PLENARY LECTURES

Lin, Fou-Lai

Stacey, Kaye

Sfard, Anna

Reimann, Peter

MODELING STUDENTS' LEARNING ON MATHEMATICAL PROOF AND REFUTATION

Fou-Lai Lin

Department of Mathematics

National Taiwan Normal University

Based on a national survey and some further studies of mathematical proof and refutation of 7th through 9th graders, this paper will show evidence of the existence of continuity between refuting as a learning strategy and the production of conjectures, and between a more effective teaching strategy and the traditional teaching strategy. A detailed analysis of students' refutation schemes will be presented, and a model of their refuting process will be described based on both their refutation schemes and an expert's thinking process on refutation.

INTRODUCTION

Connecting Teaching with Students' Cognition

Research on students' mathematics cognition usually aims to investigate students' thinking and the strategies used, and further to show what guides students' thinking and why the strategies are used. Information about students' cognition can then naturally be applied to redesigning teaching strategies for enhancing students' learning in mathematics classrooms. Both the students' mathematics cognition and the related teaching modules associated with empirical evidence on its effectiveness are meaningful resources for teachers to learn teaching. Indeed, results of research on students' mathematics cognition proved to be key resources for redesigning teaching modules and reforming curriculum to ensure effective learning (Hart, 1980, 1984; Lin, 1991, 2000; Harel, 2002; Boero et al., 1998, 2002; Duval, 2002).

This paper focuses on investigating teaching and learning strategies to connect students' mathematics cognition for enhancing learning on mathematical proof and refutation. We will analyze cognition on proof and refutation in a specific group of students (about one third of their age population). And, for easy implementation in school practices, we chose the coloring strategy for learning proving, and the refuting strategy for learning conjecturing; both strategies are economic and innovative with new thinking. The evidence of using refuting as a learning strategy to generate innovative conjectures shall be presented.

A Research Program on Argumentation and Mathematics Proof

An ongoing two-staged research program on the development of proof and proving is the main reference in this paper. The first stage (2000~03) studied junior high students' understanding of proof and proving. The second stage (2003~07) is studying teaching and learning of mathematics proof. Three phases were carried out during the first stage: instrument development, pilot study, and national survey. Six

2005. In Chick, H. L. & Vincent, J. L. (Eds.). *Proceedings of the 29th Conference of the International Group for the Psychology of Mathematics Education*, Vol. 1, pp. 3-18. Melbourne: PME.

booklets comprising of algebra and geometry questions for 7th, 8th, and 9th graders were developed for the national sampling survey, and the survey involved 1181 seventh, 1105 eighth, and 1059 ninth graders respectively from 61, 60 and 61 classes in 18 sample schools. Most of the items developed in the English study (Healy & Hoyles, 1998) were adopted and modified based on Taiwan students' responses in the pre-pilot study during the first phase of the first stage. In addition, some new tasks were evolved from our interviews, which enabled the features of students' pre-formal reasoning to come through in both the instrument and coding system.

The second stage, teaching and learning mathematics proof, is comprised of an integrated project and four subprojects focusing on algebra (Lin, et al., 2004), geometry (Cheng & Lin, 2005), reading comprehension of geometry proof (Yang & Lin, 2005), and teaching and learning the validity of conditional statements (Yu Wu et al., 2004). The studies are strongly influenced by the work of many current researchers, such as the classification of student proof scheme (Harel & Sowder, 1998) and its application on teacher education (Harel, 2002), the cognitive analysis of argumentation and mathematical proof (Duval, 1998, 1999, 2002), the framework of proof and proving (Healy & Hoyles, 1998), the complexity of students understanding proving (Balacheff, 1987), the function and value of proof (Hanna, 1996, de Villiers, 1991, Hanna & Jahnke, 1993), and the theoretical validation approach of the Italian school (Garuit, Boero & Lemut, 1998).

ONE MORE STEP TOWARD AN ACCEPTABLE PROOF

The Incomplete Proof Group

When the national survey was administered in December 2002, the 9th graders had just learned formal proof in geometry for three months, while the 7th and 8th graders had not yet learned it. Based on the detailed coding schemes, students' performances on geometry proving were regrouped into four types: acceptable, incomplete, improper and intuitive proof. Students missing one step in their deductive reasoning is a typical incomplete proof. Students reasoning non-deductively or based on incorrect properties or with correct properties that do not satisfy with the given premises are codes of the improper proof. Students reasoning based on visual judgment or authority are typical codes of the intuitive proof.

The terminology "acceptable proof" derived from a statement by Clark and Invanik (1997): "Writing, for both students and researchers, is not just about communicating mathematical subject matter. It is also about communicating with individual readers, including powerful gatekeepers such as examiners, reviewers and editors." We took into account teachers' views for assessing whether a proof was acceptable or not.

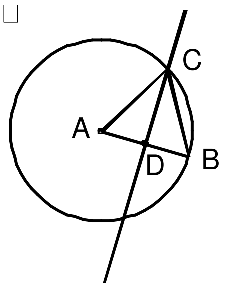
Students in the incomplete proof category were able to recognize some crucial elements for their reasoning (Kuchemann & Hoyles, 2002). They were able to distinguish premises from conclusions in the task setting. Particularly, on the two-step proof items, they were even mindful to check conditions of the theorems applied, i.e., micro reasoning (Duval, 1999.) They were also able to organize statements according to the status, premise, conclusion and theorem into a deductive step. Duval (2002) named such competency as the first level in geometrical proof. The second level is the organization of deductive steps into a proof. From the first step conclusion to the target conclusion, valid deductive reasoning generally moves forward through either successive substitution of intermediary conclusion or coordination of some conclusions. Duval (2002) pointed out that students might have "gaps in the progress of reasoning which makes the attempt of proving failed." This arises either from misunderstanding of the second level organization or from the context of the problem. We shall carefully examine Duval's statement above for the group of students who performed incomplete proofs in the two-step proof tasks.

The data from our national survey showed that one quarter of 9th graders could construct acceptable proofs in a two-step unfamiliar item; approximately one third was able to perform incomplete proofs; and one third did not have any responses at all.

It is obvious that educators would like to focus on this one third of 9th graders who were able to perform incomplete proofs, and to develop a learning strategy for them to fill the gap, i.e., develop one more step toward an acceptable proof. An effective learning strategy should promise that nearly a half of 9th graders will be able to construct a two-step unfamiliar geometry proof.

Incapability of Students with Incomplete Proof Performance

The two-step unfamiliar question used in the survey is as follows.



A is the center of a circle and AB is a radius. C is a point on the circle where the perpendicular bisector of AB crosses the circle. Please prove that triangle ABC is always equilateral.

Two types of incomplete proofs were observed. One type was missing the ending process. Students showed that AC=BC and AC=AB, but did not conclude that the three sides were equal. From a deductive point of view, they were ritually incomplete with the ending process, i.e., if a=b and b=c then a=b=c. Do these students who performed two valid deductive steps still have difficulty in the ending process, a classical syllogism? Or might these students simply be thinking that the two conclusions were too obvious for implying the target conclusion? Should one write this obvious step down? Would this be just an issue in the conventions of mathematical writing? Studies of students' understanding of proof by contradiction (Lin et al., 2002) and mathematical induction (Yu Wu, 2000) showed that senior high students who concluded their proofs without the ending process using either method, very often developed a ritual view about the methods. And the principle of the methods was not understood (Lin et al., 2002). If a teacher considers the two valid deductive steps as an acceptable proof, would the teacher create learning difficulties on mathematical proof for some students? A general question can be asked: How many students who can perform every valid deductive step necessary for a proof task also have difficulty organizing the deductive steps into a proof? Interview data showed that there were students behaving as such.

The other type of incomplete proof was missing one step, either AB=AC or AC=BC. The information "AC is a radius" was implicitly situated within the given premise. This information was invisible for students who did not conclude AB=AC. The property of the perpendicular bisector of a segment seemed unclear for students who did not draw the conclusion AC=BC. Some students of this type might not be aware of the need to derive the equality of all three sides for an isosceles triangle. Thus, the group of students with incomplete proof performance might not be able to:

(1) organize the deductive steps into a proof, or

(2) visualize some implicit information in the given premise, or

(3) recognize a needed mathematics property, or

(4) be aware of all necessary statements/deductive steps.

These four cognitive gaps are due not only to:

(1) misunderstanding of the organization of deductive steps into a proof,

(2) the content of a problem, but also

(3) the context knowledge, and

(4) the epistemic value, i.e., the degree of trust of an individual in a statement, from likely or visually obvious, to a statement becomes necessary (Duval, 2002).

For teaching experiments, one needs to rethink a learning strategy to ensure that students can cross these cognitive gaps.

A Learning Strategy for Promoting One More Deductive Step

Using X as learning strategy for students within their mathematics proof activities is an active research issue. Fifteen paper presentations that dealt with this issue in PME 22~28 are reviewed. The different Xs used in those papers include: arranging the context of proof situations (Garuti et al., PME26) and encouraging interactive discursion to create students' cognitive confliction (Boufi (PME26), Krummheuer (PME24), Douek et al. (PME24), Sackur et al. (PME24), Antonini (PME28)), learning within an ICT environment for conjecturing (Miyazaki (PME24), Gardiner (PME22), Hoyles et al. (PME23), Sanchez (PME27), Hadas (PME22)), emphasizing teachers' questioning as scaffolding (Blanton et al. (PME27), Douek et al. (PME27)), and using metaphors (travel) for setting target goals (Sekiguchi (PME24)). Note that the notation (PME24) indicates the paper appeared in the Proceedings of PME24. We exercised a “thought experiment”(Gravemeijer, 2002) with each of those strategies in addition to typical geometry teaching strategies used in Taiwan secondary mathematics classroom, to match the characterization of the incomplete proof group and enhance them to move one more deductive step. Finally, we chose two strategies that are commonly observed in typical Taiwanese 9th grade geometry classrooms and tested them for helping students achieve one more deductive step. The reading and coloring strategy means that students are asked to read the question, label the mathematical terms, and draw or construct this information on the given figure by color pens. The analytic questioning strategy means that students are asked to reply on what the question asked you to prove, and what conditions in the premise can be useful.

Several phases were conducted in our teaching and learning study:

- *Phase (1):* A three-item diagnostic assessment paper was developed for identifying sample subjects of the focus group. All three items share a common feature with implicitly necessary information.
- *Phase (2):* An instructional interview was conducted on 9 samples individually to examine the effectiveness of implementing both learning strategies simultaneously.
- *Phases (3) and (4):* A small group teaching experiment was carried out to study the effectiveness of only implementing one of the two learning strategies.
- *Phase (5):* A set of learning tasks on geometry proving was developed.

Based on the data resulting from phase (3), we will analyze the function of coloring the mathematical terms in proving. Turning implicit information into explicit information is definitely one function of the strategy. What else happened so that the subjects were able to complete an acceptable proof? It is noteworthy to interpret this with the data collected in the phase (3).

The three items, including the two-step unfamiliar item (G2) used in the national survey, were used in both phases (1) and (2). Nine samples were identified and interviewed. Their performances before the instructional interviews (Pre-I) and after intervening with the reading and coloring strategy (R-C) and analytic questioning strategy (A-C), respectively, during the interviews are presented in Table 1.

The notation (31) denotes the sample who performed an incomplete proof without the ending process due to omission (sample 02) or students’ epistemic value that the ending process is unnecessary (sample 05, 06, 09). The notation 31* indicates that sample 01 would not agree with the syllogistic rule “if a=b and b=c then a=b=c” during the interviews, but agreed that “a=b and b=c” are the conditions for an equilateral triangle with sides a, b and c. The behavior of sample 01 on the syllogistic rule reveals one kind of reason for missing the ending process.

Sample	Performance	G1	G2	G3	G1	G2	G3	Sample
01	Pre-I	0	32	0	4	32	32	06

	R-C	4	32	4		32	32	
	A-Q		31*			(31)	(31)	
	Pre-I	0	(31)	0	0	32	0	
02	R-C	4	4	32	4	32	32	07
	A-Q			4		4	4	
	Pre-I	0	0	32	0	32	4	
03	R-C	4	32	4	4	32		08
	A-Q		4		4			
04	Pre-I	32	21	32	0	(31)	32	
	R-C			4			4	09
	A-Q							
	Pre-I	4	(31)	32				
05	R-C			(31)				
	A-Q							

Note: Definition of codes: 4 denotes an acceptable proof; 31 denotes incomplete, missing the ending process; 32 denotes incomplete, missing one deductive step; 21 denotes improper, using an incorrect property; 0 denotes no response.

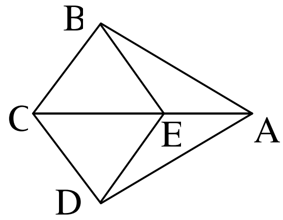
Table 1: Students' performance with/without the learning strategies R-C and A-Q

Table 1 shows that among the 24 (27-3) positions of students' performances which need to move towards an acceptable proof, 15 positions were successfully moved before or after the intervening of only the reading and coloring strategy. Since this coloring strategy is procedural in nature, the cognitive demand on learners for using this strategy is much lighter than using the analytic questioning strategy, which demands quite heavy analytical thinking. So, it is worthy to further explore the extent to which the reading and coloring strategy can enhance students' proving performance. Which kind of proof content will be effective by using this strategy? And a further interpretation of the effectiveness also seems interesting. This is the phase (3) study.

Effects of the Coloring Strategy

During the phase (3) study, four two-step unfamiliar new items were developed for 8 new participants. Before intervening with the reading and coloring strategy, out of 32 (8×4) performances, 10 were acceptable proofs and 22 were unacceptable, i.e., incomplete, or improper or had no response. Each participant had at least two unacceptable performances. One week later, 8 participants worked on the same items after intervening with the reading and coloring strategy. As a result, 16 out of the 22 unacceptable proofs had progressed to acceptable proofs. However, 4 out of 10 acceptable proofs became unacceptable, in which 3 out of 4 negative effects were coded from the same item 3.

Item 3.

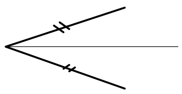


Points A, E, C are collinear,

and △ABC is congruent to △ADC.

Show that: BE=DE

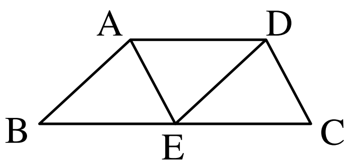
Two students misinterpreted the equality signs labelled on ∠ABC and ∠ADC as ∠CBE =∠CDE. The other student associated the sign around point C, with the angle bisector theorem and applied it improperly. Indeed, colored signs labelling on sub-



figures which cross each other would generate a disturbance that affects visualizers' interpretation on the explicit information transmitted from the sub-figures.

Among the non-effected performance, all six were collected from item 2.

Item 2.



Points B, E, C are collinear,

and △ABE is congruent to △DEC.

Show that: AD//BC

When the equality signs were colored on the six elements, sides and angles of each triangle, the colored signs produced superfluous relations among the elements. Whenever a relation matching his/her target goal was observed by a student, it became active and operational. Students then applied it without justifying deductively. This seemed to be the pattern among those non-effected unsuccessful performances. Analyzing the negative effects and non-effects of the coloring label strategy, a criterion could be used by teachers to restrict the tasks on using the strategy. If a disturbance or superfluous relation from the coloring strategy were intentionally generated onto an item, it may backfire and result in negative effects or non-effects; in this case, the strategy may not be suitable for this item.

Transmission of the Subfigure with Relation to the Theorem Image

In spite of the negative and non-effects of the coloring strategy, we are interested in how the effectiveness (16/22) of the reading and coloring strategy takes place. From neuro-psychological perspectives, "Learning occurs… when transmitter release rate increases make signal transmission from one neuron to the next easier. Hence learning is, in effect, an increase in the number of 'operative' connections among neurons" (Lawson, 2003).

Learning was indeed achieved by those subjects who applied the coloring strategy and were able to perform an acceptable proof. How were the operative connections increased among the statements according to specific status and the use of theorems? The necessary theorems existed previously in the subjects' mental structure, but were inoperative before they applied the coloring strategy. The result of the coloring process revealed subfigures with notable relations that may also correspond to the theorem. If this happens, then learners have increased the relation between the subfigure and the needed theorem. To make it clear, we shall use the term theorem image, similar to the term concept image (Tall & Vinner, 1981), to describe the total cognitive structure that is associated with the theorem, which includes all the mental pictures and associated examples, relations, process and applications. A theorem image is built up over years of learning experiences. It is personal and constantly changing as the individual meets new stimuli. Different stimuli can activate different parts of the theorem image. The stimulus resulting from coloring of mathematical terms in the premise is functioning to lead the transmitter of the revealed subfigure with relation to the corresponding part of his/her theorem image. This leads the effect of the organization of one deductive step.

MAKING DECISIONS ON FALSE CONJECTURES

Some items in each of the six booklets were connected to how students reason to make their decisions on a given false conjecture. Students were asked to make a decision among two (three) choices – agree, disagree, or uncertain (algebraic item) – and then give explanations on their choices. A unity of coding schemes was evolved for both geometry and algebra surveys. The coding schemes were used to analyze the students' performances. Based on this coding scheme, a model of refuting will be discussed. Firstly, for researchers to make sense of the thinking process in mathematical refutation, an expert was interviewed.

Mr. Counter-Example's Thinking Process on Refutation

A mathematician, nicknamed Mr. Counter-Example by his peers during his graduate studies, was interviewed to reveal the thinking process of an expert on refutation.

> "Suppose an unfamiliar mathematics proposition is proposed by myself or peers. Reading it and without having much sense with the proposition, the doubtfulness of its truth usually does not arise in my mind. To make sense of the proposition, very often I'll substitute some individual examples. Then, I will find more and more examples to satisfy the premise. Naturally those examples will be classified according to certain mathematical property. As long as the property is grasped, all kinds of examples will be considered. Finally, a specific kind of example will be identified to counter the conclusion if the proposition is false."

According to Mr. Counter-Example's description, his refuting process covers five sequential processes:

1. Entry
2. Testing some individual examples point-wisely for sense making
3. Testing with different kinds of examples
4. Organizing all kinds of examples
5. Identifying one (kind of) counterexample when realizing a falsehood

This expert's thinking process on refutation can be inferred to analyze students' reasons on refuting.

On Geometrical False Conjectures

Two conjectures in geometry were adopted from the English study (Healy & Hoyles, 1998):

> "Whatever quadrilateral I draw with corners on a circle, the diagonals will always cross at the center of circle?" (7G1, Geometry)
>
> "Whatever quadrilateral I draw, at least one of diagonals will cut the area of the quadrilateral in half?" (8G1, Geometry)

Three false conjectures were evolved from the interviews carried out during the pilot study phase of the first stage. The following one was included in geometry booklets for both 7th and 8th graders who were the subjects concerned in this section.

> "A quadrilateral, in which one pair of opposite angles are right angles, is a rectangle." (7&8 G5, Geometry)

This coding scheme was evolved according to the performances of the national representative sample and the expert's thinking process on refutation, and is more detailed than the schemes developed in the English study (Hoyles & Kuchemann, 2002), which only focused on high-attainers (top 20~25% of the student population).

On geometrical false conjectures, students either confirmed or refuted it. Comparing the frequency on G5 of 7th and 8th graders' performances, there is no evidence of progress with correct decisions over the year (37% for 8th graders, even more than 26% for 7th graders). Based on the words provided by students who ticked disagree, we classified them into three subcategories: rhetorical argument, correcting the given information, and generating counterexamples. Duval (1999, 2002) classified the relationship between a given statement A and another statement B into two types – the derivation relationship and the justification relationship. For each type, there are two kinds of reasoning that are practiced or required in mathematics teaching and learning. Semantic inference and mathematical proof support the derivation relationship; heuristic argument and rhetorical argument support the justification relationship. In our code scheme, codes c2, c3, c4, g1, g2 are the so-called heuristic arguments that take into account the constraints of the situation in the task. Generally, an argument is considered to be anything that is advanced or used to justify or refute a proposition. This can be the statement of a fact, the result of an experiment, or even simply an example, a definition, the recall of a rule, a mutually held belief or else the presentation of a contradiction (Duval, 1999). Reasons relative to the person spoken to or beliefs of the interlocutor are the rhetorical arguments. Therefore, code d4 is a rhetorical argument, and d3 is a heuristic argument.

False Conjectures if P then Q				
Code	Frequency (%)			
	7G1	7G5	8G1	8G5
Confirmation	44	26	31	37
d_0 – Misunderstanding the given information	1	2	2	1
d_1 – Much ado about nothing	23	5	12	9
d_2 – Confirm Q with incorrect reason	9	3	8	6
d_3 – Giving P' s.t. P'→Q	3	12	3	17
d_4 – Authority	0.1		0.2	1.1
Refutation	52	67	68	59
Rhetorical argument	8	8	17	11
Correcting the given information	15	51	12	33
c_0 – Criticizing the given information	9	13	3	5
c_1 – Non-example	3	3	5	9
c_2 – Providing alternative Q		32		16
c_3 – Characterizing Q s.t. P' $\rightarrow Q$	2	3	3	2
c_4 – Empirical decision	0.3		0.5	0.1
Generating (a) counterexample(s)	24	4	34	11
g_0 – Do not believe it is always true	3	1	5	3
g_1 – Giving the possibility of a counterexample	5	0.6	13	4
g_2 – Giving the way of generating a counterexample	4	0.3	4	1
g_3 – Explicit, clear counterexample	12	2	10	3
g_4 – Counterexample with mathematical proof		0.1	0.9	0.1

Note: Non-responses are not included

Table 2: 7th and 8th graders Code Frequencies on items G1 and G5 (N7=1146, N8=1050)

Our coding scheme with code frequencies cover three out of four kinds of reasoning practiced by our 7th and 8th graders on refuting false conjectures: rhetorical argument, heuristic argument and mathematical proof (clear counterexample counts). The relatively high frequency of code c2 in 7G5 was contributed by students who reasoned that under the assumption, a quadrilateral can be either a square or a rectangle. This reason reflects the prevalence of students who misunderstand the inclusion relationship between squares and rectangles. Putting the number of students with codes c2, c3, c4, g1 and g2 together, and computing its frequency, we found that 11% and 36% of 7th graders and 20% and 24% of 8th graders were able to make a heuristic arguments for refuting G1 and G5, respectively.

On Algebraic False Conjectures

Three false conjectures in algebra survey for 7th and 8th graders were chosen for discussion.

A3 "If the sum of two whole numbers is even, their product is odd?" (Both 7th and 8th graders, adopted from Küchemann & Holyes, 2002.)

A6b "The sum of a multiple of 3 and a multiple of 6 must be a multiple of 6?" (8th graders)

The data (3,6,6) in A6b was replaced by (3,6,9) in A6c for 8th graders, and respectively by (2,4,4) and (2,4,6) in A6b and A6c for 7th graders. Students' works on algebraic false conjectures were analyzed with this code scheme: "g3: explicit, clear counter example, can be distinguished into three subcodes," "g31: counterexample without reason," "g32: both supporting and counterexamples," and "g33: counterexample with analytic reasons," which often is a rule for generating a specific counterexample. Referring to the expert's thinking process on refutation, both processes (2) and (5) will be coded by g31. Thus, without words, code g31 could result from primitive or advanced thinking.

Instead of presenting the national survey data, we'll present a brief description of the students' words to model their refutation schemes on algebra. On confirmation: (1) "I believe that only true statements will be presented in my learning" (code d_1); (2) "I consider it correct, because its familiar format is akin to statements in textbooks" (code d_4); (3) "I had supporting examples, e.g., 3+6=9 and 3×2+6×2=18, they are multiples of 9" (A6c) (code d_3). On uncertain responses: (1) "I am not certain because the multiple is not given," students interpreted the term multiple in "a multiple of 3" as specific numbers, a misconception (code r_1); (2) "I had both supporting and counterexamples," in ordinary language, this statement is uncertain (code g_{32}). On refutation performances: (1) "The statement is so elegant, I must have learned it before. But, I did not. So it can't be always correct" (code g_0); (2) Simply adding a negation without reasons (code r_1). Beyond the above beliefs and rhetorical arguments, the students' refutation schemes are coded by g1, g2, g31, g32, g33 and g4. Their thinking process then is similar to certain points in the expert's thinking process.

Refuting Generates Conjectures

When students gave their explanations for refuting, many gave heuristic arguments and explicit counterexamples with reasons, and we observed that some of these students had even produced relations, known properties evidences, general rules, etc. Buying the notion of "Cognitive Unity of Theorems" from the Italian school (Garuti et al., 1998; Boero, 2002), instead of the concerns of the possible continuity between some aspects of the conjecturing process and some aspects of the proving process, we would like to investigate the possible production of conjectures derived from the aspect of the refuting process.

The activity of refuting in mathematics is considered an economic way of helping students to develop competency in critical thinking. Competency of critical analyses has been recognized as a deficit in Taiwan education and is now emphasized in the school curriculum (Ministry of Education, 2003). Two refuting-conjecture tasks in algebra and geometry respectively were developed for the investigation. Each task is comprised of several items. The first item is making decisions on relatively easy false conjectures that aim to motivate students to be aware that the task is on refuting. The second item is given some false conjecture used in the national survey for refuting. The third and fourth items ask students to produce one conjecture and more conjectures, based on their refuting processes.

All nine 7th graders who participated in the investigation with the algebra task produced meaningful conjectures. Three of them even produced a general rule for a whole number m that is divisible by the linear combination of whole numbers $ax+by$.

Seventy-five 9th graders from two classes were asked to participate in the geometry task investigation. The four false conjectures used in the tasks were 7G1 (denotes item G1 in the 7th grade survey), 8G1, 8G5, 9G6, respectively. According to the code of frequencies of refutation schemes, 76%, 73%, 53%, and 60% of their performances were in the category “generating counterexamples” with respect to those false conjectures 7G1, 8G1, 8G5, and 9G6 respectively. The conjectures produced by this group are presented in Table 3.

%	7G1	8G1	8G5	9G6
Thm.	33	20	52	7
New statement	17	8	7	1
Innovation	5	33	8	56
Total	55	61	67	64

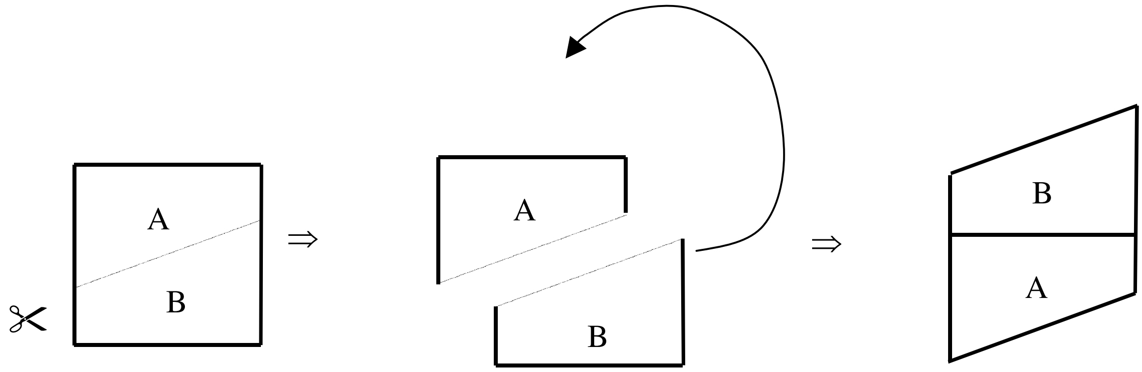
Note: Thm. denotes the conjecture is a theorem. New Statement denotes the conjecture is a new writing of learned properties. Innovation denotes the conjecture is an innovative one.

Table 3: Frequency (%) of different type of conjectures. N=75, 9th graders

Table 3 shows that the success rate for producing correct conjectures on these four tasks was approximately 60% or more. Different frequencies of each type of conjectures imply that 8G1 and 9G6 are excellent for creating brand new conjectures by 9th graders. The item 9G6 is quoted here.

9G6.

A square is cut along the dotted line, then inverted. Is the resulting figure a rhombus?



The conjectures produced by students were further distinguished into “correlating” or “not correlating” to their explanations for refuting.

The relatively high percentages in Table 4 show the continuity of the refuting process and conjecturing process. This claims that refuting is an effective learning strategy for generating conjectures. To create innovative conjectures, the content in the given false conjecture needs to be well-designed, and 9G6 is a good example.

	7G1	8G1	8G5	9G6
T1	40	57	38	69

Table 4: The percentages of conjectures that correlate to refuting

Boero (2002) reported that the Italian school has identified four kinds of inferences, intervening in conjecturing processes: (1) inference based on induction, (2) inference based on abduction, (3) inference based on a temporal section of an exploration process, and (4) inference based on a temporal expansion of regularity. Reading students’ productions in the refuting-conjecture tasks, we observed that false conjectures in numbers 7A3 and 8A6 can enhance the generation of conjectures that are inferences based on induction, abduction (e.g., a narrative) and even deduction (e.g., $3h+6k=3(h+2k)$); the task with figure dissection 9G6 can generate conjectures that are inferences based on a temporal section of an exploration process (the dissection), and tasks with 7G1 and 8G1 are relatively effective on generating conjectures that are based on the expansion of regularity (such as new statements of some properties). The following excerpt is from 9G6.

> If a line cuts a rectangle along the pair of longer sides into two parts so that the cross segment is equal to the longer side, then the two parts can be inverted to form a rhombus.

This conjecture is produced in association with sequential operations on a rectangle.

CONCLUSION

Based on our study, there is evidence showing the existence of continuity in different aspects of mathematics education. In the mathematics learning aspect, a rather high percentage of students were able to produce correct conjectures when working on refuting-conjecture tasks; this shows the existence of continuity between the refuting process and the production of truth statements. For some students, this continuity can even extend to their proving process. Indeed, some students have already provided counterexamples with analytic or mathematical proofs to refute false conjectures. In the mathematics teaching aspect, the effectiveness of the reading and coloring strategy on geometrical two-step proving shows that teachers can keep their traditional teaching approach, in which they can encourage students to label meaningful information within the given premise and conclusion and then seek linkages between the premise and the conclusion. Without disturbing their approach but suggesting students to use color pens for labelling, teachers can enhance students' proving competencies. This demonstrates continuity between a more effective teaching strategy and the traditional teaching strategy. In the aspect of research in mathematics education, there is continuity between the investigating processes by educators in mathematics education research and by mathematicians in mathematics proving. The six phases of mathematicians in proving identified by Boero (1999) is indeed shared by mathematics educators in their studies, such as the study presented in this paper. Formulating on-going investigating issues is always considered to be connected with reflections on previous phases.

Carrying out more testing on the effectiveness of the refuting-conjecture tasks will create an equilibrated set of conjecturing tasks suitable for activating different types of inferences.

Several phases of research in mathematics education presented in this paper are rather traditional, such as (1) Identifying 1/5~1/3 of students in their age population, whose mathematics understanding are more likely to be enhanced. (2) Characterizing those students' competencies. (3) Carrying out an experimental study with a redesigned learning strategy that connects to the characteristics of their cognition.

This approach can frame local (geological and societal) education issues in the wider context of collaborative international studies, for the purpose of improving mutual education. The experience seems to be a very healthy and effective approach.

Acknowledgement

This paper is supported by the National Science Council of Taiwan (# NSC 93-2521-S-003-009). Any opinions expressed here are those of author and do not necessarily reflect the views of the sponsoring institution. The author would like to thank the research team members for their extensive contributions to the study: C.Y. Chen, Y.H. Cheng, I.E. Chen, J.Y. Wu, and K.L. Yang, C.L. Lin, H.C. Hao, and C.H. Chang.

References

Balacheff, N. (1987). Processus de preuve et situations de validation. *Educational Studies in Mathematics* 18(2), 147-176.

Boero, P. (2002). The Approach to Conjecturing and Proving: Cultural and Educational Choices. *Proceedings of 2002 International Conference on Mathematics: Understanding Proving and Proving to Understand,* pp. 248-254.

Boero, P. (1999). Argumentation and mathematical proof: a complex, productive, unavoidable relationship in mathematics and mathematics education. *International Newsletter on Teaching and Learning in Mathematics Proof.* (990708)

Clark, R., & Ivanik, R. (1997). *The politics of writing.* London: Routledge.

Cheng, Y. H. & Lin, F. L. (2005) *One more step toward acceptable proof in geometry.* Symposium Proposal EARLI 2005, Cyprus.

Duval, R., (1998). Geometry from a cognitive point of view. In C. Mammana & V. Villani (Eds.) *Perspectives on the Teaching of Geometry for the 21st Century* (pp. 37-52). Dordrecht/Boston: Kluwer Academic Publishers.

Duval, R. (1999). Questioning argumentation. *International Newsletter on Teaching and Learning in Mathematics Proof.* (991112)

Duval, R. (2002). Proof understanding in MATHEMATICS: What ways for STUDENTS? *Proceedings of 2002 International Conference on Mathematics: Understanding Proving and Proving to Understand,* pp. 61-77.

Garuti, R., Boero, B., & Lemut, E. (1998). Cognitive unity of theorems and difficulty of proof. *Proceedings of PME22,* Vol.2, 99345-352.

Gravemeijer, K. (2002). Developmental research, a course in elementary data analysis as an example, *Common Sense in Mathematics Education*, pp. 43-68.

Hanna, G., & Janke, N. (1993). Proof and application *Educational Studies in Mathematics* 24(4), 421-438.

Hanna, G., & Janke, N. (1996). Proof and proving. In: Bishop A. *et al.* (Eds.) *International handbook of mathematics education* (pp. 877-908). Dordrecht: Kluwer Academic Publishers.

Harel, G., & Sowder, L. (1998). Students' proof schemes: Results from exploratory studies. In: Schonfeld, A., Kaput, J., & E. Dubinsky E. (Eds.) *Research in collegiate mathematics education III.* (Issues in Mathematics Education, Volume 7, pp. 234-282). American Mathematical Society.

Harel, G. (2002). DNR-Based Instruction: Its Application in Developing Mathematics Teachers' Knowledge Base, Particularly their Proof Schemes. *Proceedings of 2002 International Conference on Mathematics: Understanding Proving and Proving to Understand*, pp. 78-90.

Healey, L., & Hoyles, C. (1998). *Justifying and proving in school mathematics. Summary of the results from a survey of the proof conceptions of students in the UK.* Research Report Mathematical Sciences, Institute of Education, University of London.

Hoyles, C., & Küchemann, D. (2002). Students' explanations in geometry: insights from a large-scale longitudinal survey. *Proceedings of 2002 International Conference on Mathematics: Understanding Proving and Proving to Understand,* pp. 9-22.

Lawson, A. (2003). *The Neurological Basis of Learning, Development and Discovery*. Kluwer: Academic Publishers.

Lin, F. L. (1991). Characteristic of 'Adders' in Proportional Reasoning. *Proceedings of the National Science Council Part D: Mathematics, Science, and Technology Education,* Vol.1 (1), pp.1-13.

Lin, F. L. (2004). Features and Relationship of Reasoning, Proving and Understanding Proof in Number Patterns. *International Journal of Science and Mathematics Education, 2*(2), 227-256.

Tall, D., & Vinner, S. (1981). Concept Image and Concept Definition in Mathematics with particular reference to limit and continuity. *Educational Studies in Mathematics, 12*, 151-169.

Yang, K. L. (2005). Conceptualizing reading comprehension of geometry proof. *Proceedings of 2005 International Conference on Reading and Writing in Science and Mathematics.* p. 61

Yu, Jya-Yi Wu (2001). *Reasoning Difficulty Related to Validity of Argument and Truth of Assertion in Mathematics.* Unpublished Doctoral Dissertation, University of Pittsburgh.

Yu, Jya-Yi Wu, Chin E.-T., & Lin C.-J. (2004). Taiwanese Junior High School Students' Understanding about the Validity of Conditional Statement. *International Journal of Science and Mathematics Education, 2*(2), 257-282.

TRAVELLING THE ROAD TO EXPERTISE: A LONGITUDINAL STUDY OF LEARNING

Kaye Stacey

University of Melbourne, Australia

A longitudinal study of students' developing understanding of decimal notation has been conducted by testing over 3000 students in Grades 4 to 10 up to 7 times. A pencil-and-paper test based on a carefully designed set of decimal comparison items enabled students' responses to be classified into 11 codes and tracked over time. The paper reports on how students' ideas changed across the grades, which ways of thinking were most prevalent, the most persistent and which were most likely to lead to expertise. Interestingly the answers were different for primary and secondary students. Estimates are also given of the proportion of students affected by particular ways of thinking during schooling. The conclusion shows how a careful mapping can be useful and draws out features of the learning environment that affect learning.

In this presentation, we will travel on a metaphorical seven year journey with over 3000 students. As they progress from Grades 4 to 10, learning mathematics in their usual classrooms, we will think of these students as travelling along a road where the destination is to understand the meaning of decimals. The noun "decimal" means a number written in base ten numeration with a visible decimal point or decimal comma. It may be of finite or infinite length. Different students take different routes to this destination, and we will follow these different routes through the territory that is the understanding of decimal numbers and numeration. Of course, the students are simultaneously travelling to many other mathematical and non-mathematical destinations, but our information enables us to follow just one of these journeys. The benefit in following one journey derives from the knowledge that we gain of their paths on this journey, how to help them reach the destination securely and also from being able to generalise this knowledge to understanding their likely paths on their other mathematical journeys.

Our travelling companions: the students

In preparation for our journey, we need to find out about our travelling companions, the transport that is available to them, how we will map their progress, the nature of their destination and the territory through which they travel. Our travelling companions are 3204 Australian students from 12 schools in Melbourne. The schools and teachers volunteered their classes for the study. The youngest students were in Grade 4, the grade when most schools are just beginning to teach about decimals. The oldest students were in Grade 10, two or three years after teachers generally expect their students to have fully developed understanding of decimals. The data is from a cohort study, which tracked individual students for up to 4 years, testing them with the same test each semester (i.e. twice per year). Students entered the study at any grade between Grade 4 and 10, and continued to be tested until they left Grade

2005. In Chick, H. L. & Vincent, J. L. (Eds.). *Proceedings of the 29th Conference of the International Group for the Psychology of Mathematics Education*, Vol. 1, pp. 19-36. Melbourne: PME.

10, or until they left the schools or classes in the study, or until the end of the data collection phase of the study. In total, the 3204 students completed 9862 tests, and when allowing for absences from class on the testing days, the tests were an average of 8.3 months apart. The schools come from a representative range of socio-economic backgrounds, and were chosen in six geographical groups so that many students could be tracked across the primary-secondary divide. Nearly 60% of the 1079 students who were first tested in primary school (i.e. elementary school, Grades 4 to 6) were also tested in secondary school. More than 600 students completed 5, 6 or 7 tests during the study. The detailed quantitative analyses of the test results presented in this paper are taken from the PhD thesis of Vicki Steinle (2004), whose careful and imaginative contribution to our joint work on students' understanding of decimals is acknowledged with gratitude and admiration.

The transport: their teaching

The transport available to the students along this journey is principally the teaching of decimals that was provided at their schools. In the absence of a prescriptive national curriculum or recommended textbooks in these schools, teaching approaches are selected by teachers. This variety makes it difficult to give a comprehensive picture. Instruction will generally begin by introducing one place decimals as an alternative notation for tenths (e.g. 0.4 is 4 tenths, 1.8 is one plus 8 tenths) in Grades 3 or 4. Dienes' multibase arithmetic blocks and area models are the most common manipulatives used. In some programs, calculations are done with one place decimals (e.g. 0.24, 4.79) in the early years, followed by calculations with two place decimals treated exclusively later. In secondary school, textbooks very frequently ask that all decimal calculations are rounded to two decimal places. Brousseau (1997) is among the authors who have commented that teaching which works exclusively with decimals of a fixed length is likely to support overgeneralisation of whole number properties. In the course of our wider work on teaching and learning decimals, our team has designed and trialled a range of teaching interventions, including use of novel manipulatives based on a length model (Stacey, Helme, Archer & Condon, 2001b) and we have created a set of computer games using artificial intelligence techniques (Stacey, Sonenberg, Nicholson, Boneh & Steinle, 2003b), but only a very tiny percentage of students from the cohort study were involved in trialling any of these interventions. The teaching that the students received in the longitudinal study can therefore be assumed to be a representative sample of teaching across Melbourne.

The destination: understanding decimal notation

What is the destination for this journey? Students will have arrived at the final destination when they have a full understanding of the meaning of decimal notation. For the purpose of our wider work on teaching and learning about decimals, full understanding means that they should be able to interpret a number such as 17.373 in terms of place value in several ways (as 17 + 3 tenths + 7 hundredths + 3 thousandths or as 17 + 373 thousandths, etc) and to appreciate that it is less than halfway between

17 and 18, close to 17.4 but with an infinite number of numbers between it and 17.4. At this point, it is worth noting that decimal notation, as a mathematical convention, involves a mix of arbitrary facts that have to be learned and deep mathematical principles. It is not *merely* a convention. Some aspects are completely arbitrary, for example identifying the units column by the contiguous placement of a decimal point (or a decimal comma in many countries) or placing the larger place value columns on the left rather than the right. However, the notation also embodies deep mathematics, such as the uniqueness of the decimal expansion, with the consequence that all decimals of the form 2.37xxxx are larger than all decimals of the form 2.36xxxx except that $2.36\dot{9} = 2.37 = 2.370$ etc. It is this property that makes the decimal comparison task so easy for experts. In the sense of Pea (1987), decimal notation is an invented symbolic artefact bearing distributed intelligence.

Early explorers mapping the territory

The description of the territory through which students pass is strongly linked to the way in which their progress can be mapped. This is a basic feature of science: there is a two-way interaction between knowledge of a phenomenon and having instruments to observe it. In mathematics education, knowledge of students' thinking depends on asking good questions, and we only know what the good questions are by understanding students' thinking. In the context of students' understanding of decimals, Swan commented on this phenomenon in 1983:

> "It is only by asking the right, probing questions that we discover deep misconceptions, and only by knowing which misconceptions are likely do we know which questions are worth asking", (Swan, 1983, p65).

Cumulative research on students' understanding of decimals has broken this cycle to advantage. The task of comparing decimal numbers (e.g. deciding which of two decimals is larger, or ordering a set) has been used since at least 1928 (Brueckner, 1928) to give clues as to how students interpret decimal notation. Refinements to the items used, especially since 1980, improved the diagnostic potential of the task and provided an increasingly good map of the territory of how students interpret decimal notation. For example, Foxman *et al* (1985), reporting on large scale government monitoring of mathematics in Britain, observed a marked difference in the success rates of apparently similar items given to 15 year old students. Asked to identify the largest in the set of decimals {0.625, 0.5, 0.375, 0.25, 0.125}, the success rate was 61%. Asked to identify the smallest, the success rate was a surprisingly much lower 37%. Note that this paper presents all sets from largest to smallest, not in order presented. Further analysis led to the first confirmation in a large scale study that whilst some students consistently interpret long decimals (e.g. 0.625, 0.125) as *larger* numbers than short decimals (e.g. 0.5), which was well known at the time, a significant group interpret them as *smaller* numbers.

> "Despite the large proportions of pupils giving this type of response very few teachers, advisors, and other educationalists are aware of its existence – the monitoring team were

among those unaware of the 'largest is smallest' response at the beginning of the series of surveys." (Foxman *et al*, 1985, p851)

Asking students to identify the smallest from this set of decimals was used again as an item by the international "Trends in Mathematics and Science Study" (TIMSS-R, 1999) Table 1 gives the percentage of the international and Australian students giving each response, alongside Foxman *et al*'s 1985 data. The existence of the same general patterns in the selection of responses across countries and times shows that there is a persistent phenomenon here to be studied. There is also a good fit between the results from the TIMSS-R random Australian sample and a prediction made from the Grade 8 sample of the present longitudinal study (re-calculated from Steinle, 2004, Appendix 4, Table 19), which confirms that the results of the longitudinal study presented in this paper are representative of today's Australian students.

Table 1: Percentage response to the item: *Which of these is the smallest number? {0.625, 0.5, 0.375, 0.25, 0.125}* from TIMSS-R (age 13), APU (age 15) and with prediction from present longitudinal study (Grade 8).

Option	TIMMS-R International	TIMMS-R Australia	Foxman *et al.* APU, age 15	Prediction (Grade 8)
0.125	46%	58%	37%	60%
0.25	4%	4%	3%	2%
0.375	2%	1%	2%	2%
0.5	24%	15%	22%	18%
0.625	24%	22%	34%	17%

Working at a similar time to Foxman *et al*, Sackur-Grisvard and Leonard (1985) demonstrated that examination of the pattern of responses that a student makes to a carefully designed set of comparison or ordering tasks could reveal how the student was interpreting decimal notation reasonably reliably and they documented the prevalence of three "errorful rules" which students commonly use. This provided a rudimentary map of the territory through which students pass on their way to expertise in understanding decimal notation. Sackur-Grisvard and Leonard's test was later simplified by Resnick *et al* (1989) and has been steadily refined by our group to provide an instrument which can map where students are on their journey to expertise. Current researchers, such as Fuglestad (1998), continue to find that decimal comparison tasks provide a useful window into students' thinking and progress.

The territory and the mapping tool

Measuring the progress of a large cohort of students along the journey to understanding decimal notation required a mapping tool that is quick and easy to administer, and yet informative. The version of the instrument used in our longitudinal study is called Decimal Comparison Test 2 (DCT2). It consists of 30 pairs of decimals with one instruction: “circle the larger number in each pair”. The pattern of responses (not the score) on 5 item-types (subsets of items with similar mathematical and psychological properties) enables classification of students into 4 “coarse codes” (A, L, S and U) which are further broken down into 11 “fine codes” (A1, A2, L1, etc) to describe likely ways of thinking about decimals. Figure 1 gives one sample item from each item-type in DCT2 and shows how students in 7 of the fine codes answer these items. Students are classified into the coarse codes on the basis on their answers to the first two item-types (shaded in Figure 1) whereas the fine codes use all item-types. In summary, we map where students are on their journey by administering a test that is simple to do, but has a complex design and a complex marking scheme. Details of the sampling, the test and its method of analysis and many results have been described elsewhere; for example, Steinle and Stacey (2003) and Steinle (2004). We can think of the 11 fine codes as the towns that students might visit on the journey, although, as in most adventure stories, these towns are mostly not good places to be. The 4 course codes are like shires; administrative groupings of towns (fine codes) that have some connections.

Comparison Item		**A1**	**A2**	**L1**	**L2**	**S1**	**S3**	**U2**
4.8	4.63	√	√	×	×	√	√	×
5.736	5.62	√	√	√	√	×	×	×
4.7	4.08	√	√	×	√	√	√	×
4.4502	4.45	√	×	√	√	×	×	×
0.4	0.3	√	√	√	√	√	×	×

Figure 1. Sample items from DCT2 and the responses for the specified codes.

Some of the ways of thinking that lead to these patterns of responses are briefly summarised in Table 2. In the presentation, some of these ways of thinking will be illustrated with case studies from Steinle, Stacey and Chambers (2002). The L behaviour (generally selecting a longer decimal as a larger number) was widely known long before the S behaviour (generally selecting a shorter decimal as a larger number) was documented as reported above. Neither coarse code A nor U students choose on length. Students coded A are correct on straightforward comparisons, and U is a mixed group making other responses. The ways of thinking that lie behind these behaviours (other than U) have been identified by interviews with students, supported by close analysis of response patterns to identify the characteristics of apparently similar items to which groups of students react differently. Behind the codes, there are often several different ways of thinking that result in the same patterns of responses to the DCT2. Later refinements of the test enable some of these different ways of thinking to be separated. Space forbids a full description here.

Table 2: Matching of codes to the ways of thinking

Coarse Code	Fine Code	Brief Description of Ways of Thinking
A apparent expert	A1	Expert, correct on all items, with or without understanding.
	A2	Correct on items with different initial decimal places. Unsure about 4.4502 /4.45. May only draw analogy with money. May have little understanding of place value, following partial rules.
L longer-is-larger	L1	Interprets decimal part of number as whole number of parts of unspecified size, so that 4.63>4.8 (63 parts is more than 8 parts).
	L2	As L1, but knows the 0 in 4.08 makes decimal part small so that 4.7>4.08. More sophisticated L2 students interpret 0.81 as 81 tenths and 0.081 as 81 hundredths etc resulting in same responses.
S shorter-is-larger	S1	Assumes any number of hundredths larger than any number of thousandths so 5.736 < 5.62 etc. Some place value understanding.
	S3	Interprets decimal part as whole number and draws analogy with reciprocals or negative numbers so 0.3>0.4 like 1/3>1/4 or -3>-4.
U	U2	Can "correctly" order decimals, but reverses answers so that all are incorrect (e.g. may believe decimals less than zero)
	U1	Unclassified – not fitting elsewhere. Mixed or unknown ideas.

How adequate is DCT2 as an instrument to map where students are on their journeys to full understanding? Clearly it has limitations, but it also has many strengths. Its ease of administration made the longitudinal study of a large number of students possible. The test can reliably identify a wide range of student responses, as illustrated in Table 2. Test-retest agreement is high. Even after one semester, when one would expect considerable learning to have occurred, 56% of students re-tested in the same fine code (calculation from data in Steinle 2004, Table 5.17). Where we have interviewed students shortly after testing, they generally exhibit the diagnosed way of thinking in a range of other items probing decimal understanding. There is one important exception. Very frequently, students whom the test diagnoses as expert (A1) are (i) not experts on other decimal tasks and (ii) it is also sometimes the case that they can correctly complete comparison items but do *not* have a strong understanding of decimal notation. For this reason our code for expertise is A1, with A standing for *apparent task expert*. In relation to point (i), our intensive use of one task has highlighted for us that expertise in one task does not necessarily transfer to related tasks without specific teaching. For example, A1 students being expert in the comparison test would be able to order books in a library using the Dewey decimal system. However, they may have little idea of the metric properties of decimals: that 0. 12345 is very much closer to 0.12 than it is to 0.13, for example, and they may not be able to put numbers on a number line. We therefore make no claim that our *apparent task experts* in A1 are expert on other decimal tasks. In relation to point (ii), students with either good or poor understanding can complete DCT2 correctly by following either of the two expert rules (left-to-right digit comparison or adding zeros and comparing as whole numbers e.g. compare 63 and 80 to compare 4.63 and 4.8). DCT2 therefore *over-estimates* the number of experts. As a tool to map students' progress it overestimates the numbers who have arrived at the destination. Its strength is in identifying the nature of erroneous thinking. Some mathematics educators may be inclined to dismiss DCT2 as "just a pencil-and-paper test" and take the position that only an interview can give reliable or deep information about student thinking. I contend that carefully designed instruments in any format with well studied properties, are important for advancing research and improving teaching. Many interviews also miss important features of students' thinking and unwittingly infer mastery of one task from mastery of another.

THE JOURNEYS

Some sample journeys

Table 3 shows the journeys of 9 students in the longitudinal study. It shows that Student 210403026 completed tests each semester from the second semester of Grade 4 to the first semester of Grade 7, and was absent on one testing day in Grade 5. Student 300704112 always tested in the L coarse code, which is an extreme pattern that sadly does not reveal any learning about this topic in two and a half years of school attendance. Student 310401041 completed 7 tests, being diagnosed as either unclassified or in the L coarse code. Student 410401088, however, moved from L behaviour to expertise in Grade 7. Some of the students in Table 3 have been chosen to illustrate how many students persist with similar ways of thinking over several years. The average student showed more variation than these. In addition, there is always the possibility that changes between tests have been missed, since students were tested at most twice per year. Some students show movement in and out of A1.

Table 3: A sample of students' paths through the study

ID	Grade 4		Grade 5		Grade 6		Grade 7		Grade 8		Grade 9		Grade 10	
210403026		L1		A1	S3	S5	S1							
300704112							L1	L4	L4	L2	L1			
310401041		L2		L1	U1	U1	L4	U1	U1					
390704012							L1	A1	U1	A1	S3			
400704005							A1	A2	A1	A2	A1			
410401088		L1	L1		L4	L1	L2	A1	A1					
500703003								S1	S5		S3	S3	U1	
500703030								S3	S5		S1	A2		
600703029								A1		U1	A1	A1	A3	

Prevalence by grade: where the students are in each year of the journey

Figures 2, 3a and 3b show the percentage of students who are in each of the codes by grade level. This data is the best estimate available from the longitudinal study (technically, the improved test-focussed prevalence of Steinle (2004)). As expected, the percentage of experts on the test (A1 in Figure 2) grows markedly in the early years, rising steadily until Grade 8. However, at Grade 10, which is regarded as the end of basic education, it is still only at 70% indicating that there are likely to be many adults without a strong understanding of decimal numbers. This observation is reinforced by studies of teacher education students (Stacey *et al*, 2001c) and nurses where "death by decimal" (Lesar, 2002) is a recognised phenomenon. Measuring expertise with the DCT2 over-estimates, we summarise by noting that one quarter of students attain expertise within a year or so of first being introduced to decimals (i.e. in grade 5), a further half of students attain expertise over the next 5 years, leaving a quarter of the school population who are not clear on these ideas by the end of Grade 10.

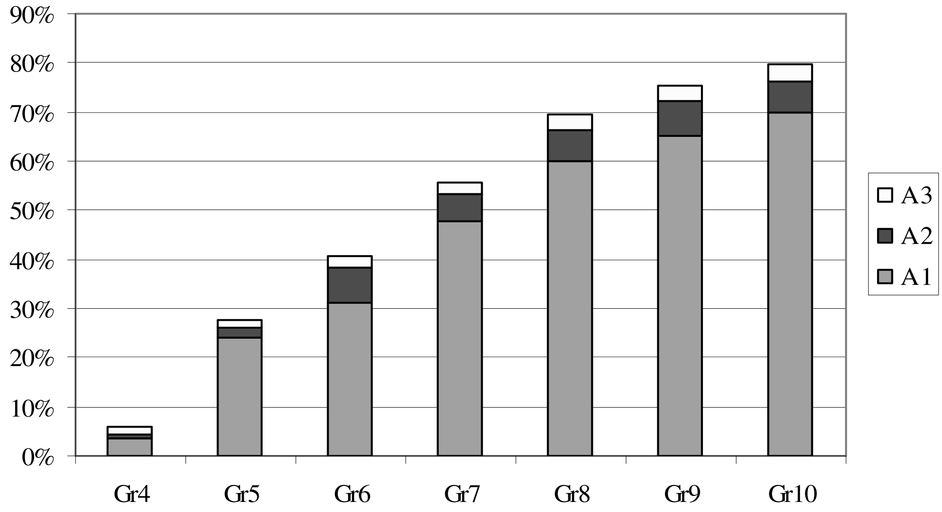


Figure 2: Best estimate of the prevalence of A codes by grade
(from Figure 9.3, Steinle, 2004)

Figure 2 also shows that the percentage of students in the non-expert A group remains (i.e. A2/A3) at about 10% from Grade 6 throughout secondary school, and for reasons related to the test construction, we know this to be an under-estimate. These students operate well on the basic items, but make errors on what could be expected to be the easiest comparisons, such as 4.45 and 4.4502. We believe there are several causes: an over-reliance on money as a model for decimal numbers; over-institutionalisation of the practice of rounding off calculations to two decimal places; and use of partially remembered, poorly understood rules for comparing decimals. A2 and A3 students function well in most circumstances, but may in reality have very little understanding. We have several times overheard teachers describing their A2 students as having “just a few more little things to learn”. In fact these students may have almost no understanding of place value.

Figure 3a shows how that the prevalence of L codes drops steadily with grade. As might be expected, the naïve misconception that the digits after the decimal point function like another whole number (so that 4.63 is like 4 and 63 units of unspecified size and 4.8 is 4 plus 8 units of unspecified size), is an initial assumption about decimal numbers, and Foxman *et al* (1985) demonstrated that it is exhibited mainly by low achieving students. The fairly constant percentage of students in category L2 (around 4% up to Grade 9) provides an example of how students’ knowledge sometimes grows by just adding new facts to their accumulated knowledge, rather than building a consistent understanding based on fundamental principles. One cause of code L2 is that L1 students simply add an extra piece of information to their pre-existing way of thinking – commonly in this case, the information that a decimal number with a zero in the tenths column is small so that 4.08 < 4.7 even though 8>7.

Figure 3b shows the best estimate of prevalence of the S codes. These codes are less common, but there is no consistent trend for them to decrease: instead about 15% of students in most grades exhibit S behaviour at any one time. The largest group is in code S3, which is again a naïve way of thinking not appreciating place value. That over 10% of Grade 8 students (those in S3) will consistently select 0.3 as smaller than 0.4 is an extraordinary result. Earlier studies had omitted these items from tests, presumably because they were thought to be too easy. We believe that S thinking grows in junior secondary school largely because of interference at a deep psycholinguistic or metaphorical level from new learning about negative numbers, negative powers (e.g. 10^(-6) is a very small number) and more intense treatment of fractions, and a strange conflation of the spatial spread of place value columns with number-lines. These ideas are explained by Stacey, Helme & Steinle (2001a).

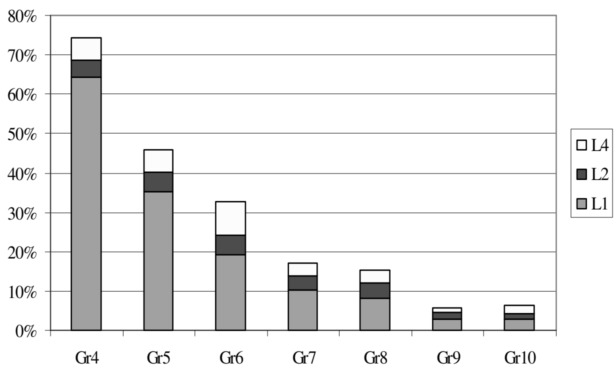


Figure 3a: Prevalence of L codes by grade (from Figure 9.7, Steinle, 2004)

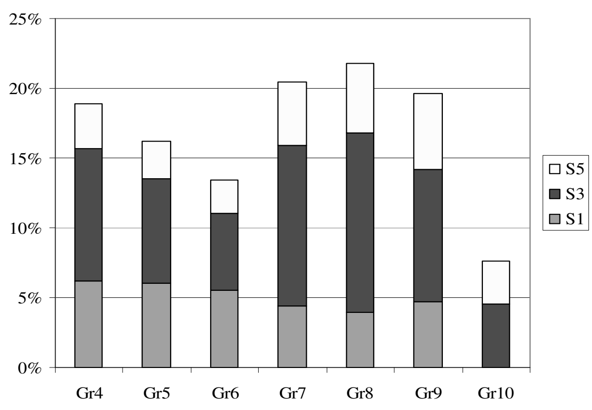


Figure 3b: Prevalence of S codes by grade (from Figure 9.10, Steinle, 2004)

Student-focussed prevalence: how many students visit each town?

The data above have shown the percentage of students testing in various codes – in the journey metaphor, a snapshot of where the individuals are at a particular moment in time. This is one way to answer to the question "how prevalent are these ways of thinking". However, it is also useful to see how many students are affected by these ways of thinking over their schooling, which is analogous to asking how many students visited each town sometime on their journey. Figure 4 shows the percentage of students who tested in each coarse code at some time in primary school, or at some time in secondary school. These percentages add up to more than 100% because students test in several codes. This data in Figure 4 is based on the 333 students in primary school and 682 students in secondary school who had completed at least four tests at that level of schooling. Had any individual been tested more often, he or she may have also tested in other codes. Hence it is evident that the data in Figure 4 are all under-estimates.

This new analysis gives a different picture of the importance of these codes to teaching. For example, less than 25% of students exhibited S behaviour at any one test, but 35% of students were affected during primary school. Similar results are evident for the fine codes, although not presented here. For example, Fig. 3b shows that about 6% of students were in S1 at any one time, but at least 17% of primary and 10% of secondary students were in S1 at some time. As noted above, these are underestimates.

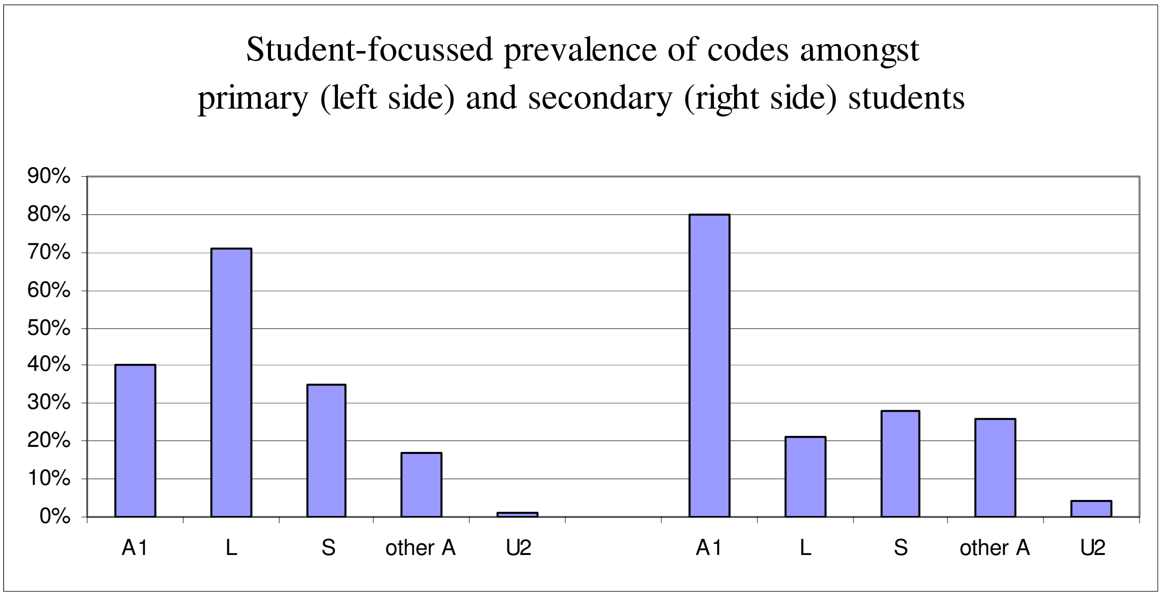


Figure 4: The percentage of students who test in given codes at some stage in primary and secondary school (derived from Steinle, 2004, Ch 9).

Persistence: which towns are hard to leave?

The sections above show where students are at various stages on their journeys. In this section we report on how long they stay at each of the towns on their journey. These towns are not good places to be, but how attractive are they to students? Figure 5a shows that around 40% of students in the L and S codes retested in the same code at the next test (tests averaged 8.3 months apart). The figure also shows that after 4 tests (averaging over two and a half years) still about 1 in 6 students retest in the same code. It is clear from this data that for many students, school instruction has insufficient impact to alter incorrect ideas about decimals.

Fortunately, expertise is even more persistent than misconceptions. On a test following an A1 code, 90% of A1 students rested as A1 and the best estimate from Steinle (2004) is that 80% of A1 students always retest as A1. This means that about 20% of the DCT2 "experts" achieve this status by less than lasting understanding (e.g. by using a rule correctly on one occasion, then forgetting it).

Figure 5b shows an interesting phenomenon. Whereas persistence in the L codes decreases with age (Figure 5b shows L1 as an example), persistence in the S and A2 codes is higher amongst older students. This might be because the instruction that students receive is more successful in changing the naive L ideas than S ideas but it is also likely to be because new learning and classroom practices in secondary school incline students towards keeping S and A2 ideas. The full data analysis shows that this effect occurred in nearly all schools, so it does not depend on specific teaching.

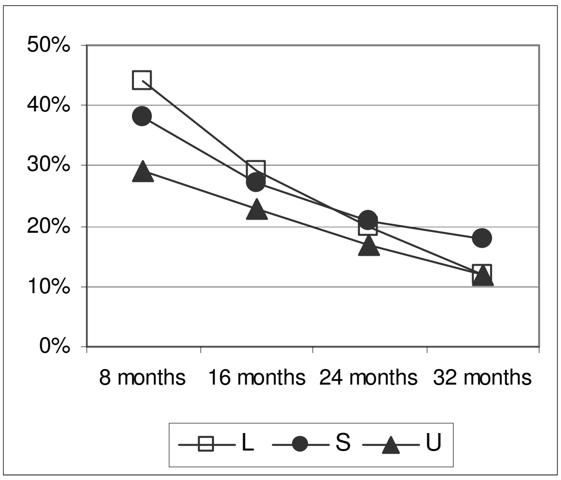


Figure 5a: Persistence in L, S and U codes after 1, 2, 3 or 4 semesters (adapted from Steinle, 2004, Fig. 6.5)

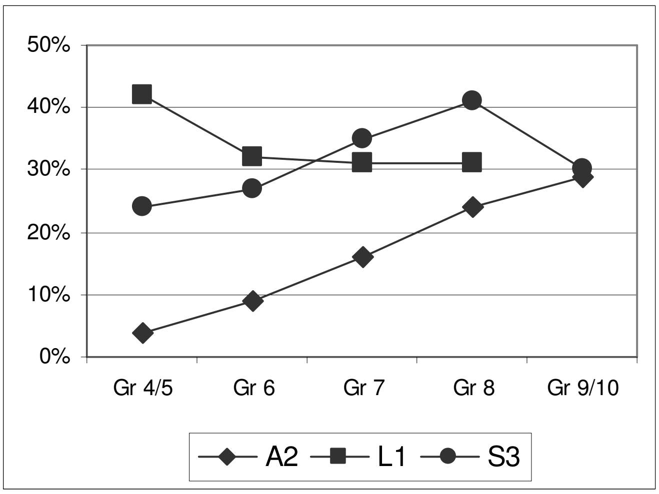


Figure 5b: Persistence in A2, L1, S3 and S5 over one semester by grade of current test (adapted from Steinle, 2004, Fig. 6.1)

Proximity to expertise: which town is the best place to be?

A final question in describing students' journeys is to find which town is the best place to be. In other words, from which non-A1 code is it most likely that a student will become an expert on the next test? Figure 6 shows the best estimates of Steinle (2004) from the longitudinal data. For both primary and secondary students the A codes and the U codes have the highest probabilities. The case of the A codes will be discussed below. The vast majority of students in U ("unclassified") do not respond to DCT2 with a known misconception: they may be trying out several ways of thinking about decimals within one test, or simply be guessing. Figure 5a shows that the U coarse code is the least persistent, and the data in Figure 6 shows that there is a relatively high chance that U students will be expert on the next test. It appears that it is worse to have a definite misconception about decimals than to be inconsistent, using a mix of ideas or guessing. Perhaps these students are more aware that there is something for them to learn and are looking for new ideas.

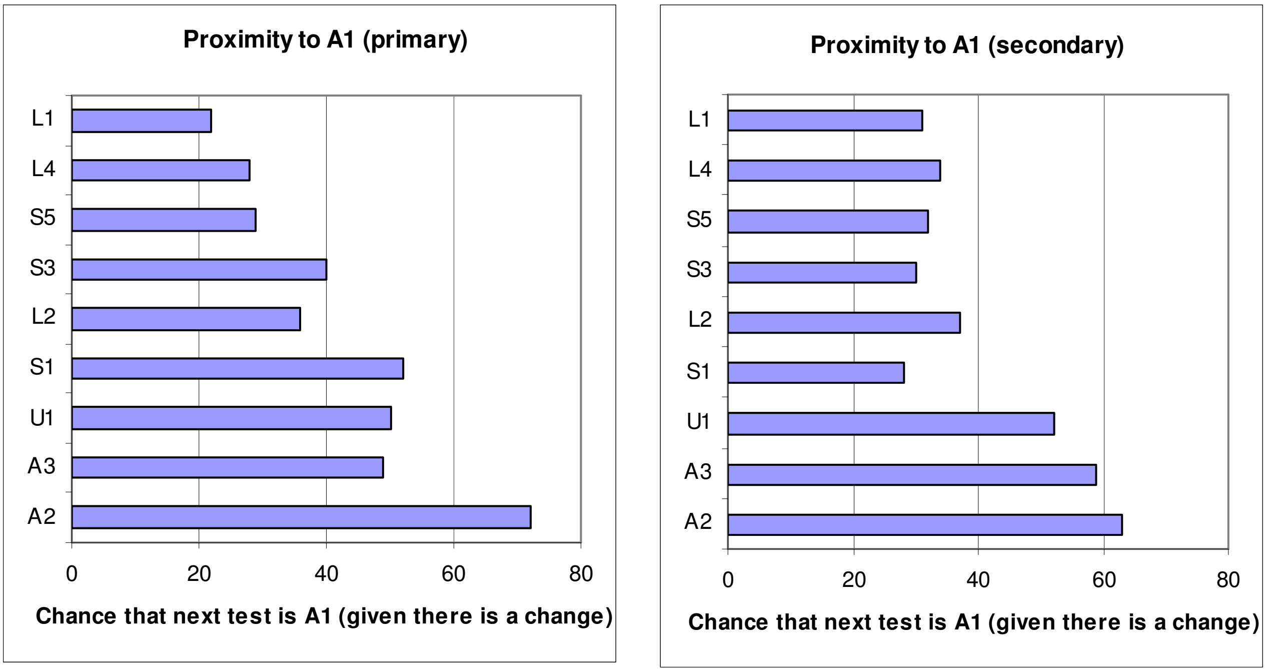


Figure 6: Chance that the next test is A1, given there is a change of code, for primary and secondary cohorts. (Codes ordered according to combined cohort proximity.)

Students in the L codes generally have only a low chance of moving to expertise by the next test. This bears out predictions which would be made on our understanding of the thinking behind the L codes. Since L1 identifies students who generally think of the decimal part of the number as another whole number of parts of indeterminate size, L1 is rightly predicted to be far from expertise. The L2 code (see Table 2) consists of at least two groups: one who graft onto L1 thinking an isolated fact about numbers with a zero in the tenths columns and a more sophisticated group of students with some place value ideas. Is the much greater chance of L2 students becoming expert over L1 students attributable to both or to the more sophisticated thinkers only? This is an example of a question that needs a more refined test than DCT2.

In the above section on persistence, I commented that the S codes behave differently in primary and secondary schools. This is again the case in Figure 6. Whereas primary students in S codes have a better chance than L students to become experts, this is not the case in secondary school. This is *not* because S students are more likely to stay in S, because the analysis has been done by removing from the data set those students who do not change code. Exactly what it is in the secondary school curriculum or learning environment that makes S students who change code more likely to adopt ideas which are not correct, is an open question.

The A codes have very high rates of progression to A1. This is of course good, but there is a caution. As noted above, students who have tested as A1 on one test generally stay as A1 on the next test, but 10% do not (see for example, students 400704005 and 600703029 from Table 3). The A2 and A3 codes are over-represented in these subsequent tests. This indicates to us that some of the A1 students are doing well by following partly understood and remembered versions of either of the two expert rules, possibly so partial as to simply make a decision on the first one or two decimal places (e.g. by analogy with money), truncated or rounded. In a "tricky" case such as the comparison 4.4502/4.45, these partially remembered rules fail. Truncating or rounding to one or two decimal digits gives equal numbers and to carry out the left-to-right digit comparison rule, the 0 digit has to be compared with a blank. Poorly understood and remembered algorithms are likely to fail at this point, resulting in *ad hoc* guessing. As students complete subsequent tests in A1, A2 and A3, moving between them, we see examples of Brown and VanLehn's (1982) "bug migration" phenomenon. There is a gap in students' understanding or in their memorised procedures, and different decisions about how to fill this gap are made on different occasions. Our work with older students (e.g. Stacey *et al*, 2001c) shows that these problems, evident in comparisons such as 4.45/4.4502, remain prevalent beyond Grade 10. The movement between the A codes is evidence that a significant group of the DCT2 "experts" have little place value understanding.

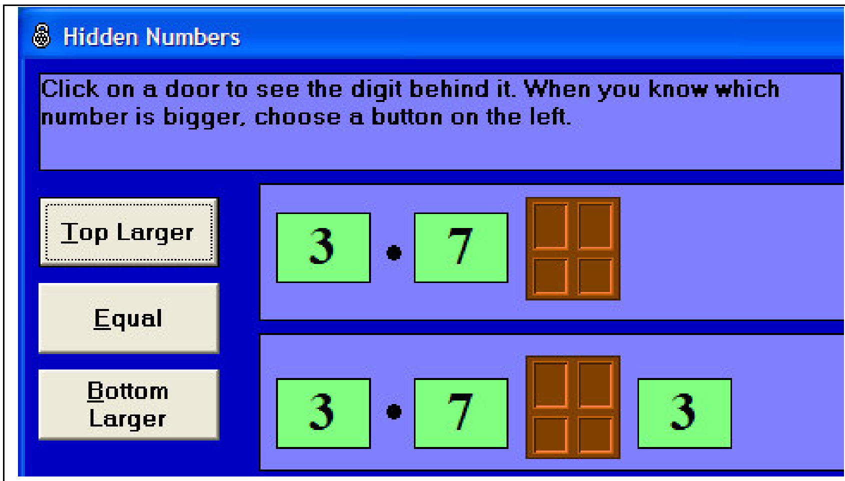
The study of student's thinking especially in the A and S codes has highlighted difficulties associated with zero, both as a *number* and as a *digit*, that need attention throughout schooling (Steinle & Stacey, 2001). Zeros can be visible or invisible and represent the number between positive and negative numbers, or a digit. As a digit, zero operates in three ways numbers; to indicate there are zero components of a given place value, as a place holder to show the value of surrounding digits, and also to indicate the accuracy of measurement (e.g. 12 cm vs 12.0 cm) although the latter interpretation has not been explored in our study. Improved versions of the decimal comparison test, especially for older students, include more items involving zeros in all of these roles, and allow the comparisons to be equal (e.g. 0.8 with 0.80).

HOW IS A DETAILED MAP OF LEARNING USEFUL?

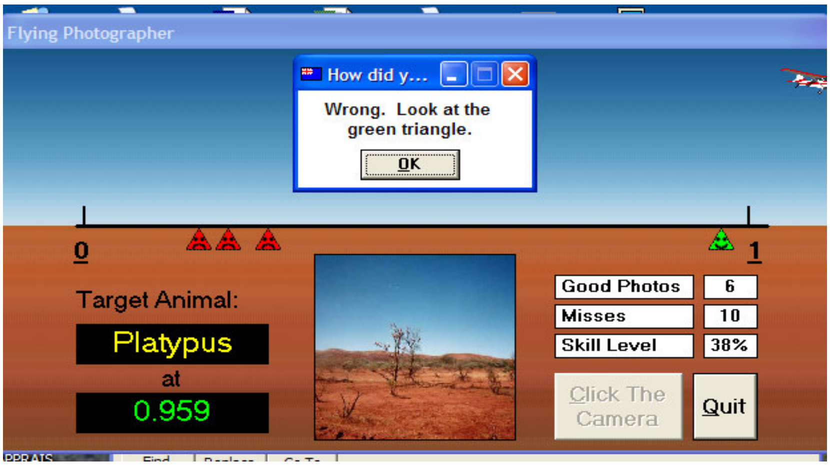
The research work in the 1980s using comparison of decimals identified three "errorful rules". The map of the territory of learning decimals at that stage therefore divided it into four regions (expertise and three others). DCT2 can diagnose students into 12 groups (the 11 of the longitudinal study and one other). As we interviewed students who tested in different codes on DCT2 and examined responses to the sets of items more closely, we came to realise that several ways of thinking lay behind some of our codes (e.g. L2, S3), which opened up the possibility of making further refinements to DCT2 to separate these groups of students. We also discovered other ways of thinking that DCT2 did not properly identify, such as problems with 0. We refined DCT2 to better identify some of these groups. However, the important question which is relevant to all work on children's thinking is how far it is useful to take these refinements. How fine a mapping tool will help students on the journey?

For teaching, it is common for people to say that only the coarsest of diagnoses is useful. The argument is that busy teachers do not have the time to carefully diagnose esoteric misconceptions, and in any case would be unable to provide instruction which responded to the information gained about an individual student's thinking. I agree. Our experience in teachers' professional development indicates that they find some knowledge of the misconceptions that their students might have to be extremely helpful to understand their students, and to plan their instruction to address or avoid misinterpretations. Hence they find that the coarse grained diagnosis available for example from the Quick Test and Zero Test (Steinle et al, 2002) is of practical use.

However, in many countries, we will soon be going beyond the time when real-time classroom diagnosis of students' understanding is the only practical method. The detailed knowledge of student thinking that has been built up from research can be built into an expert system, so that detailed diagnosis can be the province of a computer rather than a teacher. Figure 7 shows two screen shots from computer games which input student responses to a Bayesian net that diagnoses students in real time and identifies the items from which they are most likely to learn. Preliminary trials have been promising (Stacey & Flynn, 2003a). Whereas all students with misconceptions about decimal notation need to learn the fundamentals of decimal place value, instruction can be improved if students experience these fundamental principles through examples that are individually tailored to highlight what they need to learn. Many misconceptions persist because students get a reasonable number of questions correct and attribute wrong answers to "careless errors". This means that the examples through which they are taught need to be targeted to the students' thinking. An expert system can do this (Stacey et al, 2003b).



In *Hidden Numbers*, students pick the relative size of two numbers, revealing digits by opening the doors. This task reveals misconceptions e.g. when students select by length or open doors from the right. An expert system diagnoses thinking and provides tasks for teaching or diagnosis.



The *Flying Photographer* has to photograph animals (e.g. platypus) from an aeroplane, given decimal co-ordinates (e.g. 0.959). This task uses knowledge of relative size, not just order. An expert system tracks responses (e.g. if long decimals are always placed near 1) and selects new items to highlight concepts.

Figure 7. Screen shots from two games which provide diagnostic information to an expert system which can diagnose students and select appropriate tasks.

LESSONS ABOUT LEARNING

An overview of the journey

The longitudinal study has examined students' progress in a specific mathematics topic, which complements other studies that have tracked growth in mathematics as a whole or across a curriculum area. The overall results demonstrate the substantial variation in ages at which expertise is attained, from a quarter of students in Grade 5 to about three quarters in Year 10. The good alignment of data from the longitudinal study and the random sample of TIMSS-R shows that we can confidently recommend that this topic needs attention throughout the grades in most secondary schools. The fact that about 10% of students in every grade of secondary school (fig. 2) are in the non-expert A codes (A2 and A3) shows that many students can deal apparently expertly with "ordinary" decimals, which conceals from their teachers and probably from themselves, their lack of understanding of fundamental decimal principles.

Moreover, the fact that many students retain the same misconception over long periods of time (e.g. about 20% in the coarse codes over 2 years, and around 30% in some fine codes over 6 months) demonstrates that much school instruction does not make an impact on the thinking of many students. Our study of proximity to expertise provides empirical support for the notion that it is harder to shake the ideas of students who have a specific misconception than of those who do not; again this points to the need for instruction that helps students realise that there is something for them to learn, in a topic which they may feel they have dealt with over several years.

One important innovation of this study is to look not just at the prevalence of a way of thinking at one time, but to provide estimates of how many students are affected in their schooling, which provides a different view of the practical importance of phenomena.

How the learning environment affects the paths students take

Another important result of this study is that in the different learning environments of primary and secondary school, students are affected differently by various misconceptions. For example, the S misconceptions in primary school are relatively quickly overcome, being not very persistent and with high probability of preceding testing as an expert, but this is not the case in secondary school.

The very careful study of the responses to DCT2 and later comparison tests has revealed a wide range of students' thinking about decimals. As demonstrated in earlier studies, some students (e.g. L1) make naïve interpretations, overgeneralising whole number or fraction knowledge. Others simply add to a naïve interpretation some additional information (e.g. some L2, and see below). We have proposed that some false associations, such as linking numbers with whole number part of 0 with negative numbers, arise from deep psychological processes (Stacey et al, 2001a). Other students (e.g. some A2) seem to rely only on partially remembered rules, without any definite conceptual framework. We explain the rise in the prevalence and persistence of S and non-expert A codes in the secondary school mainly through reinforcement from new classroom practices, such as rounding to two decimal places and interference from new learning (e.g. work with negative numbers). This shows that other topics in the mathematics curriculum, and probably also other subjects, affect the ideas that students develop and the paths that they take among them.

Learning principles or collecting facts

Although understanding decimal notation may appear a very limited task, just a tiny aspect of a small part of mathematics, full understanding requires mastery of a complex web of relationships between basic ideas. From the perspective of the mathematician, there are a few fundamental principles and many facts are logically derived from them. From the point of view of many learners, however, there are a large number of facts to be learned with only weak links between them. This is demonstrated by the significant size of codes such as A2 (e.g. with secondary students confident only with tenths, without having made the generalisation of successive decimation). Teaching weakly linked facts rather than principles is inherent in some popular approaches, such as teaching one-place decimals first, then two-place decimals the next year, without exposing what we call the "endless base ten chain". Artificially high success in class comes by avoiding tasks which require understanding the generalisation and principles, and concentrating on tasks with predictable surface features (e.g. Brousseau, 1997; Sackur-Grisvard *et al*, 1985).

For mathematics educators, the challenge of mapping how students think about mathematical topics is made considerably harder by the high prevalence of the collected facts approach. As the case of decimal numeration illustrates, we have tended to base studies of students' thinking around interpretations of principles, but we must also check whether that current theories apply to students and teachers who are oriented to the collected facts view, and to investigating how best to help this significant part of the school population.

Tracing the journeys of students from Grade 4 to Grade 10 has revealed many new features of how students' understanding of decimals develops, sometimes progressing quickly and well, but for many students and occasionally for long periods of time, not moving in productive directions at all. The many side-trips that students make on this journey point to the complexity of the learning task, but also to the need for improved learning experiences to assist them to make the journey to expertise more directly.

Acknowledgements

The decimals work was funded through a series of projects by Australian Research Council and University of Melbourne, with contributions from V. Steinle, E. Sonenberg, A. Nicholson, S. Helme, and Melbourne and Monash research and honours students. Teachers and children generously gave their time to do the tests.

REFERENCES

Brousseau, G. (1997). Problems with Teaching Decimal Numbers, Didactical Problems with Decimals. In N. Balacheff, M. Cooper, R. Sutherland, & V. Warfield (Eds.), *Theory of Didactical Situations in Mathematics: Didactique des Mathematiques, 1970-1990* (pp. 117-232). Dordrecht, The Netherlands: Kluwer Academic.

Brown, J. S., & VanLehn, K. (1982). Towards a Generative Theory of "Bugs". In T. P. Carpenter, J.M. Moser, & T. A. Romberg (Eds.), *Addition and subtraction: A cognitive perspective* (pp. 117-135). Hillsdale, New Jersey: Lawrence Erlbaum.

Brueckner, L. J. (1928). Analysis of Difficulties in Decimals. *Elementary School Journal, 29*, 32-41.

Foxman, D., Ruddock, G., Joffe, L., Mason, K., Mitchell, P & Sexton, B. (1985). *A Review of Monitoring in Mathematics 1978 to 1982. (Vol. 1).* London: Dept of Ed. & Science.

Fuglestad, A. B. (1998). Computer Support for Diagnostic Teaching. A case of decimal numbers. *Nordic Studies in Mathematics Education (Nordisk Matematikk Didaktikk), 6*(3-4), 25-50.

Lesar, T. (2002). Tenfold medication dose prescribing errors. *Annals of Pharmacotherapy, 36*, 1833-1839.

Pea, R. (1987) Practices of distributed intelligence and designs for education. In G. Salomon (Ed) *Distributed cognitions: Psychological and educational considerations.* (pp 47 - 87) C.U.P., Cambridge.

Resnick, L. B., Nesher, P., Leonard, F., Magone, M., Omanson, S., & Peled, I. (1989). Conceptual Bases of Arithmetic Errors: The Case of Decimal Fractions. *Journal for Research in Mathematics Education, 20*(1), 8-27.

Sackur-Grisvard, C., & Leonard, F. (1985). Intermediate Cognitive Organizations in the Process of Learning a Mathematical Concept: The Order of Positive Decimal Numbers. *Cognition and Instruction, 2*(2), 157-174.

Stacey, K. & Flynn, J. (2003a) Evaluating an adaptive computer system for teaching about decimals: Two case studies. In V. Aleven, U. Hoppe, J. Kay, R. Mizoguchi, H.Pain, F. Verdejo, & K. Yacef (Eds) *AI-ED2003 Supplementary Proceedings of the 11th International Conference on Artificial Intelligence in Education*.(pp 454 – 460), Sydney: University of Sydney.

Stacey, K., Helme, S., & Steinle, V. (2001a). Confusions between decimals, fractions and negative numbers: A consequence of the mirror as a conceptual metaphor in three different ways. In M. v. d. Heuvel-Panhuizen (Ed.), *Proceedings of the 25th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 217-224). Utrecht: PME.

Stacey, K., Helme, S., Archer, S., & Condon, C. (2001b). The effect of epistemic fidelity and accessibility on teaching with physical materials: A comparison of two models for teaching decimal numeration. *Educational Studies in Mathematics, 47*, 199-221.

Stacey, K., Helme, S., Steinle, V., Baturo, A., Irwin, K., & Bana, J. (2001c). Preservice Teachers' Knowledge of Difficulties in Decimal Numeration. *Journal of Mathematics Teacher Education, 4*(3), 205-225.

Stacey, K., Sonenberg, E., Nicholson, A., Boneh, T., Steinle, V. (2003b) A teacher model exploiting cognitive conflict driven by a Bayesian network. In *Lecture Notes in Artificial Intelligence (Series: Lecture Notes in Computer Science Vol 2702/2003)*. (pp 352-362) Heidelberg: Springer-Verlag.

Steinle, V. & Stacey, K. (2003). Grade-related trends in the prevalence and persistence of decimal misconceptions. In N.A. Pateman, B.J. Dougherty & J. Zilliox (Eds.), *Proceedings of the 27th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 259 – 266). Honolulu: PME.

Steinle, V. (2004). *Changes with Age in Students' Misconceptions of Decimal Numbers*. Unpublished PhD, University of Melbourne, Melbourne.

Steinle, V., & Stacey, K. (2001). Visible and invisible zeros: Sources of confusion in decimal notation. In J. Bobis, B. Perry & M. Mitchelmore (Eds.), *Numeracy and Beyond. Proceedings of the 24th Annual Conference of the Mathematics Education Research Group of Australasia* (Vol. 2, pp. 434-441). Sydney: MERGA.

Steinle, V., Stacey, K., & Chambers, D. (2002). *Teaching and Learning about Decimals* [CD-ROM]: Department of Science and Mathematics Education, The University of Melbourne. Online sample http://extranet.edfac.unimelb.edu.au/DSME/decimals/

Swan, M. (1983). *Teaching Decimal Place Value: A Comparative Study of "Conflict" and "Positive Only" Approaches*. Nottingham: Shell Centre for Mathematical Education.

TIMSS-R. *IEA's Repeat of the Third International Mathematics and Science Study at the Eighth Grade*. Retrieved, 29th February, 2004, from the World Wide Web: http://timss.bc.edu/timss1999i/pdf/t99math_items.pdf

IDENTITY THAT MAKES A DIFFERENCE: SUBSTANTIAL LEARNING AS CLOSING THE GAP BETWEEN ACTUAL AND DESIGNATED IDENTITIES

Anna Sfard

The University of Haifa & Michigan State University

Anna Prusak

Oranim Teachers College

In the attempt to account for striking differences between learning activities of immigrant mathematics students from the former Soviet Union and of their native Israeli classmates, we introduce the notions of actual *and* designated identities. *These identities are subsequently presented as important factors that mold learning and influence its effectiveness. Since designated identities may be seen as personalized, "customized" versions of people's cultural heritages, ours is the story of the wider culture making its way into individual learning processes.*

> [For me,] school mathematics was … something that one cannot escape and must try to be done with as quickly as possible… The numbers did not scare me; rather the scary part was my complete lack of interest in them… All that I remember now is my constant effort to match formulas with exam questions.

This quote from a retrospective account of a successful university student[1] is unlikely to surprise a person who knows a thing or two about mathematics learning and teaching. We are all only too familiar with this kind of unhappy reminiscences. Much less common are reports about mathematics-related experiences of interest and joy, such as the one provided by another high-school graduate:

> Mathematics lessons were my favorites. If they were difficult, I saw them as a challenge, as a puzzle to cope with. I was ready to invest time and effort in solving special bonus problems.

What is it that makes some students learn mathematics willingly and with interest while leaving many of their peers indifferent, if not openly resistant? How does this difference influence the learning practices of the student? These questions are certainly not new. They have been fueling mathematics education research ever since its inception. The study to be presented in this talk is a result of yet another attempt to come to grips with the long-standing quandaries.

[1] This and the following excerpt are taken from autobiographical accounts of students who participated in university courses given by the first author in the Education Department at the University of Haifa.

Our research project was occasioned by the recent massive immigration from the former Soviet Union to Israel.[2] More specifically, it was triggered by a spontaneous, yet-to-be-tested observation that a disproportionately large portion of this particular group of immigrants could pride itself with impressive results in mathematics, and not just in school, but also in national and international mathematical competitions.[3] We began asking ourselves whether there was anything unique about the immigrant students' mathematics learning and if there was, how this uniqueness could be accounted for. The conjecture we wished to test while launching our investigation was that dissimilarities in learning processes, rather than being a simple outcome of cognitive differences between individual learners, are a mixed product of individual and collective doing. Such differences, we believed, are often reflective of differing sociocultural histories of the learners.

In what follows, we try to substantiate this hypothesis on the basis of our findings. We begin with detailed examples of the two types of learning, the *ritualized* and the *substantial*, signaled by the students' testimonies quoted above. In our study, both kinds of learning have been found in one class consisting of native Israelis and immigrant mathematics students. The dissimilarities in learning paralleled the difference in the students' sociocultural background. In the attempt to understand how sociocultural factors made their way into the learners' individual activities, we introduce the notions of *actual* and *designated identities* which then serve as the "missing link" between culture and learning.

TWO TYPES OF LEARNING: SUBSTANTIAL AND RITUALIZED

Example to think with: NewComers and OldTimers as mathematics learners

The study began in fall of 1998 and focused on one 11th grade class that followed an advanced mathematics program. 9 out of the 19 students were *NewComers* – recent immigrants from big cities in the former Soviet Union such as Moscow, Kiev and Tbilisi. The rest were native Israelis, whom we call *OldTimers*. All of the students came from well-educated families. The second author, a one-time immigrant from the Soviet Union, served as the teacher. In the course of the entire school year all classroom processes were meticulously observed and documented. Numerous interviews with the students, with their parents and with other teachers constituted additional data.

[2] According to the leading Israeli newspaper *Haaretz*, "Approximately 200 thousand children immigrated to Israel in 11 years, most of them from the former Soviet Union; they constitute 15% of the Israeli youth"(31.08.2001).

[3] This conjecture should not be misread as saying that the immigrants from the former Soviet Union are generally highly successful in mathematics. This said, "[t]here are [immigrant] children who arrive at the highest places in international competitions in mathematics and physics and thanks to them, Israel climbed from 24th to 13th place in the 1995 international championship" (*Haaretz,* 2 August 1996).

The salience of the differences between the learning processes of the two groups exceeded our expectations. In this article we present only a tiny vignette from this extensive research project (the full report can be found in Prusak 2003). It must be stressed, however, that the striking intra-group homogeneity and the significant inter-group difference reported on these pages is representative of all our results, whatever the particular aspect of learning considered in the analyses.

The sub-study in question focused on independent learning. Our story begins in the tenth week of the school year, on the day when the class got the unusual homework assignment: After having learned trigonometry for two months and, in particular, after being introduced to the theorem known as *law of sines*, the learners were asked to study the new subject, *law of cosines and its applications*, with the help of a textbook. To guide their independent learning, the teacher proposed a work plan, which was presented as a series of questions to be answered in the course of the study*: (1) How can the law of cosines be presented in words?(2) How can it be formulated in the language of algebra? (3) How can it be proved? (4) What is its importance?* The teacher advised that the students write their answers to the questions once they were sure they understood the subject.

The first difference between the two groups has shown when, a few days later, the teacher asked to see the notes made by the learners as a part of their homework assignment. This request surprised some students. After all, the teacher did not request the written answers, she had only recommended them as potentially helpful. And yet, whereas only 4 out of the 9 NewComers had anything written to show, the OldTimers, with no exception, were able to come up with the kind of notes the teacher was asking for. The two groups differed further in the nature of the available record. As a rule, the OldTimers' answers to the teacher's questions were simply the relevant passages copied from the textbook. Of the four NewComers who did make notes, only one answered all four questions, whereas the sole focus of the other three sets of records was the proof of the cosine law (question 3 in the work plan.) Two of these proofs were quite unlike anything that could be found in other students' notebooks, so it was clear that these were students' reconstructions rather than quotes from the book.

Impressed by this visible disparity, the teacher asked whether anybody in the class felt a need for an additional explanation. This time, there was no difference between the OldTimers and NewComers: All the students felt that the topic has been understood. In spite of this, the teacher declared her wish to probe a bit further. She asked the class to formulate the law of cosines and to prove it in writing. The request was accompanied by a blackboard drawing of a triangle, marked with letters different from those that appeared in the textbook. The following passage from the teacher's journal presents students' reaction to the previously unannounced test:

> Several OldTimers started complaining: "We learned at home with the letters A, B, C and we got used to them"… The Newcomers did not show any sign of surprise. All of them, even Boris, usually the slowest, finished quickly.

Type of response	Number of responses	
	OldTimers	**NewComers**
Full proof, textbook version	1	6
Full proof, modified version	-	2
Partial, erroneous proof	1	-
No proof	8	1

Table 1: Students' responses to the request to prove the law of cosines

As shown in Table 1, the results attained by the two groups could hardly be more dissimilar: While all NewComers but one succeeded in the task, only one of the OldTimers was able to produce a reasonable proof. Moreover, two of the NewComers came up with their own versions of the proof, the type of response that is usually taken as the most persuasive evidence of understanding.

OldTimers (translated from Hebrew)	**NewComers** (translated from Russian)
***Ada**, who* did not *succeed in reproducing the proof:* I read the chapter in the book and tired to understand When I felt I understood, I copied the proof to the notebook	***Sonya,** who succeeded in reproducing the proof:* I read the proof a number of times, trying to remember and making notes on a separate page. I reproduced the proof without writing and I wrote the proof from memory with the book closed. I compared the proof to the one in the book. I then read and tried to understand the examples [of application] in the book
***Liora,** who* did not *succeed in reproducing the proof:* Copied the verbal formulation [of the cosine law], drew a triangle *in the head* [the student's own emphasis], read the verbal presentation and translated to letters in the head. Compared the formula to the one in the book and copied into the notebook. Read the proof and understood what they did. Solved the problems with the help of the formula. In case [I] could not do it, read the solved example.	***Misha,** who succeeded in reproducing the proof:* I began by translating [to Russian] of all the words in the theoretical text that were unclear to me. I read the theorem again until I understood its proof. When I was sure I understood the theorem, I drew a triangle with vertices marked differently than in the book and I wrote the new proof without looking into the book. After I finished, I checked the correctness of the proof with the help of the book. I read and understood the solved examples [of problems] in the book and began solving the homework problems.

Table 2: Representative responses to the question
How did you learn? Describe the process in some detail.

Once they completed their proofs, the students were asked to describe in writing the steps they performed while implementing the homework assignment. The

NewComers were allowed to respond in Russian. The English version of representative answers can be found in Table 2. The two columns give rise to two strikingly different pictures of the learning process: Whereas the OldTimers satisfied themselves with reading the book and answering the teacher's four questions by copying the relevant passages from the book, the NewComers intertwined reading the textbook exposition with their own independent attempts to formulate and prove the theorem.

We may now sum up and say that the OldTimers and NewComers differed in a consistent manner both in the way they learned and in the results attained. The learning process of the NewComers was clearly associated with their greater success on the test. The fact that the sequence of steps performed by the only OldTimer who managed to produce a correct proof was closer to that of NewComers than to that of OldTimers confirms this latter claim: There seems to be a tight correspondence, perhaps even a causal relationship, between the way NewComers learned and the effectiveness of their learning.

DEFINING SUBSTANTIAL AND RITUALIZED LEARNING

The first thing that strikes the eye in our data is that NewComers' and OldTimers' actions seem to have been directed at different recipients. The fact that the OldTimers implemented all the tasks required by the teacher apparently without asking themselves why they were performing these particular steps shows that, for these learners, the teacher was the ultimate addressee. NewComers, unlike OldTimers, did not perform all the prescribed tasks, and if they did, they did not leave any written records, evidently not being bothered about showing their work to the teacher. Thus, whatever these latter students did at home, they did it for themselves, according to their own assessment of its importance. In this activity, they were their own judges, and we have grounds to suspect that in this role, some of them were more exacting than anybody else, including the teacher.

Activities that have different addressees are usually perceived as having different goals. Clearly, in the eyes the OldTimers the process of learning was the end in itself, whereas the only thing that really counted for the NewComers was a certain product of the process, one that could be trusted to outlast the activity itself. In other words, the NewComers wanted the learning-induced change to be robust and durable. The desired lasting transformation can best be described in terms borrowed from what Harré & Gillet (1995) call *discursive psychology* and what was named *communicational approach to cognition* by other writers (Sfard 2001, Sfard & Lavi 2005; Ben Yehuda et al. 2005). According to the basic tenet of this approach, *thinking can be usefully conceptualized as a form of communication*, with this latter term signifying interaction that does not have to be audible, verbal, synchronic or directed at others. Within this framework, school *learning becomes the activity of changing one's discursive ways in a certain well defined manner*. In particular, learning to think mathematically is tantamount to being initiated into a special form of discourse, known as mathematical. Armed with this conceptual apparatus we may now say that for the NewComers, learning was the activity of introducing a lasting change into their own discursive activity, whereas for OldTimers it meant an episodic, ritualized participation in a discourse initiated by others.

We decided to call the two types of learning *substantial* and *ritualized*, respectively. In ritualized learning the learner engages in the mathematical discourse only in response to other person's request and for this other person's sake. In contrast, substantial learning may be defined as one that results in turning the new discourse from its initial status of a *discourse-for-others* into a *discourse-for-oneself*, that is, into a discourse in which this person is likely to engage spontaneously while solving problems and trying to answer self-posed questions.[4] This special kind of learning has a lasting effect on one's communication with oneself, that is, on this person's thinking.

The NewComer' strenuous effort toward substantial learning, noticed in the learning episode reported above, could be observed all along our extensive study, whatever the aspects of learning considered in its different segments. This effort was clear whether we were watching the students simplifying a complex algebraic expression, proving a trigonometric identity or trying to collaborate with others in solving a non-standard problem. On these diverse occasions, the NewComers' wish to turn the new discourse into a communication with themselves was evidenced also by their constant backtracking and self-examination, by their conspicuous preference for individual work, by their care for the appropriateness of their mathematical expression, and more generally, by their insistence on following all those rules of communication which they considered as genuinely 'mathematical'.

DEFINING IDENTITY[5]

Why talk about identity?

The striking dissimilarities between the OldTimers' and NewComers' learning called for explanation. Although we had a basis on which to claim the existence of some systematic differences in the teaching practices in the former Soviet Union and in Israel, these differences did not seem to tell the whole story. A teaching approach might have been responsible for the NewComers' acquaintance with certain techniques, but this fact, *per se*, did not account for the students' *willingness* to use these methods. We felt that to complete the explanation, we needed to clarify why the

[4] The term *discourse-for-oneself* is close to Vygotsky's idea of *speech-for-oneself*, introduced to denote a stage in the development of children's language (see e.g. Vygotsky 1987, p.71). Our terms also brings to mind the Bakhtinian distinction between *authoritative discourse*, a discourse that "binds us, quite independently of any power it might have to persuade us internally"; and *internally persuasive* discourse, one that is "tightly woven with 'one's own world.' (Bakhtin, 1981, pp. 110-111.)

[5] For a more extensive presentation of the topic see Sfard & Prusak 2005.

participants of our study were among those students who actually took advantage of the learning opportunities created by their teachers.

Yet another obvious explanation for the effectiveness of the NewComers' learning was that their immigrant status amplified their need for success.[6] Although certainly true, this account did not seem to be telling the whole story since it did not explain why school mathematics was singled out by the immigrant participants of our study as the medium through which to exercise their pursuit of excellence. Indeed, no other immigrant population, of which Israel has always had many, displayed a comparable propensity for mathematics. We decided to turn to the notion of identity, viewing it as a conceptual link between the collective and the individual.

Although the term "identity" is not new, it is only quite recently that it began drawing attention of educators at large, and of researchers in mathematics education in particular (see e.g., Boaler & Greeno, 2000; Nasir & Saxe, 2003; Cobb, 2004; Roth, 2004). Its new prominence is reflective of the general sociocultural turn in human sciences. The related time-honored notions of *personality*, *character*, and *nature,* being irrevocably tainted with connotations of natural givens and biological determinants, are ill-suited to the sociocultural project. In contrast, *identity*, which is thought of as man-made and as constantly created and re-created in interactions with others (Holland & Lave, 2003), seems just perfect for the task. Together with the acceptance of identity as the pivotal notion of the new research discourse comes the declaration about humans as active agents who play decisive roles in determining the dynamics of social life and in shaping individual activities.

We believe that the notion of identity is a perfect candidate for the role of "the missing link" in the researchers' story of the complex dialectic between learning and its sociocultural context. However, we also believe that this notion cannot become truly useful unless it is provided with an operational definition.

Defining identity

Its current popularity notwithstanding, the term 'identity' is usually employed without being operatively defined. The few defining attempts that can be found in the literature appear to be a promising beginning, but not much more than that. Gee (2001), who declares that "Being recognized as a *certain 'kind of person*' in a given context" (p. 99) is what he means by 'identity' also relates this notion to "the person's own narrativization" (p. 111), that is, to stories a person tells about herself. The motif of "person's own narrativization" recurs in the description proposed by Holland et al. (1998), even if formulated in different terms:

[6] As observed by Ogbu (1992), the status of minority is a doubly-edged sword. As shown by empirical findings, belonging to minority may, in some cases, motivate hard work and eventual success, whereas in some others it would have an opposite effect. Immigrants, whom Ogbu calls "voluntary minorities" as opposed to those whose minority status was imposed rather than chosen, are more likely than the others to belong to this former group.

> People tell others who they are, but even more importantly, they tell themselves and they try to act as though they are who they say they are. These self-understandings, especially those with strong emotional resonance for the teller, are what we refer to as identities. (p. 3)

If we said that these two descriptions are "promising beginnings" rather than fully satisfactory definitions, this is because of one feature that they have in common: They rely on the expression "who one is" or its equivalents. Unfortunately, neither Gee nor Holland and her colleagues make it clear how one can decide about "who" or "what kind of person" a given individual is. This said, their descriptions have an important insight to offer: By foregrounding "person's own narrativizations" and "telling who one is," these definitions link the notion of identity to the activity of communication. In an attempt to arrive at a more operational definition of identity we decided to build on the idea of identifying as communicational practice, thereby rejecting the notion of identities as extra-discursive entities which we merely "represent" or "describe" while talking.

In concert with the vision of identifying as a discursive activity, we suggest that identities may be defined as collections of stories about persons or, more specifically, as those narratives about individuals that are *reifying*, *endorsable* and *significant*. The reifying quality comes with the use of verbs such as *be*, *have* or *can* rather than *do*, and with the adverbs *always*, *never*, *usually*, etc. that stress repetitiveness of actions. A story about a person counts as *endorsable* if the identity-builder is likely to say, when asked, that it faithfully reflects the state of affairs in the world. A narrative is regarded as *significant* if any change in it is likely to affect the storyteller's feelings about the identified person. The most significant stories are often those that imply one's memberships in, or exclusions from, various communities.

As a narrative, every identifying story may be represented by the triple ${}_{B}A_{C}$, where A is the identified person, B is the author and C the recipient. Within this rendering it becomes clear that multiple identities exist for any person. Stories about a given individual may be quite different one from another, sometimes even contradictory. Although unified by a family resemblance, they depend both in their details and in their general purport on who is telling the story and for whom this story is meant. What a person endorses as true about herself may be not what others see enacted. To ensure that this last point never disappears from our eyes, we denote the different identities with names that indicate the relation between the hero of the story, the storyteller, and the recipient: ${}_{A}A_{C}$, a story told by the identified person herself, will be called A's *first-person* identity (1st P); ${}_{B}A_{A}$, a story told to its main character, will be named *second-person* identity (2nd P); finally, ${}_{B}A_{C}$, a story told by a third party to a third party, will be referred to as *third-person identity* (3rd P). Among all these, there is one special identity that comprises the reifying, endorsable, significant 1st P stories the storyteller addresses to herself (${}_{A}A_{A}$). It is this last type of stories that is usually intended when the word *identity* is used unassisted by additional specifications. Being a part of our ongoing conversation with ourselves, the first-person self-told identities are likely to have the most immediate impact upon our actions.

With the narrative definition, human agency and the dynamic nature of identity are brought to the fore, whereas most of the disadvantages of the traditional discourses on "personality", "nature" or "character" seem to disappear. The focus of the researcher's attention is now on things said by identifiers and no essentialist claims are made about narratives as mere "windows" to an intangible, indefinable entity. As stories, identities are human-made and not God-given, they have authors and recipients, they are collectively shaped even if individually told, and they can change according to the authors' and recipient' perceptions and needs. As discursive constructs, they are also reasonably accessible and investigable.[7]

Toward a theory of (narrative defined) identity

Since questions about identity can now be translated into queries about the dynamics of narratives, and since this latter phenomenon is amenable to empirical study, the narrative definition may be expected to catalyze a rich theory of identity. Much can now be said about identities simply by drawing on what is known about human communication and on how narratives interact one with another. Let us present some initial, analytically derived thoughts on how identities come into being and develop.

Actual and designated identities. The reifying, significant narratives about a person can be split into two subsets: *actual identity*, consisting of stories about the actual state of affairs, and *designated identity*, composed of narratives presenting a state of affairs which, for one reason or another, is *expected* to be the case, if not now then in the future. Actual identities are usually told in present tense and are formulated as factual assertions. Statements such as *I am a good driver*, *I have an average IQ*, *I am army officer* are representative examples. Designated identities are stories believed to have the potential to become a part of one's actual identity. They can be recognized by their use of the future tense or of words that express wish, commitment, obligation or necessity, such as should, *ought*, *have to*, *must*, *want*, *can/cannot*, etc. Narratives such as *I want to be a doctor* or *I have to be a better person* are typical of designated identities.

The scenarios that constitute designated identities are not necessarily desired, but are always perceived as binding. One may expect to "become a certain type of person,"

[7] For all these obvious advantages, one may claim that "reducing" identity to narratives undermines its potential as a sense-making tool. Story is a text, the critic would say, and identity is also, maybe even predominantly, an experience (see e.g. Wenger, 1998). Although we agree that identities originate in daily activities and in the "experience of engagement", we also posit that it would be a category mistake to claim that these characteristics disqualify our narrative rendering of identity. Indeed, it is our *vision* of our own or other people's experiences, and not these experiences as such, that constitutes identities. Rather than viewing identities as entities residing in the world itself, our narrative definition presents them as *discursive counterparts* of one's lived experiences.

that is, to have some stories applicable to oneself, for various reasons: because the person thinks that what these stories are telling is good for her, because these are the kinds of stories that seem appropriate for a person of her sociocultural origins or just because they present the kind of future she is designated to have according to others, in particular to those in the position of authority and power. More often than not, however, designated identities are not a matter of a deliberate rational choice. A person may be led to endorse certain narratives about herself without realizing that these are “just stories” and that they have alternatives.

Designated identities give direction to one’s actions and influence one’s deeds to a great extent, sometimes in ways that escape any rationalization. For every person, some kinds of stories have more impact than some others. *Critical* stories are those core elements, which, if changed, would make one feel as if one’s whole identity changed: The person’s ‘sense of identity’ would be shaken and she would lose her ability to tell in the immediate, decisive manner which stories about her are endorsable and which are not. A perceived persistent gap between actual and designated identities, especially if it involves critical elements, is likely to generate a sense of unhappiness.

Where do designated identities come from? The role of significant narrators. Being a narrative, the designated identity, although probably more inert and less context-dependent than actual identities, is neither inborn nor entirely immutable. Like any other story, it is created from narratives that are floating around. One individual cannot count as the sole author even of those stories that sound as if nobody has told them before.

To put it differently, identities are products of discursive diffusion – of our tendency to recycle strips of things said by others even if we are unaware of these texts’ origins. Paraphrasing Mikhail Bakhtin, we may say that any narrative reveals to us stories of others.[8] Identities coming from different narrators and being addressed at different audiences are in a constant interaction and feed one into another. These stories would not be effective in their relation-shaping task if not for their power to contribute to the addressees’ own narratives about themselves and about others. Thus, the people to whom our stories are told, as well as those who tell stories about us, may be tacit co-authors of our own designated identities. Either by animating other speakers or by converting their stories about us to the first person, we incorporate our 2nd and 3rd person identities into our self-addressed designated identities.

Another important sources of one’s own identity are stories about others. There are many possible reasons for turning such narratives into first person and incorporating them into one’s own designated identity. Thus, for example, the identity-builder may be attracted either to the heroes of these narratives or to their authors. Another reason may be one’s conviction about being “made” in the image of a certain person (e.g., of

[8] Bakhtin (1999) spoke about utterances and words rather than stories.

socially deprived parents, alcoholic father or academically successful mother) and "doomed" to a similar life. Whether a story told by somebody else does or does not make it into one's own designated identity depends, among other things, on how significant the storyteller is in the eyes of the identified person. *Significant narrators,* the owners of the most influential voices, are carriers of those cultural messages that will have the greatest impact on one's actions.

Learning as closing the gap between actual and designated identities. It is now not unreasonable to conjecture that identities are crucial to learning. With their tendency to act as self-fulfilling prophecies, identities are likely to play a critical role in determining whether the process of learning will end with what counts as success or with what is regarded as failure.

These days, in our times of incessant change, when the pervasive fluidity of most social memberships and of identities themselves is a constant source of fears and insecurities, the role of learning in shaping identities may be greater than ever. Learning is our primary means for making reality in the image of fantasies. The object of learning may be the craft of cooking, the art of appearing in media or the skill of solving mathematical problems, depending on what counts as critical to one's identity. Whatever the case, learning is often the only hope for those who wish to close a critical gap between their actual and designated identities.

IDENTITY AS AN INTERFACE BETWEEN CULTURE AND LEARNING

The designated identities of NewComers and of OldTimers

Let us go back to our study on NewComers and OldTimers learning mathematics together and show how our conceptual apparatus helps us in answering the question about cultural embeddeddness of learning. Below we argue that designated identities of the OldTimers and of NewComers were the channel through which these students' cultural background was making its way into their mathematical learning.

To map NewComers' and OldTimers' designated identities, we listened to their stories about themselves told to their teacher on various occasions. True, what we really needed were *self-addressed* stories of the type ${}_{A}A_{A}$ rather than ${}_{A}A_{Teacher}$, because this former type of narrative was more likely to interact significantly with one's actions. This preference notwithstanding, we were confident that the teacher-addressed designated identities would prove informative, especially if they displayed diversity paralleling the observed differences in learning. Further, we made certain deductions regarding the NewComers' and OldTimers' expectations from themselves on the basis of their self-referential remarks, of their comments about others (e.g. the teacher of fellow students), and of our own observations on the ways they acted. As a background, we used interviews with the students' parents and with other teachers. What was found with the help of this multifarious evidence displayed intra-group uniformity and inter-group differences comparable in their salience to those observed previously in the context of the students' learning.

	OldTimers	NewComers
Future plans ("What do you want to do in future?")	• [What I want to do] changes, because I change • For me, the only important thing is to be happy, and I don't have any particular profession in mind.	• In Russia I knew all the time that I'll follow in my brother's footsteps and learn computers. • From the earliest childhood I dreamt to be a medical doctor, like my mother.
The reasons for learning mathematics ("I learn mathematics because…)	• matriculation certificate with advanced mathematics will help me to get to the university, especially if the grade is high • I have to pass matriculation examination if I want to achieve anything in life. • it is obligatory	• I need knowledge and good education, and I love learning. • mathematics is my favorite school subject • I need to be a "full-fledged human being" and I want to feel I did something in life. • for me learning mathematics means creativity • mathematics is important and I like it very much

Table 3: Elements of OldTimers' and NewComers' designated identities

As can be seen from the students' responses to the question "What do you want to do in future?" presented in Table 3, probably the most obvious critical element of the NewComers' vision of themselves in the future was their professional career. Their tendency to identify themselves mainly by their designated professions stood in stark contrast to the OldTimers' declarations on their need "to be happy" and the latter interviewees' adamant refusal to specify any concrete plans for the future. The professions desired by the NewComers (e.g., computer scientist, medical doctor, engineer) were all related to mathematics, and this appeared to account for these students' special mathematical proclivity. And yet, there seemed to be more to these students' inclination toward mathematics than just the wish to promote their professional prospects. According to the NewComers' frequent remarks, the special attraction of mathematics was in the fact that its rules could be seen as universal rather than specific to a particular place or culture. While explaining why they chose to learn advanced mathematics (see students' completions of the sentence "I learn mathematics because…" in Table 3), the NewComers spoke about the knowledge of mathematics as a necessary condition for her becoming "a fully-fledged human being." We have thus reason to claim that mathematical fluency as such, and not just anything that could be gained through it, constituted the critical element in the NewComers' 1st P designated identities. In contrast OldTimers, in explaining their choice of advanced mathematics course, stressed the fact that matriculating in this subject with high grades would largely increase their chances for being accepted to the university. In other words, if OldTimers were attracted to mathematics it was mainly, perhaps exclusively, because of its role as a gatekeeper.

To sum up, the NewComers' designated identities portrayed their heroes as exemplars of what the immigrant students themselves described as "the complete humans," with this last term implied to have a timeless, universal, generally accepted meaning, and with mathematical fluency being indispensable for the completeness. In contrast, the OldTimers expected to have their future life shaped by their own wishes and needs, which, at this point in time, were seen as fluid and, in the longer run, unforeseeable. This also points to a distinct meta-level difference between the two groups: Whereas the NewComers saw their highly prescriptive designated identities as given and apparently immutable, just like the mathematics they wanted to master, the OldTimers' expected their 1st P identities to evolve with the world in tandem.

In accord with our expectations, all this seemed to account, at least in part, for our former findings about the difference between OldTimers' and NewComers' learning. The NewComers needed mathematical fluency in order to close the critical gap between their actual and designated identities. For the OldTimers, this fluency was something to be shown upon request, like an entrance ticket that could be thrown away after use and that had no value of its own. Since mathematical skills did not constitute a critical element of the OldTimers' designated identities, these skills' absence or insufficiency did not create any substantial learning-fuelling tension.

On the cultural roots of designated identities

Where does the disparity between NewComers' and OldTimers' designated identities come from? was the last question we had to address in order to complete our story of designated identity as a link between learning and its sociocultural setting. More specifically, we needed to account for the fact that mathematical fluency constituted the critical element of the NewComers' designated identities but did not seem to play this role in the identities of OldTimers.

The first thing to say in this context is that given the NewComers' immigrant status, their being well versed in mathematics appeared of a redemptive value: The universality of mathematical skills was likely to constitute an antidote to these students' sense of local exclusion. To put it in terms of identity, we conjecture that whereas NewComers were bound to identify themselves as outsiders to their local environment, mathematical prowess was one of those properties that compensated them with the more prestigious, place-independent status of "people of education and culture."

Clearly, the idea that education at large, and the fluency in mathematics in particular, might counterbalance the less advantageous elements of their identity was not the young NewComers' original invention. In general, what the participants of our study expected for themselves was not unlike what their parents and grandparents wished for them. This is what transpired in both groups from the students' assertions about the full accord between their own and their parents' expectations, and from their remarks about the parents' impact on their choices (see sample responses to the question about the parents' expectations in Table 4). This said, there was an important difference between our two populations. Unlike in the case of NewComers, the OldTimers' parents were described as willingly limiting the area of their influence and leaving most decisions in the young people's own hands. We also found it quite telling that parents were rarely mentioned in the OldTimers' autobiographical testimonies, whereas the NewComers' accounts were replete with statements on the elders' authority and with explicit and implicit assertions on the parents' all-important role in their children's lives. Obviously, the OldTimers' parents' stories about their children's future were not as prescriptive as those of the NewComers, nor was the influence of these stories equally significant.

OldTimers	**NewComers**
• My parents want for me what I want myself. They want me to do what I want. • What is good for me – that's what they want for me. I also think that they find my plans appropriate. • My parents want me to be happy, so it is not so important for them what I'm going to do. • They want me to be what I want to be.	• My mother wants me to get good education. The process of learning itself, this is what is important to her. But a good matriculation certificate too, of course. She also wants me to study in the university. • I chose studying computers because my parents "pushed" in this direction. • My parents know best what's good for me. • For me, my grandma is the greatest authority • My mother tells me that if I meet an obstacle, I'll fail because of my laziness. I am lazy.

Table 4: Students' responses to the question about the parents' expectations regarding their children's future

Narratives about education as a universal social lever and about knowledge of mathematics as one of the most important ingredients of education evidently constituted a vital part of the NewComers' cultural tradition. In their native countries, their families belonged to the Jewish minority. According to what we were told both by the students and by their parents, these families had typically identified themselves as locally excluded but globally "at home" thanks to their fine education. Their sense of only partial attachment to the ambient community was likely the reason for the young people's relative closeness to their families. In the interviews, both the parents and the children sounded fully reconciled with their status of local outsiders. Proud of their cultural background and convinced about its universal value, they seemed to consider this kind of exclusion as the inevitable price for, and thus a sign of, the more prestigious, more global cultural membership. It seems, therefore, that the NewComers' identities as local outsiders destined to overcome the exclusion with the help of place-independent cultural assets such as mathematics were shaped by their parents' and grandparents' stories prior to the students' immigration to Israel.

Since significant narrators can count as voices of community, all these findings corroborate the claim that designated identities are products of collective storytelling – of both deliberate molding by others and of incontrollable diffusion of narratives that run in families and in communities. This assertion completes our empirical instantiation of the claim on designated identity as "a pivot between the social and the individual" aspects of learning (Wenger, 1998, p. 145).

CONCLUDING REMARKS

In this study, the narrative-defined notion of identity allowed us to get an insight into the mechanism through which the wider community, with its distinct cultural-discursive traditions, impinges on its members' mathematics learning. On this occasion, we presented substantial learning as an activity propelled by the tension between actual and designated identities. Let us conclude this talk with two comments on practical and methodological implications of this study.

First, although our account may sound as a praise of the NewComers' learning, there is, in fact, no side-taking in this report. Even if the NewComers' practices can count as somehow superior to those of the OldTimers in that they proved more effective in attaining the official goals of school instruction, we are well aware that the goals themselves may be a subject to critical reappraisal. In addition, the price to be paid for this type of learning practice may, for some students, be too high to be worthy. Although carefully crafted stories about one's "destiny" may sometimes work wonders, they are also likely to backfire when the burden of too ambitious, too tightly designated, or just ill-adjusted identities becomes unbearable.

Second, while constructing the conceptual framework supposed to help us in justifying the claim about the cultural embeddeddness of mathematics learning, we switched from the talk about identity as a "thing in the world" to the discourse in which this term refers to a type of narrative. The difference between these two renderings is subtle. The kinds of data the narratively-minded researcher analyzes in her studies is the same as everybody else's: these are stories people tell about themselves or about others to their friends, teachers, parents, and observers. The only distinctive feature of the narrative approach is that rather than treat the stories as windows to some other entity that stays the same when "the stories themselves" change, the adherent of the narrative perspective is interested in the stories as such, accepting them for what they appear to be: Words that are taken seriously and shape one's actions. Mapping the intricate relations between different kinds of narratives and fathoming the complex interplay between stories told and deeds performed was the sole focus of this study. By taking a close look at the narratives' movement between one generation to another and between the level of community to that of an individual and back, we hoped to be able to account for both the uniformity and the diversity typical of human ways of acting.

References

Bakhtin, M. (1981). *The dialogic imagination.* Austin, TX: University of Texas Press.

Ben-Yehuda, M., Lavy, I., Linchevski, L., & Sfard, A. (2005), Doing wrong with words or What bars students' access to arithmetical discourses. *The Journal for Research in Mathematics Education*, *36(3)*, 176 – 247.

Boaler, J., & Greeno, J. (2000). Identity, agency, and knowing in mathematical worlds. In J. Boaler (Ed.), *Multiple perspectives on mathematics teaching and learning* (pp. 45-82). Stamford, CT: Ablex.

Cobb, P. (2004). Mathematics, literacies, and identity. *Reading Research Quarterly, 39*, 332-337.

Gee, J. P. (2001). Identity as an analytic lens for research in education. *Review of Resrarch in Education, 25*, 99-125.

Harré, R., & Gillett, G. (1995). *The discursive mind*. Thousand Oaks, CA: Sage.

Holland, D. & Lave, J., Eds. (2001). *History in person: Enduring struggles, contentious practice, intimate identities.* Santa Fe: School of American Research Press & Oxford: James Currey.

Holland, D., Lachicotte, W. Jr., Skinner, D., & Cain, C. (1998). *Identity and agency in cultural worlds.* Cambridge, Massachussetts: Harvard University Press.

Nasir, N. S., & Saxe, G. B. (2003). Ethnic and academic identities: a cultural practice perspective on emerging tensions and their management in the lives of minority students. *Educational Research, 32(5)*, 14-18.

Ogbu, J.U. (1992). Understanding cultural diversity and learning. *Educational Researcher*, *21(8)*, 5-14.

Prusak, A. (2003). *The nature and role of cultural factors in the learning of mathematics.* Unpublished doctoral dissertation (in Hebrew). The University of Haifa, Israel.

Roth, W-M. (2004). Identity as dialectic: Re/making Self in urban school. *Mind, Culture, and Activity. 11(1)*, 48-69

Sfard, A. (2001). There is more to discourse than meets the ears: Learning from mathematical communication things that we have not known before. *Educational Studies in Mathematics, 46(1/3),* 13-57.

Sfard, A. & Lavi, I. (2005). Why cannot children see as the same what grownups cannot see as different? – Early numerical thinking revisited. *Cognition and Instruction*, *23(2)*, 237-309.

Sfard, A. & Prusak, A. (2005). Telling identities: In search of an analytic tool for investigating learning as a culturally shaped activity. *Educational Researcher, 34(4)*, 14-22.

Wenger, E. (1998). Communities of practice: Learning, meaning, and identity. Cambridge: Cambridge University Press.

CO-CONSTRUCTING ARTEFACTS AND KNOWLEDGE IN NET-BASED TEAMS: IMPLICATIONS FOR THE DESIGN OF COLLABORATIVE LEARNING ENVIRONMENTS

Peter Reimann

Research Centre for Computer-Supported Learning and Cognition (CoCo)
University of Sydney

Computer-based learning environments for science and mathematics education support predominantly individual learning; from first generation drill and practice programs to today's advanced, knowledge-based tutorial systems, one learner interacting with one computer has been the typical setting. Mathematics educators, however, increasingly appreciate the value of collaborative learning and include team-learning activities in their lessons. In this presentation, drawing on our research in science and design areas, an overview is provided of the approaches and lessons learned regarding computer-supported collaborative learning and a number of design guidelines for computer-supported collaborative learning environments are suggested. Since equations and graphs are so important in mathematics, particular attention is paid to the role of external representations (and their co-construction) for computer-mediated collaboration.

APPROCHES TO FOSTER COLLABORATIVE LEARNING

Why foster collaboration? There are two arguments for supporting individuals as well as groups in cooperative behavior. First, cooperative behavior and, thus, collaborative learning leads to better performance of students compared to individual or competitive learning (Barron & Sears, 2002; Johnson & Johnson, 2004). Second, individuals in a group do not automatically cooperate and act as a group. A huge amount of contributions is dedicated to enhance collaborative learning in computer-mediated and residential cooperative learning. Johnson and Johnson (2004) distinguish four different basic types of cooperative learning: formal cooperative learning, informal cooperative learning, cooperative base groups and academic controversy. Mostly, formal and informal cooperative learning are addressed by methods fostering collaborative behavior. In some cases, the different types of cooperative learning represent several steps in the progress of a group (e.g., a group starts with informal cooperative learning, establishes formal cooperative learning afterwards and, finally, builds a cooperative base group). While informal cooperative learning according to the definition of Johnson and Johnson (2004) is restricted to short time intervals, most programs and assistance focus on the enhancement of formal cooperative learning.

Numerous methods of assisting learners in small group formal cooperative learning have been proposed. Some approaches are on the level of instructional design demanding specific cooperation patterns such as Group Jigsaw, Reciprocal Teaching

2005. In Chick, H. L. & Vincent, J. L. (Eds.). *Proceedings of the 29th Conference of the International Group for the Psychology of Mathematics Education*, Vol. 1, pp. 53-68. Melbourne: PME.

or Problem-Based Learning. Other approaches are direct teaching of cooperative behavior, modeling, or scripting (e.g. Rummell et al., 2002). Especially for groups that are beginning a "collaborative episode" (i.e., there are no or little experiences in cooperative learning and the building of social relationships is at its beginning) such direct intervention is appropriated in order to avoid frustration and to reduce cognitive load. Even more experienced learners may benefit form assistance in cooperation: Especially in groups with many degrees of freedom related to cooperation and task fulfillment little or poor interaction is reported (e.g. Cohen, 1994).

The problem of poor peer interaction is well known in residential collaborative learning, but with the use of typed text-based computer-mediated communication this problem is likely to be increased. It is much more difficult to establish, perform and maintain basic cognitive mechanisms like turn-taking or grounding. But also and in particular social mechanisms like building positive interrelationships, establishing a group identity etc. are afflicted. Major causes for these difficulties derive from a lack of external cues as described in models of cues-filtered out and canal reduction.

Recent research in CMC-based (computer-mediated communication) collaborative learning has contributed a variety of technological/instructional approaches and solutions to overcome these problems. Especially scripting of collaboration (as a scaffolding mechanism) has gained attention in order to enhance turn-taking (Pfister & Mühlpfordt, 2002; Reiserer, Ertl & Mandl, 2002), design rationale (Buckingham-Shum, 1997) or reflection (Diehl, Ranney & Schank, 2001). Reiser (2002) differentiates between two basic mechanisms of these scaffolding techniques: Providing structure and problem orientation. Structured communication is one method to guide learners in the sense of an optimized behavioral model (e.g. problem solving heuristics) or a coordinated exchange between several learners. Furthermore, attention of learners can be drawn to relevant aspects or elements of a collaborative problem-solving process. Thus, scaffolding and scripting can avoid irrelevant or distracting tasks, strategies and processes.

Scripting as a scaffolding mechanism, however, is not always beneficial. Learner guidance in problem solving can also limit the degrees of learners' freedom. Reiser (2002, p. 263) states: "However, given the importance of connecting students' problem solving work to disciplinary content, skills, and strategies, it may also be important to provoke issues in students, veering them off the course of non-reflective work, and forcing them to confront key disciplinary ideas in their solutions to problems." In addition, when structuring interaction and discourse for learners, we always run the risk of interrupting spontaneous discourse. Scripting implies external guidance on sequence or categorization of contributions, but it is very difficult to identify discourse and patterns that are generally appropriate and effective.

In our recent research, we tried to avoid such a drastic and direct intervention that limits learner control by providing an inflexible structure. Instead of pre-structuring, we pursue what we call a "post-hoc structuring", i.e., we take the data derived form interactions (and additional variables assessed from learners) and re-use them for scaffolding. This way we avoid direct interference with the communication process, provide authentic material (based on learners' own contributions) and, hopefully, help students to become more self-efficient. Furthermore, this approach provides learners with accurate information about their current status within a group and group's progress and also with information on possible further directions that can optimize group functions (e.g., communication, group-members' interrelationships and learning or problem-solving outcomes). Before we have a closer look at our methods of collaboration management, a study is presented that analyses a discourse structuring approach.

SCAFFOLDING

In this study, we[9] analysed a scaffolding approach that is typical for what Reiser (2002) coined "providing structure". In this case, structure is provided on how student can communicate with each other. In particular, we looked at three levels of structuring (electronic) communication: Unstructured – a chat tool was provided to groups of (three) students; Simple-Structure: A graphical argumentation schema was provided on a shared whiteboard with four types of "nodes" (claim, pro- and contra-argument, sub-claim; Full-Structure: in this condition, seven node types had to be used (question, pro-and contra argument, idea, decision, fact, and miscellaneous, see Fig. 1) following the IBIS notational conventions (see Buckingham-Shum, 1996).

We ran an experiment with three conditions (Chat, Simple-Structure, Full-Structure) and 5 groups of 3 participants in each condition. Participants had to develop collaboratively an argument for a "wicked" environmental issue, the benefits and risks of transporting oil on sea with tankers. Our expectation was that the higher the degree of argument structure, the better the quality of the arguments a group will produce. In order to evaluate the quality of the arguments, we used the coding scheme of Newman and colleagues (Newman, Johnson, Webb & Cochrane, 1997) that has been developed to assess the quality of arguments exchanged in computer-mediated communication. This method yields a "critical thinking index" which varies between 0.0 and 1.0, with values close to 1.0 indicating higher argument quality.

Argument quality did indeed increase as a function of scaffolding through argument structuring, with a significant differences between all three conditions. It is worth noting, however, that increasing the structure led to a decrease in the frequency of arguments.

[9] Oliver Orth helped with the experimentation and data analysis.

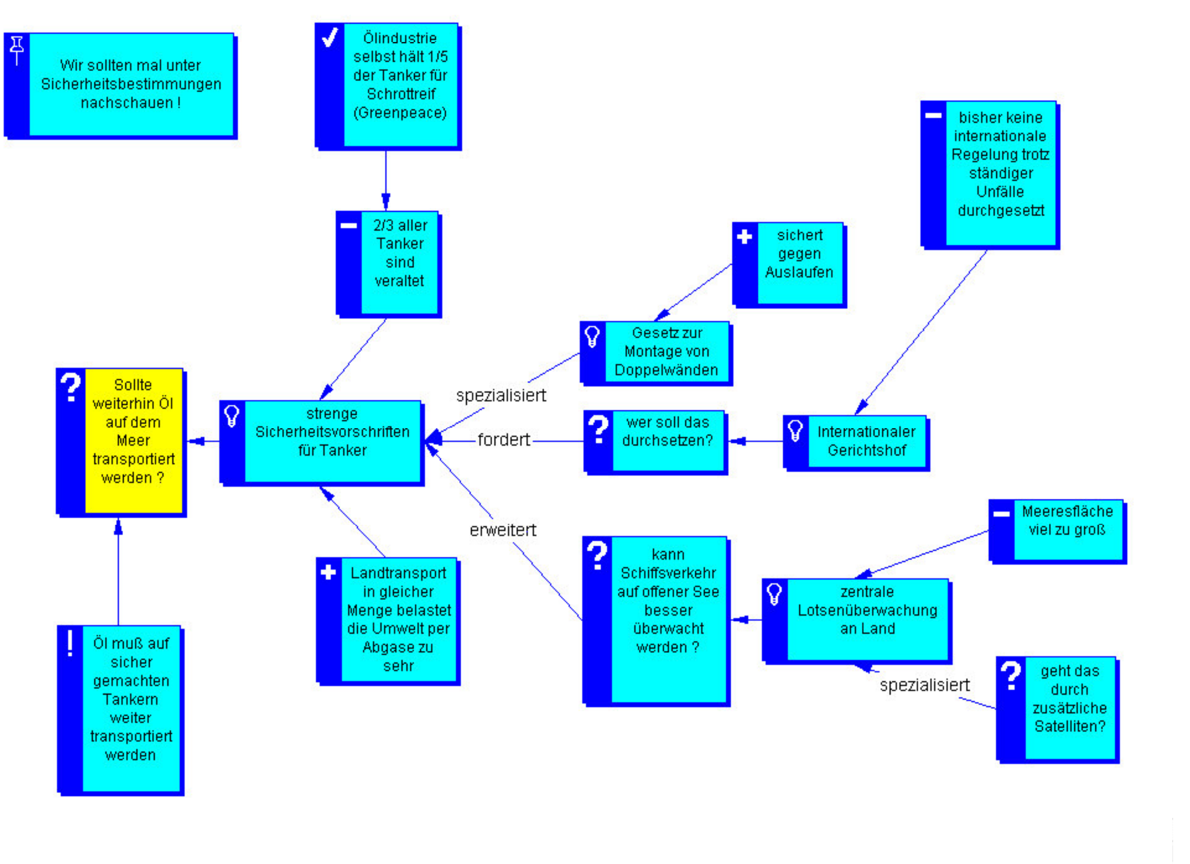


Figure 1: Argument graph using the full-structure (IBIS) notation

FEEDBACK AND GUIDANCE

Substantial research has been dedicated to find support mechanisms for online collaborators. Many authors discuss possibilities of scaffolding by structuring computer-mediated communication (e.g. Dobson & McCracken, 1997; Jonassen & Remidez, 2002; Reiser, 2002). Common to all these approaches is the provision of a structure for discourse and/or problem-solving. Instead of pre-structuring we pursue a way of post-hoc structuring interaction in online learning groups.

CMC itself provides the basis for this approach. During computer-mediated communication, all data can easily be stored and re-used for feedback purposes. In addition, software interfaces designed for CSCL (computer-supported collaborative learning) allow collecting individual quantitative data that can be used for further calculations in real time. Both data sources combined can easily be used to analyze individuals' as well as groups' performance automatically. In this way online learning groups provide the basis for feedback on their process without further interventions.

For instance, Barros and Verdejo (2000) describe an approach to provide feedback of group characteristics and individual behavior during computer-supported collaborative work based on a set of attributes that are computed out of data derived from learners' interactions. Their automatic feedback gives a qualitative description of a mediated group activity concerning three perspectives: a group's performance in reference to other groups, each member in reference to other members of the group, and the group by itself. Their approach allows extracting relevant information from online collaboration at different levels of abstraction. Although this approach seems to be very advantageous for enhancing online collaborators, Barros and Verdejo (2000) give no empirical evidence for the effectiveness of their asynchronous system. Jermann (2002, 2004) describes another possibility of providing feedback based on interaction data. He provides feedback on quantitative contribution behavior as well as learner-interaction during a synchronous problem solving task (controlling a traffic sign system). In an experiment, Jermann compared a group that received feedback about each individual learner's behavior. Another experimental group received feedback about the whole groups' success. He could show that a detailed feedback containing each individual's data enhanced learners' use of meta-cognitive strategies regarding problem-solving as well as discourse.

Our research group follows this line of feedback research. We[10] conducted studies to examine feedback effects on online collaborators during CSCL. One purpose of these investigations is to provide post-hoc scaffolding for subsequent problem solving. Another purpose is to use CMC, extract data from discourses and to provide abstracted views as a substitute for missing communication cues. In particular we investigated how the interaction in and the performance of small problem-based learning groups that cooperate via internet technologies in a highly self-organized fashion can be supported by means of interaction feedback as well as problem-solving feedback. Since the possibility of tracking and maintaining processes of participation and interaction is one of the advantages of online collaboration, ephemeral events can be turned into histories of potential use for the groups. We chose two ways to analyze how such group histories can be used for learning purposes. First, parameters of interaction like participation behavior, learners' motivation (self-ratings) and amount of contributions were recorded and fed back in an aggregated manner as an additional information resource for the group. This data could thus be used in order to structure and plan group coordination and group well-being. Second, we tracked group members' problem solving behavior during design tasks and provided feedback by means of problem-solving protocols. These protocols can be used to enhance a group's problem solving process for further tasks. Two studies testing our methodology in a synchronous and an asynchronous setting, respectively, are described next.

[10] The research reported in this section has been conducted in cooperation with Joerg Zumbach.

Automatic feedback in synchronous distributed Problem-Based Learning

The first laboratory experiment (Zumbach, Muehlenbrock, Jansen, Reimann & Hoppe, 2002) was designed as an exploratory study to test specific feedback techniques and their influence in an online collaboration learning environment.

For this purpose we designed a dPBL-learning environment. In a sample of 18 students of the University of Heidelberg we evaluated six groups of three members each. All students worked together synchronously via a computer network solving an information design problem. Each group was collaborating for about 2,5 hours (synchronously in one session). The task was to design a hypertext course for a fictitious company. All necessary task materials were provided online. In addition, all learning resources related to online information design were accessible as hypertext.

As a communication platform, the software EasyDiscussing was specifically developed for this experiment in cooperation with the COLLIDE-research group at Duisburg University, Germany. This Java-tool makes it possible to display a shared workspace to the whole group that can be modified by each member simultaneously. It contains drag-and-drop functions, thematic annotation cards like "text" (for general comments or statements), "idea", “pro” and “con” to structure the discussion, and it offers a chat opportunity as well (see Figure 2). All parameters are recorded in so-called "action protocols" and analyzed either directly or after the study. This makes it possible to check certain argumentative structures that become obvious during the course work, and also opens up the possibility to provide feedback based on the data produced.

Feedback parameters were gained in the following way: every 20 minutes students were asked about their motivation and their emotional state on a five item ordinal scale (parameters relating to the well-being function: “How motivated are you to work on the problem?” and “How do you feel actually?”). These were displayed to the whole group by means of dynamic diagrams (see Figure 3), showing each group member's motivation and emotional state with the help of a line graph. As a quantitative parameter supporting the production function two diagrams showed each group member's absolute and relative amount of contributions.

In order to test feedback effects we divided the groups into experimental groups that received feedback and into control groups which did not receive any feedback. Both groups had to do a pre- and post knowledge test, a test about attitudes towards cooperative learning (Neber, 1994), as well as some questions about their current motivation and emotional state. Besides our plan to test the techniques of how to provide feedback, we assumed that the experimental groups would be more productive since they were given parameters that would enable them to fulfill their well-being and production functions more easily, they. That means, they were assumed to contribute more ideas in an equally distributed manner, and show a greater amount of reflection, as far as interaction patterns were concerned, as opposed to the control groups.

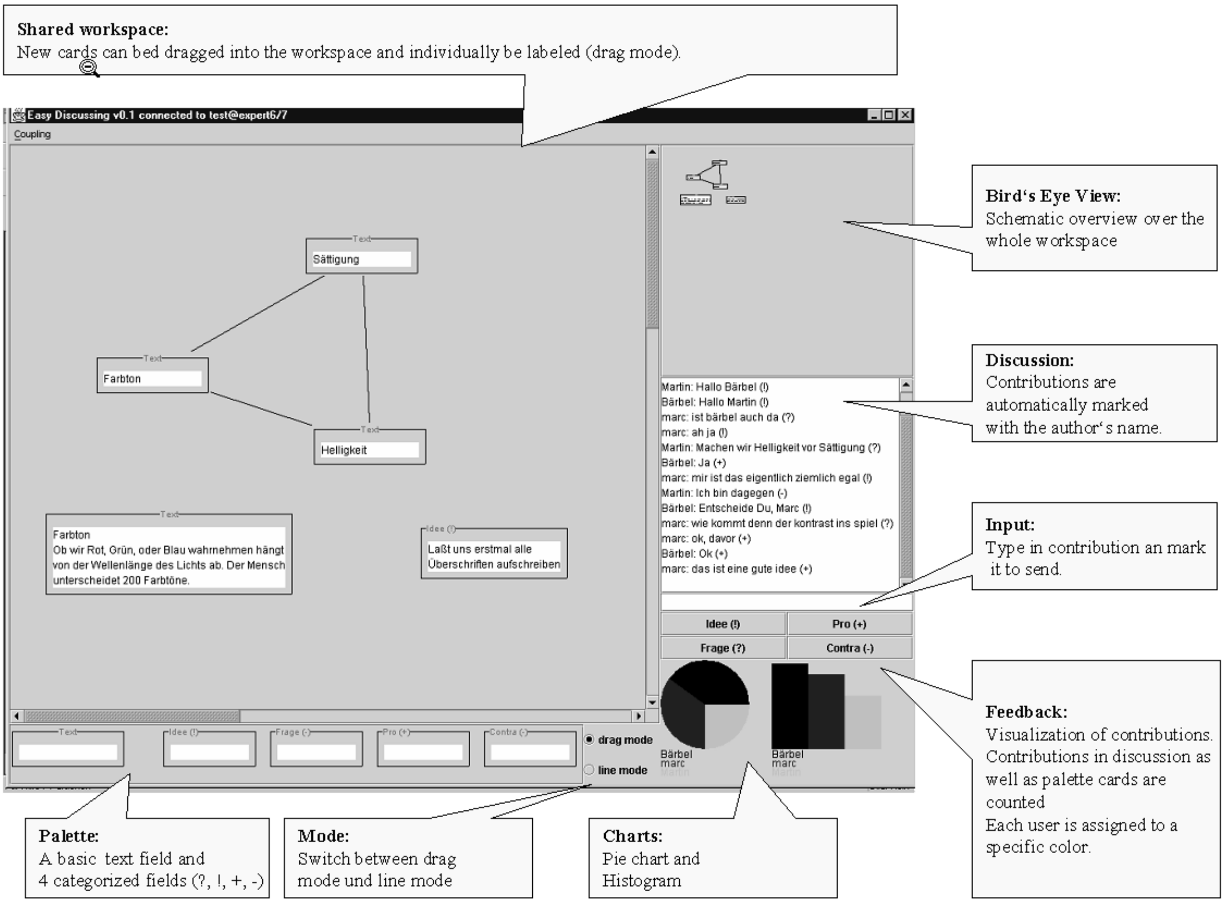
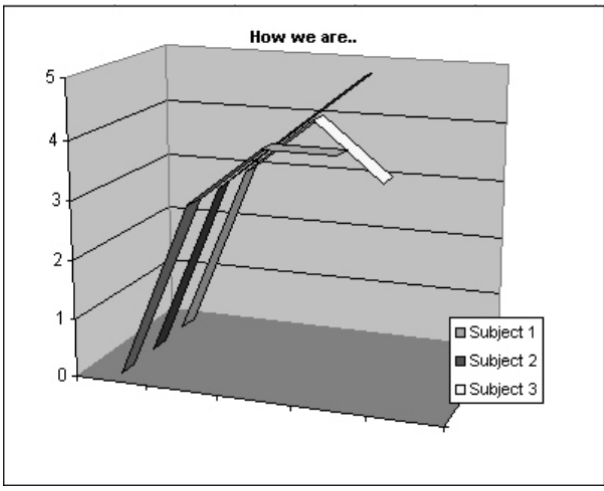


Figure 2. The design of the communication platform EasyDiscussing

The results of subjects' performance in the pre-test revealed no significant differences concerning domain knowledge. There were also no differences between both groups in post-test performance. Both groups mastered the post-test significantly better than the pre-test. There was no significant interaction between both tests and groups. We also found no significant differences regarding subjects' emotional data. The groups also showed no differences in pre- and posttests regarding motivation except a significant interaction between groups and time of measurement. While subjects in the control condition without feedback did not show differences in motivation, experimental groups had an increase from pretest to posttest. A closer look for interaction patterns in subjects' discussions revealed a significant difference in the number of dyadic interactions in groups that received feedback on their contributions.

Overall, the effects of this study indicate that some processes in computer-supported collaboration can be influenced in a positive manner by means of a steady tracking of parameters outside the task itself and immediate feedback of these to a group. Although intervention time in this experiment was short, we found positive influence of motivational feedback as well as feedback on contributions: communication patterns showed more interactive behavior for subjects of the experimental group. As a consequence of these effects, which indicate that our mechanisms have a positive influence on groups' production-function as well as group well-being, we decided to examine these feedback strategies further. For that purpose we arranged a long-time intervention study containing the same kind of visual feedback.



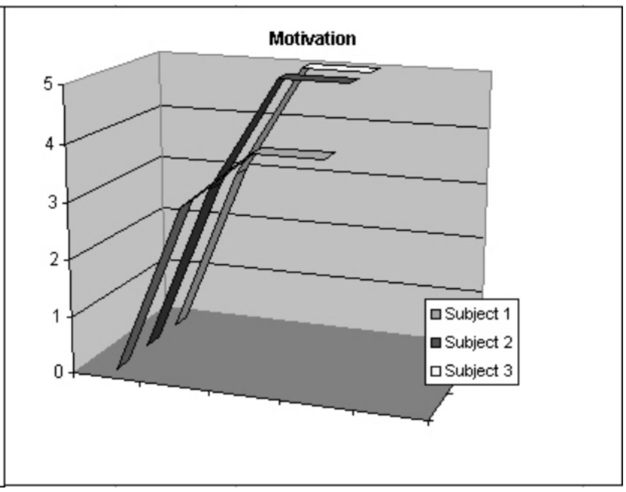


Figure 3. Feedback on emotion and motivation

Investigating the role of feedback mechanisms in long-time online learning

Our main objective in this study was to test different treatment conditions concerning feedback with groups that collaborated solely through an asynchronous communication platform over a period of four months. In this study we examined groups from three to five members – 33 participants overall. These groups participated in a problem-based course about Instructional Design that was conceived a mixture of PBL and Learning-By-Design. Learners were required to design several online courses for a fictitious company. These tasks have been presented as problems within a cover story. Each problem had to be solved over periods of two weeks (i.e. an Instructional Design solution had to be presented for the problem). As in study one, all materials were accessible online and, additionally, tutors were available during the whole course to support the students if questions emerged. At the end of each task, the groups presented their results to other groups. The asynchronous communication facility was based on a Lotus Notes® platform merging tools that can manage documents with automatic display possibilities for interaction parameters and problem-solving protocols (see Figure 4).

All documents as well as attachments were accessible over the collaboration platform. Meta-information showed when a document was created and who created it, so that interaction patterns became obvious and could be recorded. With the same technique of diagrams as in the former study, motivational and quantitative production parameters can be fed back to the user, referred to as *interaction histories*. Students' problem-solving behavior, however, had to be analyzed by the tutors themselves and had to be provided as text documents (*design histories*) in the group's workspace. Invisible for the students, a detailed action protocol was recorded in the background and was available later for analysis.

The groups were randomly assigned to one of four treatment conditions: with interaction history only, with design-history only, with both histories and without any feedback histories, i.e. a 2x2 design with the factors interaction history and design history. Several quantitative and qualitative measures to assess motivation, interaction, problem solving, and learning effects were collected before, during and after the experimental phase on different scales such as the student curriculum satisfaction inventory (Dods, 1997) or an adapted version of the critical thinking scale (Newman et al., 1997). We tried to answer the following question: What kind of influence does the administration of feedback in form of design and interaction histories, as well as their different combinations, have on students' learning? Generally, we assumed that groups with any form of histories would perform better than those without, especially as far as the motivational and emotional aspects supporting the well-being function and the production aspects supporting the production function of a group are concerned.

The results show encouraging outcomes in favor of the application of feedback within the group process. Groups that were shown design histories on their workspaces present significantly better results in knowledge tests, created qualitatively better products in the end, had produced more contributions to the task, and expressed a higher degree of reflection concerning the groups' organization and coordination. At the same time, the presence of interaction histories influenced the group members' emotional attitude towards the curriculum and enhanced their motivation for the task. Slight influences of the interaction history's visualization regarding number of contributions were also found on the production-function: Learners receiving this feedback produced more contribution than their counterparts without feedback. So far, it seems reasonable to conclude that the different kinds of feedback influence different aspects of group behavior. Whereas feedback in form of design histories seem to influence a group's production function according to McGrath's (1991) conception of group functions, feedback in form of interaction histories seems to have an effect also on the production-function, but mainly on the group's well-being function

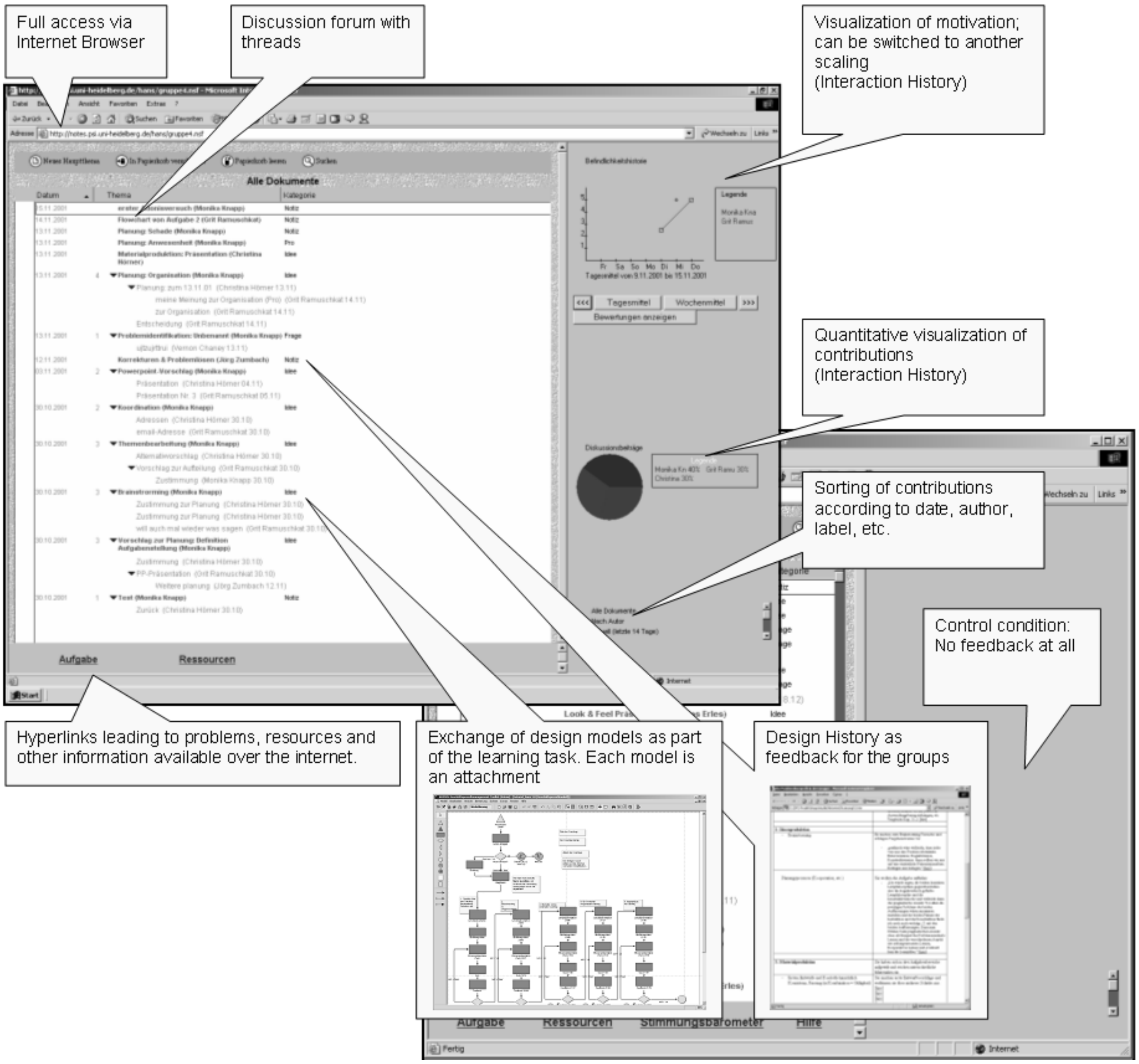


Figure 4. Asynchronous collaboration platform with feedback mechanisms.

TOWARDS ADAPTIVE VISUALISATION SUPPORT

In authentic, long term group work, it is the norm that people make use of a rich, diverse collection of communication systems, such as chat, discussion forums, and video conferencing. It is also typical that they make use of a range of tools and representational notations within one medium including, for example, written text and diagrams. We (Reimann, Kay, Yacef & Goodyear, in press) believe it is critical to begin to explore group support systems that can operate in the context of such media richness, exploiting the potentially huge amounts of data that could be available. We are particularly interested in three classes of learning that could occur in such situations:

- Learning to solve problems in a domain more effectively;
- Learning about the team, its members, and effective ways of cooperating and collaborating;
- Learning to use communication media and representational notations that match the demands of the tasks at hand, including tasks of member and collaboration management.

A number of researchers in the field of Computer-Supported Learning (CSCL) have begun to address this issue of collaboration management. Managing on-line collaboration by means of intelligent support can take a number of forms: mirroring, metacognitive and advice tools (Jermann, Soller & Muehlenbrock, 2001). They all require the ability to trace the interaction between the team members at some level of detail. We are building upon this work and intend to extend it into two directions: Firstly, in addition to supporting member interaction directly with feedback and/or advice systems, there is a need for learners to develop skills in choosing the right communication medium and tool for the situation at hand. Approaches to collaboration management that rely on a single communication medium, and/or on strongly restricted notational systems used for communicating (Conklin, 1993; Kuminek & Pilkington, 2001) need to be extended, because groups typically do not accept such limitations over longer stretches of time (Buckingham Shum, 1997). Having the choice among various communication and representation systems, however, adds to the demands groups face: they now have to deal with the additional issues of task-to-media fit (Daft & Lengel, 1984) and task-to-representation fit (Suthers, 2001). Secondly, we address human-computer interface issues extensively; not only because the management of task and interaction information distributed across various communication media raises serious attention and cognitive load issues, but also because of the social signals that come with using certain media (Robert & Dennis, 2005) and which have not been reflected sufficiently in research on computer-supported learning. We suggest an approach where the shared interface can be adapted to the needs of the work on the task as well as to the needs of interaction and member management. In the absence of a conclusive research base to derive advice from, our short term goal is to create an environment where such phenomena can be studied under controlled conditions and to experiment with various ways of visualizing information for groups and facilitators/moderators.

Adaptive Collaboration Visualisation

There has already been some work towards adaptive systems to provide advice on collaborative learning, for example (Constantino-Gonzalez, Suthers & Escamilla, 2003). There has also been recognition of the importance of social parameters, such as participation patterns (Barros & Verdejo, 2000). We will explore the use of adaptive information presentation using visualisations of the collaboration. These seem particularly promising because they are easier to implement than advice systems and no normative model of collaboration is required.

What to record. We are working on finding research-based answers to three questions around the process: (1) What to record about the learners' performance; (2) How to aggregate and then analyse the traced information; (3) What and how to visualize the results from step 2, in a manner that is adapted to the group's needs. With respect to question (1), we propose to capture *all* task- and group-related exchanges available, regardless of whether these involve the whole group, sub-groups, or individual members. Since we expect to be able to motivate the group members to help monitor their own interactions, we will be able to encourage the use of tools that we have set up to capture a rich record of interactions.

How to aggregate. An immediate effect of this is that we have to deal with large amounts of information. This must be analysed and summarised. Our approach with respect to question (2) is to collect the full set of available, un-interpreted data and then to perform a series of analyses to create both individual learner models and collective group models. We will use machine learning and data mining techniques (association rules, classification and clustering techniques such as hierarchic clustering, k-means, decision trees and data visualisation in particular) to identify patterns in groups' performance and relate those to outcome measures such as the quality of the groups' decision models and participants' satisfaction with the group process. Data mining and machine learning techniques have been successfully used for user modelling and, to a lesser extent, in education contexts. In particular, mining data based upon learners' interactions with a learning environment is promising (Bull, Brna, & Pain, 1995a).

Since a user model captures the system's beliefs about the learner's knowledge, beliefs, preferences and other attributes, it has the potential to play an important role in providing external representations of the individual and group learner models relevant to the group interaction and learning. There has been a growing appreciation of this possibility, with learner models being shared with learners in order to support reflection (Bull et al., 1995a; Bull, Brna, & Pain, 1995b, Crawford & Kay, 1993; Kay, 1995) and to help learners work collaboratively (Bull & Broady, 1997). The challenges in this project are to mine the available data sources to support the construction of a student model (Kay & Thomas, 1995), to provide natural interfaces that enable learners to see and understand the externalised form of that model (Uther & Kay, 2003), to explicitly contribute to it and, finally, but most importantly, to improve our understanding of the ways that this externalised user model can support learning and as well as the operation of the group.

What and how to visualize. Once relevant information is identified, the challenge remains how to communicate this back to the group (question 3). While the question of information visualisation has been researched before, including our own work (Uther & Kay, 2003; Zumbach & Reimann, 2003), research has so far been mainly limited to analysing individual displays of task and participation parameters (Jermann, 2004). The overall configuration of information displays – the interface elements that make up the shared work space – has been assumed as being static. We propose to dynamically adapt not only the content of individual information displays, but the *overall configuration* of information displays. For instance, when the group has to work on complex information together, social information should be reduced (in the absence of conflicts or member problems) so that all the cognitive resources can go into task information processing. Similarly, if interaction problems require attention, then the task information should temporarily be reduced and social information should be displayed with greater salience and detail. If both the task representation(s) and the social information representation(s) are properly adapted, then it should be feasible to provide suitable tradeoffs between the cognitive effort for the core task versus that for processing group and member information.

We also propose to differentiate more systematically between 'person awareness' and 'team awareness'. For instance, the video/audio display of a user – as a "rich" medium (Daft & Lengel, 1984) – primarily provides information about an *individual* group member. It does not depict information about the team as such. The user lists that are part of most chat tools, however, are a rudimentary *team awareness* component – showing who is currently "in" the group activity. Visualisations can, and probably should, play a much stronger role in supporting team awareness. For instance, Erikson and Kellogg (2000) make a number of suggestions on how to visualize social configurations of team members in digital spaces such as chat rooms.

Our current prototype collaboration environment comprises various synchronous and asynchronous communication and information representation tools, including a "digital table" that allows for co-located teamwork. We are experimenting with a number of computational approaches to aggregate collaboration information and identify psychologically and pedagogically meaningful patterns and trajectories. We are also developing means for visualising information relevant for task-, team-, and person-awareness. Building on these, we will experiment with ways to dynamically modify the respective information displays to make the overall interface adaptive to situational parameters (cognitive load, social conflicts, member problems) and to group members' preferences and individual needs.

CONCLUSIONS

In this paper, we have mainly looked at factors that apply to all forms of distributed collaborative learning, and have in particular dealt with issues that result from a lack of social awareness. While net-based group learning offers exciting opportunities to foster communication and reflection, one should not ignore the psychological challenges that arise from loosing face-to-face contact. In our recent work, we are also devoting increasing attention to the management of the user interface since adding all kinds of meta-information (helpful for reflection) to an already crowded screen space raises serious usability issues.

More would need to be said about the function of shared external representations, such as the symbols that appear on a shared whiteboard. Such shared representations do not only serve as a representation of shared knowledge, and thus play an pivotal role for grounding, they also help the group members to co-ordinate their work and to drive the agenda. The relation between such representations and the actions taken by group members need more attention in future research.

References

Barron, B. & Sears, D. (2002). Advancing understanding of learning in interaction: How ways of participating can influence joint performance and individual learning. In G. Stahl (Ed.), *Computer support for collaborative learning: Foundations for a CSCL community* (pp. 593-594). Hillsdale, N.Y.: Erlbaum.

Barros, B. & Verdejo, M. F. (2000). Analysing student interaction processes in order to improve collaboration. The DEGREE approach. *International Journal of Artificial Intelligence in Education, 11*, 221-241.

Buckingham-Shum, S. (1996). Design argumentation as design rationale. In A. Kent & J. G. Williams (Eds.), *The encyclopedia of computer science and technology* (Vol. 35, Supp. 20, pp. 95-128). New York: Marcel Dekker.

Buckingham Shum, S. (1997). Balancing formality with informality: User centred requirements for knowledge management technologies. *AAAI Spring Symposium (Mar. 24-26, 1997), Stanford University, Palo Alto, CA*. AAAI Press.

Bull, S., Brna, P., & Pain, H. (1995a). Extending the scope of the student model. *User Modelling and User-Adapted Interaction, 5*(1), 44-65.

Bull, S., Brna, P., & Pain, H. (1995b). Mr Collins: a collaboratively constructed, inspectable student model for intelligent computer assisted language learning. *Instructional Science, 23,* 65-87.

Bull, S. & Broady, E. (1997). Spontaneous peer tutoring from sharing student models. In B. du Boulay & R. Mizoguchi (Eds.), *Proceedings of the International Conference on Artificial Intelligence in Education* (pp. 143 – 150). Amsterdam: IOS Press.

Cohen, E. G. (1994). Restructuring the classroom: Conditions for productive small groups. *Review of Educational Research, 64,* 1-35.

Conklin, E. J. (1993). Capturing organizational memory. In R. M. Baecker (Ed.), *Groupware and computer-supported cooperative work* (pp. 561-565). San Francisco: Morgan Kaufman.

Constantino-Gonzalez, M., Suthers, D. & Escamilla, J. (2003). Coaching web-based collaborative learning based on problem solution differences and participation. International *Journal of Artificial Intelligence in Education, 13*(2-4), 263-299.

Crawford, K., & Kay, J. (1993). Metacognitive processes and learning with intelligent educational systems. In P. Slezak, T. Caelli & R. Clark (Eds.), *Perspectives on Cognitive Science* (pp. 63-77). Ablex Publishers.

Daft, R. L. & Lengel, R. H. (1984). Information richness: A new approach to managerial behavior and organizational design. *Research in Organisational* Behavior, 6, 191-233.

Diehl, C., Ranney, M., & Schank, P. (2001). Model-based feedback supports reflective activity in collaborative argumentation. In P. Dillenbourg, A. Eurelings, & K. Hakkarainen (Eds.), *European perspectives on computer-supported collaborative learning* (pp. 189-196). Universiteit Maastricht, the Netherlands.

Dobson, M., & McCracken, J. (1997). Problem based learning: A means to evaluate multimedia courseware in science & technology in society. In T. Muldner & T. C. Reeves (Eds.), *Educational Multimedia & Hypermedia*. Calgary: AACE.

Dods, R. F. (1997). An action research study of the effectiveness of problem based learning in promoting the acquisition and retention of knowledge. *Journal of the Education of the Gifted, 20*(4), 423-437.

Erikson, T. & Kellogg, W. A. (2000). Social translucence: An approach to designing systems that mesh with social processes. *ACM transactions on computer-human interaction, 7*(1), 59-83.

Jermann, P. (2002). Task and interaction regulation in controlling a traffic simulation. In G. Stahl (Ed.), *Computer Support for collaborative learning: Foundations for a CSCL community* (pp. 601-602). Hillsdale, NJ: Erlbaum.

Jermann, P. (2004). *Computer support for interaction regulation in collaborative problem-solving*. University of Geneva, Switzerland.

Jermann, P., Soller, A. & Muehlenbrock, M. (2001). From mirroring to guiding: a review of the state of the art technology for supporting collaborative learning. In P. Dillenbourg, A. Eurelings & K. Hakkarainen (Eds.), *European Perspectives on computer-supported learning* (pp. 324-331). Maastricht: University of Maastricht, the Netherlands.

Johnson, D. W. & Johnson, R. T. (2004). Cooperation and the use of technology. In D. H. Jonassen (Ed.), *Handbook of research for educational communication and technology* (pp. 785-812). Mawah, NJ: Lawrence Erlbaum Associate.

Jonassen, D. & Remidez, H. (2002). Mapping alternative discourse structures onto computer conferences. In G. Stahl (Ed.), *Computer Support for collaborative learning: Foundations for a CSCL community* (pp. 237-244). Hillsdale, NJ: Erlbaum.

Kay, J. (1995). The um toolkit for cooperative user modelling. *User Modelling and User-Adapted Interaction, 4*(3), 149-196.

Kay, J., & Thomas, R. C. (1995). Studying long term system use. *Communications of the ACM, 38*(7), 131–154.

Kuminek, P. A., & Pilkington, R. M. (2001). Helping the tutor facilitate debate to improve literacy using CMC. In T. Okamoto, R. Hartley, Kinshuk & J. P. Klus (Eds.), *Proceedings of the IEEE International Conference on Advanced Learning Technologies* (pp. 261-263). Madison, USA: IEEE Computer Society.

McGrath, J. E. (1991). Time, interaction, and performance (TIP). A theory of groups. *Small Group Research, (22)*2, 147-174.

Neber, H. (1994). Entwicklung und Erprobung einer Skala für Präferenzen zum kooperativen und kompetitiven Lernen [Developing and testing a scale for kooperative and competitive learning]. *Psychologie in Erziehung und Unterricht, 41*, 282-290.

Newman, D. R., Johnson, C. Webb, B., & Cochrane, C. (1997). Evaluating the quality of learning in computer supported co-operative learning. *Journal of the American Society for Information Science, 48*, 484-495.

Pfister, R., and Mühlpfordt, M. (2002). Supporting discourse in a synchronous learning environment: The learning protocol approach. In G. Stahl (Ed.), *Computer Support for collaborative learning: Foundations for a CSCL community* (pp. 581-582). Hillsdale, N.J.: Erlbaum.

Reimann, P., Kay, J., Yacef, K., & Goodyear, P. (in press). Adaptive visualisation support for self-managed learning groups. *Proceedings of the 12th International Conference on AI in Education.* Amsterdam, the Netherlands.

Reiser, B. (2002). Why scaffolding should sometimes make tasks more difficult for learners. In G. Stahl (Ed.), *Computer Support for collaborative learning: Foundations for a CSCL community* (pp. 255-264). Hillsdale, N.J.: Erlbaum.

Reiserer, M., Ertl, B. & Mandl, H. (2002). Fostering collaborative knowledge construction in desktop videoconferencing. Effects of content schemes and cooperation scripts in peer-teaching settings. In G. Stahl (Ed.), *Computer Support for collaborative learning: Foundations for a CSCL community* (pp. 379-388). Hillsdale, N.J.: Erlbaum.

Robert, L. P. & Dennis, A. R. (2005). Paradox of richness: a cognitive model of media choice. *IEEE Transactions on Professional Communication, 48*(1), 10-21.

Rummel, N., Spada, H., Hermann, F., Caspar, F. & Schornstein, K. (2002). Promoting the coordination of computer-mediated interdisciplinary collaboration. In G. Stahl (Ed.), *Computer support for collaborative learning: Foundations for a CSCL community* (pp. 558-560). Hillsdale, N.J.: Erlbaum.

Suthers, D. D. (2001). Towards a systematic study of representational guidance for collaborative learning discourse. *Journal of Universal Computer Science, 7*(3), 1-23.

Uther, J. & Kay, J. (2003). VlUM, a web-based visualisation of large user models. In *Lecture Notes in Computer Science* (Vol. 2702, pp. 198-202). Springer-Verlag GmbH.

Zumbach, J., Mühlenbrock, M., Jansen, M., Reimann, P. & Hoppe, H.-U. (2002). Multidimensional tracking in virtual learning teams. In G. Stahl (Ed.), *Computer support for collaborative learning: Foundations for a CSCL community* (pp. 650-651). Hillsdale, N.J.: Erlbaum.

Zumbach, J. & Reimann, P. (2003). Influence of feedback on distributed problem based learning. In B. Wasson, S. Ludvigsen & U. Hoppe (Eds.), *Designing for change in networked learning environments (pp. 219-228).* Dordrecht: Kluwer.

PLENARY PANEL

WHAT DO STUDIES LIKE PISA MEAN TO THE MATHEMATICS EDUCATION COMMUNITY?

Graham A. Jones

Griffith University, Gold Coast, Australia

In a real sense, PISA 2003 has touched the mathematics education community by stealth rather than by storm. Although PISA brings "baggage" commonly associated with international assessments, it takes some refreshing perspectives especially in the way that it envisions and assesses mathematical literacy. In this panel discussion we *focus on some of the issues associated with PISA: scrutiny of student performance, construct and consequential validity, what makes items difficult for students and the potential impact of PISA on mathematics education research. In selecting these issues we merely begin the debate and open the way for your participation.*

WHAT IS PISA?

The *Programme for International Student Assessment* ([PISA], OECD, 2005) is an international standardized assessment in reading literacy, mathematical literacy, problem-solving literacy and scientific literacy. It started in 1997 when OECD countries began to collaborate in monitoring the outcomes of education and, in particular, assessed the performance of 15-year-old school students according to an agreed framework. Tests have typically been administered to 4,500-10,000 students in each country. The first assessment in 2000 which focused mainly on reading literacy surveyed students in 43 countries while the second assessment in 2003 involved 41 countries and focused mainly on mathematics and problem solving. The third assessment in 2006 will largely emphasize scientific literacy and is expected to include participants from 58 countries. In this panel discussion we will concentrate on PISA 2003 and those aspects of it that deal with mathematical literacy.

THE PISA MATHEMATICAL LITERACY ASSESSMENT

In describing their approach to assessing mathematical performance, PISA documents (e.g., OECD, 2004a) highlight the need for citizens to enjoy personal fulfilment, employment, and full participation in society. Consequently they require that "all adults–not just those aspiring to a scientific career–be mathematically, scientifically, and technologically literate" (p. 37). This key emphasis is manifest in the PISA definition of mathematical literacy: " …an individual's capacity to identify and understand the role that mathematics plays in the world, to make well-founded judgments and to use and engage with mathematics in ways that meet the needs of that individual's life as a constructive, concerned, and reflective citizen" (OECD, p. 37; see also Kieran, plenary panel papers).

Reflecting this view of mathematical literacy, PISA documents (e.g., OECD, 2004a) note that real-life problems, for which mathematical knowledge may be useful, seldom appear in the familiar forms characteristic of "school mathematics." The

PISA position in assessing mathematics was therefore designed "to encourage an approach to teaching and learning mathematics that gives strong emphasis to the processes associated with confronting problems in real-world contexts, making these problems amenable to mathematical treatment, using the relevant mathematical knowledge to solve problems, and evaluating the solution in the original problem context" (OECD, 2004a, 38). In essence, mathematical literacy in the PISA sense places a high priority on mathematical problem-solving and even more sharply on *mathematical modelling*.

Although PISA's devotion to mathematical modelling has my unequivocal support, my experience tells me that it is not easy to incorporate effective mathematical modelling problems in a test that has fairly rigid time constraints. In addition, although the term mathematical modelling is relatively new in school mathematics (Swetz & Hartzler, 1991), there are instances of mathematical modelling even in the notorious public examinations of more than 50 years ago. I well remember the following problem in an examination that I took in 1953. It seems to me that it is a genuine modelling problem and it was certainly not a text book problem or a problem that anyone of that era had practised. Moreover, the fact that less than 10% percent of the 15 to 16-year-old students taking the examination solved the problem is both déjà vu and prophetic for those setting the directions for the PISA enterprise.

> In a hemispherical bowl of radius 8 inches with its plane section horizontal stands water to a depth of 3 inches. Through what maximum angle can the bowl be tilted without spilling the water? Give your answer to the nearest degree (University of Queensland, 1953)

Accordingly, even though members of our panel valued the PISA emphasis on real-world problems and mathematical modelling, there was no shortage of issues to debate. In particular, there were issues about the framework, the validity of the assessment, the construction of items, the measurement processes, the conclusions and the interpretations especially interpretations that cast the findings into the realm of an international "league table". Consequently, we faced a problem in selecting which issues to examine. Let me presage the papers of the other panellists by providing an entrée of the issues that reverberated over our internet highways.

WHAT ISSUES DOES PISA RAISE FOR MATHEMATICS EUDCATION?

As the conference theme was *learners and learning* we questioned whether PISA assessment really was designed to support a real-world approach to mathematics teaching and learning. We also raised questions about whether student performance in the PISA assessments mirrored student performance in other mathematics education research on learning and teaching. Although appropriate data was not easily accessible, we wondered what the PISA study told us about patterns of classroom activity in different cultures. Yoshinori Shimizu (plenary panel papers) did examine this from a cultural perspective by scrutinizing Japanese students' responses to some PISA items.

Issues associated with item validity, item authenticity, and item difficulty were consistently part of our discussions. The "triangular park problem" (see Williams, plenary panel papers) was hotly debated and members of the team even spent considerable time looking for triangular parks or car parks. This was part of our conversation on *real world* or *authentic assessment* and this issue is taken up further by Julian Williams under the broader topic of construct validity. Carolyn Kieran (see plenary panel papers) takes up the issue of "what makes items difficult for students?" She observes that the *difficulty levels* of some PISA items are problematic and raises doubts about how much we know about what students find difficult in certain mathematical tasks.

The politics of international assessment studies like PISA (OECD, 2004a) and *Trends in International Mathematics and Science Study* ([TIMSS], Mullis et al., 2004) were high on our debate list. Not only do these debates raise highly volatile issues and national recriminations, they also generate profound questions for those countries that are doing well and for those who are not. In addition to issues that focus specifically on the international league, assessment studies like PISA produce a range of related debates about factors such as gender, ethnicity, socio-economic status, systemic characteristics, approaches to learning, student characteristics and attitudes, and of course fiscal support (OECD, 2004b). Julian Williams (see plenary panel papers) tackles a number of these political issues especially those related to accountability: managing targets, dealing with league tables, and performance-related reviews.

There was considerable interest in discussing the impact of international assessment studies on mathematics education research. At the forefront of such issues is the question: What does PISA say to researchers interested in assessment research? Yoshinori Shimizu (see plenary panel papers) will talk about this more specifically as he refers to the benefits that can be gleaned by researchers through an examination of PISA's and TIMSS's theoretical frameworks, methodologies, and findings. For example, he notes that the detailed item scales and maps in PISA will enable researchers to perform a secondary analysis of students' thinking and accordingly gain a deeper understanding of learners and learning. Michael Neubrand (see plenary panel papers) also looks at the potential of PISA to stimulate research in mathematics education. He focuses on the structure of mathematical achievement especially in the way that PISA conceptualizes achievement through the aegis of a *mathematical literacy* framework. This gives rise to an interesting dialogue with respect to both individual and systemic (collective) competencies in mathematics and how they can be measured. There are of course other important questions such as "What do studies like PISA say to mathematics education researchers about methodological issues such as qualitative versus quantitative research?" Although this particular question is not directly addressed, the panel refers frequently to methodological issues and as such issues a challenge to the participants for further engagement and debate.

CONCLUDING COMMENTS

I believe that this panel discussion is most timely as I am not convinced that mathematics educators are as cognizant as they might be about the impact of the burgeoning industry that encompasses international studies like PISA (OECD, 2003) and TIMSS (Mullis et al., 2004). Although the build up and dissemination of PISA has been slow to take root in the mathematics education research community, the findings have certainly not gone unnoticed by national and state governments, educational systems, business leaders and parent groups. They know where their nation or their state came in the "league stakes" but they have little understanding of the intent and limitations of such studies. Accordingly, an important aim of this panel is to encourage mathematics education researchers to be more proactive not only in publicly illuminating and auditing research like PISA but also in identifying ways in which PISA can connect with and stimulate their own research. In the words of Sfard (2004, p. 6) we should exploit these special times in mathematics education:

> Confronting the broadly publicized, often disappointing, results of the international measurements of students' achievements, people from different countries started wondering about the possibility of systematic, research-based improvements in mathematics education

References

Mullis, I. V. S., Martin, M.O., Gonzalez, E. J., & Chrostowski, S. J. (2004). *TIMSS 2003 international mathematics report: Findings from IEA's trends in international mathematics and science study at the fourth and eighth grades*. Chestnut Hill, MA: Boston College, TIMSS & PIRLS International Study Center.

OECD (2004a). *Learning for tomorrow's world: First results from PISA 2003*. Paris: Author.

OECD (2004b). *Messages from PISA 2000*. Paris: Author.

OECD (2005, May 8). *Program for international student assessment* (PISA). Retrieved from http://www.pisa.oecd.org/document/56.

Sfard, A. (2004).There is nothing more practical than good research: On mutual relations between research and practice in mathematics education. In M. Niss (Ed.), *Plenary and regular lectures: Abstracts*. Copenhagen, Denmark: 10th International Congress of Mathematical Education

Swetz, F. & Hartzler, J. S. (1991). *Mathematical modelling in the secondary school curriculum*. Reston, VA: NCTM.

University of Queensland. (1953). *Mathematics B: Junior public examination*. Brisbane, Australia: Author.

FROM A PROFILE TO THE SCRUTINY OF STUDENT PERFORMANCE: EXPORING THE RESEARCH POSSIBILITIES OFFERED BY THE INTERNATIONAL ACHIEVEMENT STUDIES

Yoshinori Shimizu

Faculty of Education, Tokyo Gakugei University

The recent release of two large-scale international comparative studies of students' achievement in mathematics, the OECD-PISA2003 and the TIMSS2003, has the potential to influence educational policy and practice. A careful examination of their findings, theoretical frameworks, and methodologies provides mathematics education researchers with opportunities for exploring research possibilities of learners and learning.

BEYOND THE COMPETITIVE EMPHASIS IN REPORTS

The release of results of the OECD-PISA2003 (Programme for International Student Assessment, OECD, 2004) and the TIMSS2003 (Trends in International Mathematics and Science Study, Mullis, et al., 2004) in December 2004 received huge publicity through the media in Japan. The purposes of international studies such as PISA and TIMSS include providing policy makers with information about the educational system. Policy makers, whose primary interest is in such information like their own country's relative rank among participating countries, welcome a simple profile of student performance. Also, there is a close match between the objectives of PISA, in particular, and the broad economic and labour market policies of host countries. The match naturally invites a lot of public talk on the results of the study with both competitive and evaluative emphasis. This was the case in Japan.

There was one additional large-scale study in 2003 of student performance in mathematics in Japan. In the National Survey of the Implementation of the Curriculum, which has also been released recently (NIER, 2005), the students from grades 5 through 9 (N>450,000) worked on items that are closely aligned with the specific objectives and content of in Japanese mathematics curriculum. TIMSS2003 sought to derive achievement measures based on the common mathematical content as elaborated with specific objectives, whereas PISA2003 was explicitly intended to measure how well 15-years-olds can apply what they have learned in school within real-world contexts. The recent release of these studies should shed light on the new insight into learners and learning from multiple perspectives.

The large-scale studies, conducted internationally or domestically, provide a profile of a population of students from their own perspectives. We need to go beyond competitive emphasis in the reports of such studies to understand more about the profile of students' performance and to explore the possibilities of further research that such studies provide.

2005. In Chick, H. L. & Vincent, J. L. (Eds.). *Proceedings of the 29th Conference of the International Group for the Psychology of Mathematics Education*, Vol. 1, pp. 75-78. Melbourne: PME.

In this short article, a few released items of PISA2003 are drawn upon to propose that a careful examination of the findings, the theoretical framework and the methodology used as well, provides mathematics education researchers with opportunities to examine further research questions that might be formulated and addressed.

THE SCRUTINIES NEEDED

One of the distinct characteristics of the PISA2003, having mathematics as the major domain in the recent cycle of the project, is the way in which the results of student performance are described and reported. The mathematics results are reported on four scales relating to the overarching ideas, as well as on an overall mathematics scale. The characteristics of the items as represented in the map, which shows the correspondence between the item and the scale, provide the basis for a substantive interpretation of performance at different levels on the scale.

We can now take a closer look at the profile of students' response to the released items. Even the results of a few released items from PISA2003 suggest possibilities for conducting a secondary analysis and further research studies in order to develop deeper understanding of learners and learning. In particular, such items, or overarching ideas, as follows raise questions for Japanese mathematics educators, in particular, and mathematics education researcher, in general, to consider.

An Illuminating Example: SKATEBOARD

One of the items on which Japanese student performance looks differently from that of their counterparts elsewhere is in Question 1 of the item called SKATEBOARD (OECD, 2004, p.76). This short constructed response item asks the students to find the minimum and the maximum price for self-assembled skateboards using the price list of products given in the stimulus. The item is situated in a *personal* context, belongs to the *quantity* content area, and classified in the *reproduction* competency cluster. The results show that the item has a difficulty of 464 score points when the students answer the question by giving either the minimum or the maximum, which locates it at Level 2 proficiency. On the quantity scale, 74% of all students across the OECD community can perform tasks at least at Level 2. The full credit response has a difficulty of 496 score points, which places it at Level 3 proficiency. On the quantity scale, 53% of all students across the OECD community can perform tasks at least at Level 3.

When we look into the data on the students' response rate in each country, a different picture appears. Japan's mean score was significantly lower than the OECD average for the item (See Table 1) and the pattern in the percentages for students' responses look different from their counterparts in other countries.

Of note among the numbers in Table 1 is the lower percentage of correct responses from Japanese students than from their counterparts, as well as the higher no response rate. Students can find the minimum price by simply adding lower numbers for each part of the skateboard and the maximum price by adding larger numbers.

Country	Full Credit	Partial Credit	No Response	Correct
Australia	74.1	9.3	1.8	78.7
Canada	74.9	9.1	2.0	79.4
Germany	71.7	11.5	5.2	77.5
Japan	54.5	8.0	10.6	58.5
OECD Average	66.7	10.6	4.7	72.0

Table 1: The percentage of students' response for SKATEBOARD, Question1 (An excerpt from National Institute for Educational Policy Research, 2004, p. 102.)

The results suggest that some students, Japanese students, in this case, may be weak in handling multiple numbers where some judgment is required, assuming that they have little trouble in the execution of the addition procedure. We need an explanation with scientific evidence for the results.

Another Example: NUMBER CUBES

Another example comes from the result of the item called NUMBER CUBES (OECD, 2004, p.54). This item asks students to judge whether the rule for making a dice (that the total number of dots on two opposite faces is always seven) applies or not with the given four different shapes to be folded together to form a cube. The item is situated in a *personal* context, belongs to the *space and shape* content area, and classified in the *connection* competency cluster. The results show that the item has a difficulty of 503 score points, which places it at Level 3 proficiency. On the space and shape scale, 51% of all students across the OECD community can perform tasks at least at Level 3.

Country	Students' Choice of Correct Judgments					
	Four (Full)	Three	Two	One	None	No Res.
Australia	68.6	14.1	7.2	6.4	2.4	1.2
Canada	69.6	14.0	7.3	6.3	2.1	0.6
Germany	69.0	13.9	7.3	5.6	2.3	1.9
Japan	83.3	8.9	4.2	2.0	0.9	0.7
OECD Average	63.0	16.0	8.9	7.2	2.7	2.3

Table 2: The percentage of students' response for NUMBER CUBES (An excerpt from National Institute for Educational Policy Research, 2004, p. 108.)

The result shows that Japan's mean score was significantly higher than the OECD average as well as being higher than other participating countries (See Table 2). Also, the pattern of students' choice is slightly different from other countries.

In order to complete the item correctly, we need to interpret the two dimensional object back and forth by "folding" it to make the four planes of the cube mentally as a three-dimensional shape. The item requires the encoding and spatial interpretation of two-dimensional objects. Why did a group of students, once again Japanese students, perform well on this particular item? Does the result suggest that those students have a cultural practice with number cubes, or Origami, inside and outside schools? A further exploration is needed to explain the similarities and differences in students' responses among participating countries.

There are other insights offered by the recent international studies. The TIMSS2003 collected information about teacher characteristics and about mathematics curricula. The PISA2003 also collected a substantial amount of background information through the student questionnaire and the school questionnaire. These data on contextual variables as well as performance data related to the cognitive test domain give us rich descriptions of the learning environments of the learners.

As was mentioned above, the recent release of the two large-scale international achievement studies provides mathematics education researchers with opportunities for exploring research possibilities in relation to learners and learning. While we need to examine the results from each study carefully, we also need to synthesize the results from different perspectives as a coherent body of description of the reality of the learners.

References

Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., & Chrostowski, S. J. (2004). *TIMSS 2003 international mathematics report: Findings from IEA's Trends in International Mathematics and Science Study at the fourth and eight grades. Chestnut Hill, MA: Boston College, TIMSS & PIRLS International Study Center.*

National Institute for Educational Policy Research (2003). *The report of 2001 national survey of the implementation of the curriculum: Mathematics*. Tokyo: Author.

National Institute for Educational Policy Research (2004). *Knowledge and skills for life II: OECD- Programme for International Student Assessment 2003*. Tokyo: Gyosei

National Institute for Educational Policy Research (2005). *A summary of 2003national survey of the implementation of the curriculum: Mathematics*. Tokyo: Author.

Organisation for Economic Co-operation and Development (2003). *The PISA 2003 assessment framework: Mathematics, reading, science and problem solving knowledge and skills*. Paris: Author.

Organisation for Economic Co-operation and Development (2004). *Learning for tomorrow's world: First results from PISA 2003*. Paris: Author.

THE PISA-STUDY: CHALLENGE AND IMPETUS TO RESEARCH IN MATHEMATICS EDUCATION

Michael Neubrand

Dept. of Mathematics, Carl-von-Ossietzky-University, Oldenburg (Germany)

Beyond the results, a large scale study like PISA may also stimulate the area of research in mathematics education. Since an empirical study needs a sound conceptualization of the field - "mathematical literacy" in the case of PISA - mathematics education research and development may benefit from the structures of mathematical achievement defined for PISA. Further research can build upon the work done in PISA.

PISA, the "Programme for International Student Assessment" (OECD, 2001, 2004) came into the public focus mainly for the results and the prospective consequences to be drawn: "All stakeholders – parents, students, those who teach and run education systems as well as the general public – need to be informed on how well their education systems prepare students for life" (OECD, 2004, p 3). However, the PISA study deserves interest also from the point of view of research in mathematics education. This perspective is inherent to PISA: The PISA-report "considers a series of key questions. What is meant by 'mathematical literacy'? In what ways is this different from other ways of thinking about mathematical knowledge and skills? Why is it useful to think of mathematical competencies in this way, and how can the results be interpreted?" (OECD, 2004, p 36)

This paper draws attention to some of the impulses and challenges to mathematics education research coming from the PISA studies. We recognize both, the international study, and the national option in Germany which was based on an extended framework and included additional components.

SYSTEM RELATED DIAGNOSIS OF MATHEMATICAL ACHIEVEMENT

What are the aims of PISA? PISA's main focus is to measure the outcomes of the whole educational systems in the participating countries, and choses, as the most sensible group to investigate, the group of the 15 years olds in the countries. The key question therefore is on the system level: What do we know about the mathematical achievement and its conditions in an *educational system* compared to what one can observe in an international overview?

Apparently, this is not thoroughly in tune to the mainstream of mathematics education research. There are long and ongoing traditions in mathematics education which point to a contrasting aspect: What are an *individual's* thoughts, difficulties, sources, and strategies when learning mathematics? Our common interest is often more on an individual's understanding, or on the misunderstandings in the social communication among the individuals in the classroom. Thus, it does not wonder that international comparisons found and still find critical reactions, going back as far as

2005. In Chick, H. L. & Vincent, J. L. (Eds.). *Proceedings of the 29th Conference of the International Group for the Psychology of Mathematics Education*, Vol. 1, pp. 79-82. Melbourne: PME.

Hans Freudenthal's fundamental critique in the beginning of comparative studies in mathematics (Freudenthal, 1975).

Contrasting that tradition, the complementary question towards a *systems' efficiency* in mathematics teaching and learning is not less challenging. One has to define appropriate concepts and instruments to answer the question on a basis which incorporates the knowledge mathematics education research has given us so far. In fact, PISA took that challenge serious in a twofold way: The concept "mathematical literacy" forming the basis for testing mathematics achievement is explicitly bound to the mathematics education tradition (OECD, 2003; Neubrand et al., 2001); and vice versa, the PISA test gave rise to further developments of conceptualizing mathematical achievement (Neubrand, 2004). Thus, PISA provides theoretically based, and empirically working conceptualizations of mathematical achievement, which can be seen as an impetus to mathematics education research.

CONCEPTUALIZING MATHEMATICAL ACHIEVEMENT

Sources of the concept "mathematical literacy"

The specific idea of PISA is that the outcomes of an educational system should be measured by the competencies of the students. The key concept is "literacy". Three roots can be traced back: a tradition of pragmatic education (e.g., Bybee, 1997), Freudenthal's conception that "mathematical concepts, structures and ideas have been invented as tools to organise the phenomena of the physical, social and mental world" (Freudenthal, 1983), and considerations on what mathematics competencies are about (Niss, 2003). From there the PISA-framework developed that PISA aims to test the capability of students "to put their mathematical knowledge to functional use in a multitude of different situations" (OECD, 2003).

Conceptualizing „mathematical literacy" in the international PISA study

The domain "mathematical literacy" was conceptualized and related to the test items (problems) in the international PISA study by three components (Fig. 1).

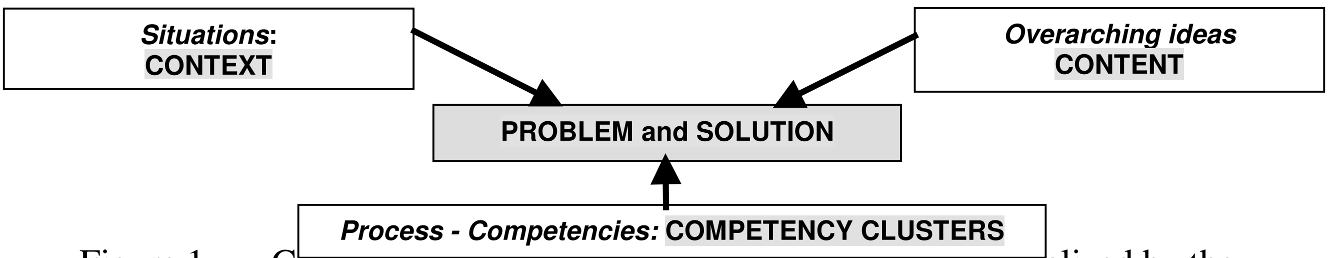


Figure 1. Components of mathematical problems as conceptualized by the international PISA framework (OECD, 2003, p. 30).

It is one of the major impetuses (and challenges) to mathematics education research that (and if) a list of mathematical competencies, accumulated in the Competency Clusters ("Reproduction" - "Connections" - "Reflection"), may hold as "key characteristics" (OECD, 2004, Annex A6) to construct an appropriate instrument to test mathematical achievement. In 2004 PISA reported countries' achievement differentiated by the content-dimension, and it will be a matter of further research to clear how far the competencies itself are present in the countries.

Conceptualizing mathematical achievement in the German national PISA option

Even stronger than PISA-international, the German national option capitalizes that an achievement test like PISA should map mathematics as comprehensively as possible. Therefore, typical ways of thinking and knowing in mathematics should be present in the test items. This model of the test tasks formed the basis (Fig. 2):

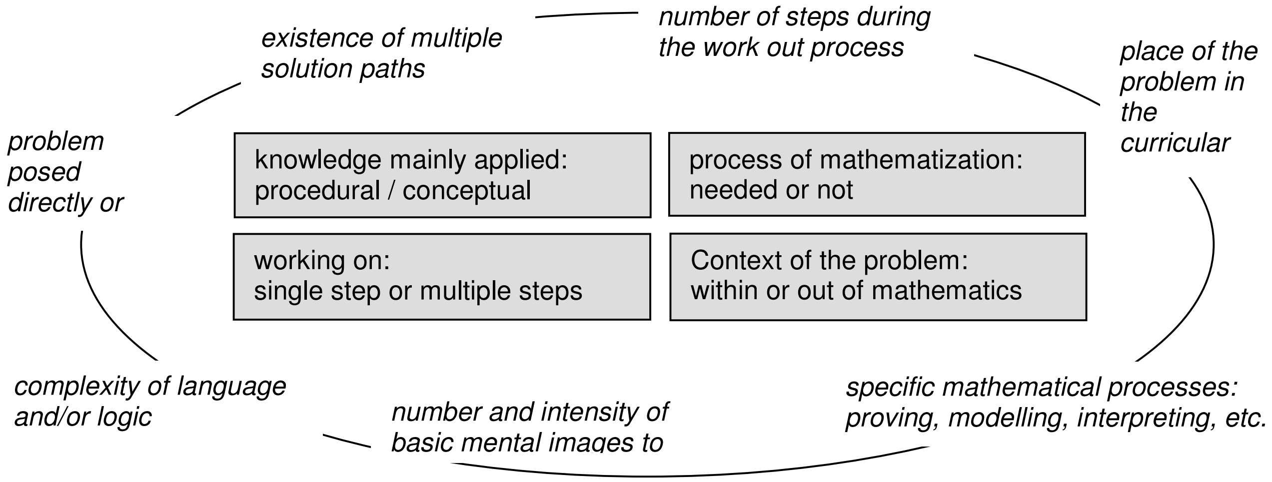


Figure 2. The model of a mathematical problem used in PISA-Germany: The core, and examples of characteristic features (Neubrand, 2004)

With the four basic features (the “core”) mathematical achievement can be structured by three “types of mathematical activities” (J. & M. Neubrand in Neubrand, 2004): (i) employing only techniques, (ii) modeling and problem solving activities using mathematical tools and procedures, (iii) modeling and problem solving activities calling for connections and using mathematical conceptions. From the cognitive and the mathematical point of view the three classes realize the full range of mathematical thinking, since one recognizes technical performance, and the essential modes of thinking, i.e., procedural vs. conceptual thinking (Hiebert, 1986).

ANALYTIC RESULTS OF PISA

The defined structures of mathematical achievement express themselves also in the data. But clearly, there remains a lot to do for further research.

International test: Countries show differences in content areas

Not surprisingly, test data show differences among countries in the performance on the defined content areas, the “overarching ideas” (OI). While the students in some countries behave quite uniformly over the content (e.g., Finland, Belgium), in some countries considerable differences appear. For example, Japan shows strengths in the OIs “change and relationships” and “space and shape”; and (relative) weaknesses in “quantity” and “uncertainty”. Germany shows weakness in the geometry and stochastics items. Results like these give hints what fields of mathematics should earn greater emphasis in curriculum and teaching. (See OECD, 2004 for details.)

Difficulty of a problem: A question of various features

Analyses done after PISA-2000 in Germany revealed some insight into the processes which make the solution of an item more difficult. However, as said in the beginning, due to the nature of the data, one can get information on mathematical learning and thinking in the whole, and not information of an individual's ways of thinking. Nevertheless, there are interesting results to obtain.

(a) *Not* the same features make a problem difficult in any of the three "types of mathematical activities" (J. & M. Neubrand in Neubrand, 2004). As a consequence, mathematic teaching cannot restrict itself to only a limited scope of mathematics.

(b) There is a competency specific to mathematics, that influences the difficulty of problems, even of those problems which call for modeling processes: the capability to use formalization as a tool (Cohors-Fresenborg & al. in Neubrand, 2004).

(c) Different didactical traditions and ways of teaching lead to different "inner structures" of mathematical achievement, made visible by different performance in the types of mathematical activities (J. & M. Neubrand in Neubrand, 2004).

References

Bybee, R.W. (1997). Towards an understanding of scientific literacy. In W. Gräber & C. Bolte (Ed.), *Scientific literacy. An international symposium* (S. 37-68). Kiel: Leibniz-Institut für die Pädagogik der Naturwissenschaften (IPN).

Freudenthal, H. (1975). Pupils' achievements internationally compared - The IEA. *Educational Studies in Mathematics 6,* 127 - 186.

Freudenthal, H. (1983). *Didactical phenomenology of mathematical structures*. Dordrecht: Reidel.

Hiebert J. (Ed.). *Conceptual and procedural knowledge: The case of mathematics*. Hillsdale, NJ: Lawrence Erlbaum.

Neubrand, M. (Ed.) (2004). *Mathematische Kompetenzen von Schülerinnen und Schülern in Deutschland: Vertiefende Analysen im Rahmen von PISA-2000*. Wiesbaden: VS - Verlag für Sozialwissenschaften.

Neubrand M. et al. (2001). Grundlagen der Ergänzung des internationalen PISA-Mathematik-Tests in der deutschen Zusatzerhebung. *Zentralblatt für Didaktik der Mathematik 33*, 33-45.

Niss, M. (2003). Mathematical competencies and the learning of mathematics: The Danish KOM Project. In A. Gagatsis & S. Papastavridis (Eds.), *3rd Mediterranean conference on mathematical education. Athens, Jan. 2003* (S. 115-124). Athens: Hellenic Math. Society.

OECD - Organisation for Economic Co-operation and Development (2001). *Knowledge and skills for life: First results from PISA 2000*. Paris: OECD.

OECD (2003): *The PISA 2003 Assessment Framework: Mathematics, reading, science and problem solving knowledge and skills*. Paris: OECD.

OECD (2004). *Learning for tomorrow's world: First results from PISA 2003*. Paris: OECD.

SOME RESULTS FROM THE PISA 2003 INTERNATIONAL ASSESSMENT OF MATHEMATICS LEARNING: WHAT MAKES ITEMS DIFFICULT FOR STUDENTS?

Carolyn Kieran

Université du Québec à Montréal

Département de Mathématiques

With the announcement of the 2003 PISA results in December 2004, we can now take a closer look at the released items and at how the 15-year-olds of the PISA assessment fared. A brief examination of item difficulty within the "change and relationship" scale suggests that we still know little about what it is that students find difficult in certain mathematical tasks.

MATHEMATICAL LITERACY IN PISA

The PISA concept of mathematical literacy is concerned with "the capacity of students to analyse, reason, and communicate effectively as they pose, solve and interpret mathematical problems in a variety of situations involving quantitative, spatial, probabilistic or other mathematical concepts" (OECD, 2004, p. 37). More precisely, mathematical literacy is defined as "an individual's capacity to identify and understand the role that mathematics plays in the world, to make well-founded judgments and to use and engage with mathematics in ways that meet the needs of that individual's life as a constructive, concerned and reflective citizen." The objective of the PISA 2003 assessment was "to obtain measures of the extent to which students presented with problems that are mainly set in real-world situations can activate their mathematical knowledge and competencies to solve such problems successfully" (OECD, 2004, p. 57).

HOW MATHEMATICAL LITERACY WAS MEASURED

Students' mathematics knowledge and skills were assessed according to three dimensions: mathematical content, the processes involved, and the situations in which problems are posed. Four content areas were assessed: shape and space, change and relationships, quantity, and uncertainty – roughly corresponding to geometry, algebra, arithmetic, and statistics and probability. The various processes assessed included: thinking and reasoning; argumentation; communication; modeling; problem posing and solving; representation; and using symbolic, formal, and technical language and operations. The competencies involved in these processes were clustered into the reproduction, connections, and reflection clusters. The situations assessed were of four types: personal, educational or occupational, public, and scientific. Assessment items were presented in a variety of formats from multiple choice to open-constructed responses.

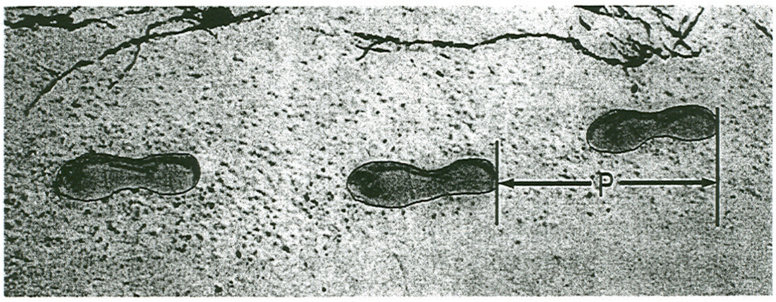
2005. In Chick, H. L. & Vincent, J. L. (Eds.). *Proceedings of the 29th Conference of the International Group for the Psychology of Mathematics Education*, Vol. 1, pp. 83-86. Melbourne: PME.

The PISA 2003 mathematics assessment set out to compare levels of student performance in each of the four content areas, with each area forming the basis of a separate scale. Each assessment item was associated with a point score on the scale according to its difficulty and each student was also assigned a point score on the same scale representing his or her estimated ability. Student scores in mathematics were grouped into six proficiency levels, representing groups of tasks of ascending difficulty, with Level 6 as the highest. The mathematics results are reported on four scales relating to the content areas mentioned above. As will be seen, an examination of item-difficulty within these scales reveals some surprises that, in turn, suggest that we, as researchers, may not really know what makes some mathematical tasks more difficult than others for students.

ITEM DIFFICULTIES FOR SAMPLE ITEMS FROM THE CHANGE AND RELATIONSHIP CONTENT AREA: THE WALKING UNIT

The Walking unit (OECD, 2004, p. 64) begins as follows:

WALKING



The picture shows the footprints of a man walking. The pacelength P is the distance between the rear of two consecutive footprints.

For men, the formula, $\frac{n}{P} = 140$, *gives an approximate relationship between n and P where:*

n = number of steps per minute, and
P = pacelength in metres.

Items 4 and 5 from this unit, along with the respective item difficulties and discussion of the competency demands, are presented in Figure 1. The level of difficulty ascribed to Item 4 is difficult to fathom: 611, which places it at Level 5 proficiency – a level at which only 15 % of OECD area students are considered likely to succeed. Yet, the item requires simply substituting n by 70 in the given formula $n/p = 140$, and then dividing 70 by 140. Its difficulty would seem closer to a Level 2 proficiency item, which according to the OECD report typically involves the "interpretation of a simple text that describes a simple algorithm and the application of that algorithm" (p. 69) – a task that 73% of OECD area students would be likely to solve. While students might attempt to solve the equation $70/p = 140$ by a cross-multiplication technique, they could also think about the task in terms of proportion ($70/p=140/1$, i.e., 70 is to 140 as p is to 1) or arithmetically in terms of division (70 divided by what number yields 140?).

WALKING

QUESTION 5

Bernard knows his pacelength is 0.80 metres. The formula applies to Bernard's walking.

Calculate Bernard's walking speed in metres per minute and in kilometres per hour. Show your working out.

Score 3 (723)

Answers which indicate correctly metres/minute (89.6) and km/hour (5.4). Errors due to rounding are acceptable.

Score 2 (666)

Answers which are incorrect or incomplete because:

- They were not multiplied by 0.80 to convert from steps per minute to metres per minute.
- They correctly showed the speed in metres per minute (89.6 metres per minute) but the conversion to kilometres per hour was incorrect or missing.
- They were based on the correct method (explicitly shown) but with other minor calculation error(s).
- They indicated only 5.4 km/hr, but not 89.6 metres per minute (intermediate calculations not shown).

Score 1 (605)

Answers which give *n* = 140 x .80 = 112 but no further working out is shown or incorrect working out from this point.

This open-constructed response item is situated in a personal context. The coding guide for this item provides for full credit, and two levels of partial credit. The item is about the relationship between the number of steps per minute and pacelength. It follows that it fits the change and relationships content area. The mathematical routine needed to solve the problem successfully is substitution in a simple formula (algebra), and carrying out a non-routine calculation. To solve the problem, students first calculate the number of steps per minute when the pace-length is given (0.8 m). This requires substitution into and manipulation of the expression: n/0.8 = 140 leading to: n = 140 x 0.8 which is 112 steps per minute. The next question asks for the speed in m/minute which involves converting the number of steps to a distance in metres: 112 x 0.80 = 89.6 metres; so his speed is 89.6 m/minute. The final step is to transform this speed into km/h - a more commonly used unit of speed. This involves relationships among units for conversions which is part of the measurement domain. Solving the problem also requires decoding and interpreting basic symbolic language, and handling expressions containing symbols and formulae. The problem, therefore, is rather a complex one involving formal algebraic expression and performing a sequence of different but connected calculations that need understanding of transforming formulas and units of measures. The lower level partial credit part of this item belongs to the connections competency cluster and with a difficulty of 605 score points it illustrates the top part of Level 4. The higher level of partial credit illustrates the upper part of Level 5, with a difficulty of 666 score points. Students who score the higher level of partial credit are able to go beyond finding the number of steps per minute, making progress towards converting this into the more standard units of speed asked for. However, their responses are either not entirely complete or not fully correct. Full credit for this item illustrates the upper part of Level 6, as it has a difficulty of 723 score points. Students who score full credit are able to complete the conversions and provide a correct answer in both of the requested units.

QUESTION 4

If the formula applies to Heiko's walking and Heiko takes 70 steps per minute, what is Heiko's pacelength? Show your work.

Score 1 (611)

Answers which indicate p = 0.5 m or p = 50 cm or p = 1/2 (unit not required).

This open-constructed response item is situated in a personal context. It has a difficulty of 611 score points, just 4 points beyond the boundary with Level 4. Everyone has seen his or her own footsteps printed in the sand at some moment in life, most likely without realising what kind of relations exist in the way these patterns are formed, although many students will have an intuitive feeling that if the pace-length increases, the number of steps per minute will decrease, other things equal. To reflect on and realise the embedded mathematics in such daily phenomena is part of acquiring mathematical literacy. The item is about this relationship: number of steps per minute and pacelength. It follows that it fits the change and relationships content area. The mathematical content could be described as belonging clearly to algebra. Students need to solve the problem successfully by substitution in a simple formula and carrying out a routine calculation: if n/p = 140, and n = 70, what is the value of p? The students need to carry out the actual calculation in order to get full credit. The competencies needed involve reproduction of practised knowledge, the performance of routine procedures, application of standard technical skills, manipulation of expressions containing symbols and formulae in standard form, and carrying out computations. Therefore the item belongs to the reproduction competency cluster. The item requires problem solving by making use of a formal algebraic expression. With this combination of competencies, and the real-world setting that students must handle, it illustrates Level 5, at the lower end.

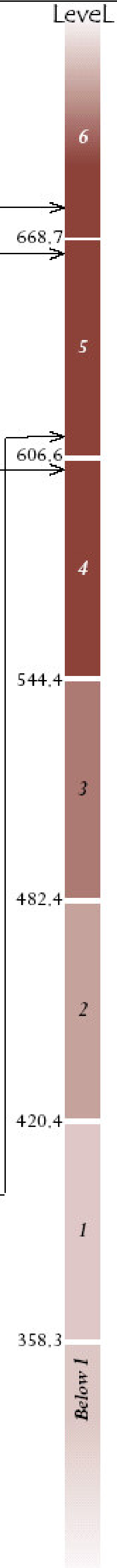


Figure 1. Items 4 and 5 of the Walking unit (OECD, 2004, p. 65)

Curiously, a response earning a partial score of 2 on the seemingly much more difficult Item 5 – at least more difficult from an *a priori* perspective – places it at Level 5 as well, albeit nearer the upper boundary of Level 5. But, it is not clear why a response that is deemed incomplete (and receives a score of 2) because the "112 steps per minute was not multiplied by .80 to convert it into metres per minute" – a conceptual demand that is at the core of Item 5 – is considered superior to the response "n = 140 x .80 = 112," which appropriately receives a partial score of 1. Notwithstanding the argument that could be made for both of these responses" to Item 5 receiving the same score of 1, the main issue concerns the conceptual demands that are inherent in Item 5, but which are lacking in Item 4. Why do students find Item 4 just about as difficult as Item 5?

While some might claim that the procedural demands of Item 4 (with the unknown in the position of denominator) explain to a certain extent why the difficulty level is 611, results from past research studies of equation-solving errors suggest that the difficulty level of this item should not be so high. For example, Carry, Lewis, and Bernard (1980) reported the following success rates for the solving of the given equations among students who covered a range from strong to very weak in algebra skills (e.g., 82%: 9(x+40) = 5(x+40); 76%: 1/3 = 1/x + 1/7; 76% 5/10 = (x-10)/(x+5)). In another study involving classes of 6th to 8th grade students, younger than those tested within PISA, Levin (1999) reported that 30% of the students correctly answered the following question by setting up and solving a proportion using cross multiplication (5/9=2/n): "On a certain map, the scale indicates that 5 cm represents the actual distance of 9 miles. Suppose the distance between two cities on this map measures 2 cm. Explain how you would fine the actual distance between the two cities." The equation was not unlike the one involved in Item 4; moreover, the students had to generate it themselves from the problem situation. One can only conclude that if the PISA results for this item and related symbolic representation items represent a trend with respect to students' abilities to handle rather simple symbolic forms, it is indeed a disturbing one. While Nathan and Koedinger (2000) noted that students find symbolically-presented problems more difficult than story problems and word-equation problems, the PISA results suggest that the discrepancy may be much greater than that reported by these researchers.

References

Carry, L. R., Lewis, C., & Bernard, J. E. (1980). *Psychology of equation solving – an information processing study* (Final Technical Report). Austin: The University of Texas at Austin.

Levin, S. W. (1999). Fractions and division: Research conceptualizations, textbook presentations, and student performances (doctoral dissertation, University of Chicago, 1998). *Dissertation Abstracts International 59*: 1089A.

Nathan, M. J., & Koedinger, K. R. (2000). Teachers' and researchers' beliefs about the development of algebraic reasoning. *Journal for Research in Mathematics Education, 31*, 168-190.

OECD (2004). *Learning for tomorrow's world – First results from PISA 2003*. Paris: OECD.

THE FOUNDATION AND SPECTACLE OF [THE LEANING TOWER OF] PISA

Julian Williams

University of Manchester

I raise questions about the construct and consequential validity of international studies such as PISA, and about PISA itself. I suggest a fault line runs through the construct 'mathematical literacy', but more importantly, through mathematics education generally, distinguishing 'Realistic' mathematics and 'Authentic' mathematics. I then ask questions about the political consequences of PISA in an audit culture in which targets beget processes. The aim to influence policy is identified with perceptible shifts in PISA discourse. As an instrument in the global education market, with its theft of critical theorists' rhetorical resources, is PISA re-invigorating the spectacle of international league tables?

INTRODUCTION

When I was a boy I visited Pisa and was very impressed by the leaning tower. I recall imagining that one could walk up the tower by spiralling up the outside, and was slightly disappointed by the reality. Later I learned that the inclination of the tower was annually increasing, and engineers feared that it would eventually fall over: they planned to strengthen the foundations to stop this, but did not straighten it. The tower has become a global spectacle, even featuring in jokes etc. (what did Big Ben say to the leaning tower of Pisa? I've got the time if you've got the inclination). The tower of Pisa became globally spectacular because of its dodgy foundations, not despite them.

I aim to raise questions about the validity of PISA (capitals now). First, I examine the construct validity of the foundation of PISA, 'mathematical literacy'; second, I address the consequential validity of PISA, its political consequences, as spectacle.

CONSTRUCT VALIDITY: THE FOUNDATION OF 'MATHEMATICAL LITERACY'

A confession: I find some of the items in PISA seductive, especially some of the Problem Solving items. In one the student is asked to diagnose a faulty bicycle pump, in another they are asked to evaluate some information on various drugs and select an appropriate pain-killer for 13 year old George, an asthmatic child with a sprained ankle. At face value, these represent a kind of functional 'literacy'. Turning to the mathematical literacy item used to explain the notion of mathematical modelling and mathematisation, one finds the park problem: where should a street-light be placed to illuminate a park? The park is mathematised as a triangle, the area lit is a circle, and the solution is the triangle's circumcentre (as long as the park is not obtuse-angled, explains PISA, 2003, p26).

2005. In Chick, H. L. & Vincent, J. L. (Eds.). *Proceedings of the 29th Conference of the International Group for the Psychology of Mathematics Education*, Vol. 1, pp. 87-90. Melbourne: PME.

I may be obtuse, but … our parks in English towns are usually locked at night, not lit. Perhaps they mean a car park? But … how many triangular car parks have you seen? I looked around and noticed that the lights were often on the perimeter of the park, which is in turn usually made up of rectangular blocks. For obvious reasons one might expect car parks to be rectangular, especially in modern countries where road systems are grid based. Perhaps one would find them in towns where road networks crystallised on the basis of clusters of medieval villages, like Chester or York? Both these towns are a long way from Manchester, so this prompted me to email my co-presenter from Japan and… he found one! (But where was the lighting?)…

Does the validity of Euclid really lie in such considerations? How has this come to be? I fantasise: Euclid, on a trip to visit the leaning tower, finds a triangular car park and noticing the light at the midpoint of one side… "Eureka: the circumcentre of a right-angled car park lies at the mid-point of the hypotenuse."

But Realistic Mathematics Education (RME) does not require that mathematics be authentic in this 'real' sense: only that the situation is realistic for the entry of the student into a world that begs to be mathematised. The validity test for RME then is (i) mathematical, rather than 'real' functionality, and (ii) empirical (i.e., do the students experience the problem in an intuitive way). Many of the PISA items appear to have this quality, at least to some degree.

I suggest that Realistic mathematics is primarily embedded in a scholastic, pedagogical activity system and is essentially embedded in the students' imaginary, experiential world: the object of activity is, in the end, to learn mathematics. On the other hand, I suggest Authentic mathematics is used as an instrument within an Activity System whose object is not essentially to learn mathematics, but to achieve some 'real' objective in a world outside mathematics. To become Authentically functional is to break out of the scholastic straitjacket and requires what Engestrom (e.g., Engestrom, 1987) called 'expansive' activity: at the very least, the class that 'plans a party' has to really have the party.

I prefer to think of this distinction as a fault line deep underneath the surface of the concept of 'mathematical literacy', rather than a dichotomy as such. Does this line undercut the mathematics education literature too?

And where is PISA? I'd say some of the best tasks are Realistic, but never quite Authentic (you would hope George's 15 year old literate elder sibling would think to ask a good pharmacist before deciding which painkiller to buy his asthmatic younger brother, wouldn't you? Sorry, 'code 0: no credit'). Could they be?

DISCOURSE AND SPECTACLE OF PISA: POLITICAL CONSEQUENCES

PISA has a political aim, that is, it seeks to influence policy. Thus on the one side, we have mathematics-literacy tasks, and the identification of learning outcomes for students. But on the other, we have summative statistics that 'count' for policy. This entails an interesting discursive shift. Initially, PISA (e.g., 2005) suggest that correlations display ‘associations’ that cannot be assumed to be ‘causal’, but later these associations become ‘influences’ that policy makers might find ‘interesting’. What is the difference for policy, i.e. what is the political difference between an influence and a cause? I see from the dictionary (OED) that an influence is in its original usage an astrological one, and later became political: it is essentially the exertion of an action whose mechanism is ‘unseen’ except in its effects.

This is significant because it determines to some extent the ‘consequences’ of PISA. How can policy makers be expected to read PISA’s results on the influence of SES or softer variables such as ‘school climate’ on learning outcomes? We see from the PISA-2000 study, for instance PISA (2005), that school climate explains significant variation in outcomes, but not that school climate is a possible ‘associate’ of high learning outcomes, and in Gill et al. (2002), associations with school background become ‘attributable’ to school background (p xvi).

Michael Power, who calls himself a professor of critical accountancy, has described the discourse of performativity in our audit culture (i.e., that of managing targets, league tables, performance-related reviews, etc.) as a Foucaultian discourse of (mis-) trust (Power, 1999). He and others have pointed to the way measurement constructs become targets and begin to dominate processes: thus as I write Prime Minister Tony Blair is felled by an angry electorate in debate on TV. He is accused of being responsible for the fact that in some doctors’ surgeries patients are not allowed to book an appointment to see their doctor more than 2 days ahead. Why? Because the government had introduced a performance target for the percentage of patients that have to wait more than 2 days. In vain he protests that this was not his intention! How will PISA measures be used, and what will be their unintended consequences?

Stronach (1999) in ‘Shouting theatre in a crowded fire’ construes the international tests and league table performance as a global spectacle, with ‘pupil warriors’ doing their sums for Britain. There’s England in the Premier league, 3 up on old rivals Germany, there’s a cluster of Confucian Pacific rim teams in the lead, but here comes Finland from nowhere suddenly challenging them. Is it social democracy or Nokia that ensures the team’s strength?

The association between PISA/TIMSS league tables and football competitions, the Olympics, horse races etc. is too strong to be denied, and ‘England’ in the tables becomes metonymically the nation and its education system per se, competing in the game with the rest of the world. One forgets that in fact the order of the names in the table are mostly not statistically significant, of course. What else is a table of scores actually for except to emphasise the ordinal at the expense of the complexity of the underlying data/reality? (That is intended to be a mathematically literate observation, if you like.)

The tabloid/redtop press are masters of this spectacle, but we all become implicated: government funding for research (at least in the UK) is increasingly predicated on ‘making a difference’ to learning outcomes in practice, and hence fulfilling political demands to become 'world class'. But how can world class be judged, except by international competition and league tables, and hence comparative measurement?

With what consequence? Is there no going back? Has the spectacle seduced our rationality? Pisa will always be the place with the leaning tower. While PISA challenges TIMSS by engaging with some 'literacy' rhetoric drawn from critical theory, the source of much that seems seductive in it, one reading of this move might be, as Gee et al. (1996) and others have suggested with 'fast capitalism', that the system steals critical theorists' rhetorical resources and emerges all the stronger for it.

So, where next? Could an expanded Authentic mathematics assessment emerge to confront the Realistic PISA, and in whose interest might that be?

References

(see www.education.man.ac.uk/lta/pme/PISA)

Engestrom, Y. (1987). *Learning by expanding: an activity-theoretical approach to developmental research*. Helsinki, Orienta-Konsultit.

Gee, J., Hull, G. and Lankshear, C. (1996). *The new work order: behind the language of the new capitalism*. Westview Press, Boulder, CO.

Gill, B., Dunn, M. & Goddard, E. (2002). *Student achievement in England*. London: HMSO.

PISA (2003). *The PISA 2003 assessment framework*. Paris: OECD.

PISA (2004). *Learning for tomorrow's world*. Paris: OECD.

PISA (2005). *School factors related to quality and equity*. Paris: OECD.

Power, M. (1999). *The audit explosion: rituals of verification*. Oxford: OUP.

Stronach, I. (1999). *Shouting theatre in a crowded fire*. Evaluation, 5(2), 173-193.

Afterword

Is the metaphoric association of Pisa and PISA – their foundations and their glorious spectacles – valid? If the consequence is that one is inclined to believe that there is a fault underlying 'mathematical literacy', I suggest yes. If one is led to think that this fault is implicated in the faux-spectacle of PISA, perhaps: the argument is that the act of global assessment becomes false *by virtue* of its becoming a political spectacle.

[Acknowledgements: to Google.com for suggesting the Pisa=PISA metaphor, and Ian Stronach for the introduction to this notion of spectacles.]