: A Search for Identity*. (pp. 87-103). Dordrecht: Kluwer Academic Publishers.

Acknowledgements

The authors would like to thank Richard Lesh, Claus Michelsen and Bharath Sriraman for useful advice and inspiration during the preparation of this paper. Furthermore the authors would express sincere thanks to the teaching assistants as well as the Lecturer who teaches the calculus course in which this teaching experiment have taken place. Without their friendly cooperation such an experiment would have been impossible to carry through.

Authors

Steffen M. Iversen
Department of Mathematics and Computer Science
University of Southern Denmark
5230 Odense , Denmark
email: iversen@imada.sdu.dk

Christine J. Larson
School of Education,
Indiana University,
Bloomington,
Indiana, 47405-1006
USAemail: chjlarso@indiana.edu