RESEARCH FORUM 3

Theme
Measuring mathematics learning and describing goals for systemic reform

Coordinators
Peter Sullivan and Dianne E Siemon

Reactor/Discussant
Stephen Lerman

Contributors
Margaret Brown, Mike Askew, Valerie Rhodes, Hazel Denvir, Esther Ranson & Dylan Wiliam, Barbara Clarke, Andrea McDonough & Peter Sullivan, Robert Wright & Peter Gould, Zemira Mevarech, Marja van den Heuvel-Panhuizen

RESEARCH FORUM

MEASURING MATHEMATICS LEARNING AND DESCRIBING GOALS FOR SYSTEMIC REFORM

Co-ordinators:

Peter Sullivan, La Trobe University, Australia
Dianne Siemon, RMIT University, Australia

FORUM OVERVIEW

In many countries, governments have initiated projects to evaluate and describe mathematical competence. Researchers within such projects make assumptions about what mathematics should be learned at particular ages, how learning goals are described, how learning is measured, and what findings may mean for curriculum and teaching. The forum is not critiquing whether this process is appropriate or necessary, but considers the solutions that researchers have sought and the rationale for those solutions.

Contributors to this forum describe their involvement in such projects, assumptions they made, the tensions they identified, and their resolutions of those tensions. Particular emphasis is on the contribution that such projects make to the mathematics education research.

Some of the decisions connected with this process include: What are the mathematical goals for students at particular ages? How can the learning of this mathematics be measured? What instruments and processes are appropriate for collecting evidence? What evidence would be characteristic of learning? Are there underlying assumptions about the way curriculum is described? Are there underlying assumptions about the way mathematics is taught? What are the tensions (e.g., suitability of instruments for all students, need for attention to have issues arising will be addressed) identified, and how are these reconciled?

MEASURING MATHEMATICS LEARNING AND DESCRIBING GOALS FOR SYSTEMIC REFORM: SOME ISSUES FOR CONSIDERATION

Stephen Lerman, South Bank University, London, UK

Large-scale projects to provide evidence for policy on developing mathematical competence are opportunities for mathematics education researchers to play a part in the improvement of students' learning of mathematics. There have been too many situations internationally where policy has proceeded without evaluative, descriptive or predictive research. Indeed in some countries educational researchers are deliberately ignored. It is most important, therefore, that such projects consider carefully as wide a range of issues concerning the quality of their project and its potential impact before commencement and as an on-going process throughout their work.

What follows is a non-exhaustive list of issues, expressed as questions, that I believe should be considered. These comprise methodological and ethical questions and one concerning engagement with the research community.

I invite the contributors to this Research Forum to respond to some or all of these issues, describing whether and how they feature in their projects. On the other hand you may have reservations about the necessity of responding to one or more of these issues. If so, I invite you to explain your point of view.

- Whose agenda is the project? Whose interests are served: students, parents, teachers, society, Government, the mathematics education community? Who owns the data? How will it be used?
- What are the theoretical frameworks (some of which may be implicit)? Are they drawn from psychological and/or sociological and/or other intellectual resources? What are the methodological justifications for the methods being used?
- What are the assumptions about the relationship between teaching and learning? What are the assumptions about the role of texts and other materials?
- What are the assumptions about the nature of curriculum?
- Where do the norms of student mathematical development come from (since both norm-referenced and criterion-referenced norms are actually norm-referenced)? Will tests be designed to measure actual development or proximal development (or both)?
- Does the sample size do justice to the research questions?
- Who benefits from the project, and how? Will there be any losers? Will the disadvantaged be further disadvantaged?
- How are insiders' perspectives taken into account (students, teachers, parents, etc.)?
- What arrangements are there for gaining, then reviewing and revising the consent of participants?
- How will the findings be critiqued within the community of mathematics education research (or other academic communities)? Will the project be discussed with the research community at the design stage, the interim findings and/or the final outcomes stages? Will the findings be expected to be generalisable internationally?

MEASURING PROGRESS IN NUMERACY LEARNING

Margaret Brown, Mike Askew, Valerie Rhodes, Hazel Denvir, Esther Ranson and Dylan Wiliam

King's College, University of London

LARGE SCALE RESEARCH AND POLICY

We focus on the longitudinal assessment strand of the Leverhulme Numeracy Research Programme in the UK between 1997 and 2002. A broader report is presented in a Plenary paper in this conference (Brown, 2002): we have tried to minimise the overlap in content.

Steve Lerman is right in his challenge to be concerned about large-scale assessment of mathematical competence; there is no doubt that the results of studies such as SIMS and TIMSS (Robitaille & Garden,1989; Reynolds & Farrell, 1996; Beaton et al., 1996; Mullis et al.,1997) have had a major influence on mathematics education in England, providing the justification for an ever-tightening centralised control of curriculum and teaching methods. Two English studies which one of us has previously managed (Concepts in Secondary Mathematics and Science (CSMS) (Hart(Ed.), 1981), and Graded Assessment in Mathematics(GAIM) (Brown (Ed.), 1992) has supplied data informing the structure and content of these reforms. While we regret some of the uses made of the results of these studies, these policies have nevertheless been developed by democratically elected governments and supported by the majority of teachers, parents and voters. We believe that it is part of the collective responsibility of mathematics education researchers in any country, and indeed internationally, to assist in the collection of reliable and valid large-scale data to inform international, national and local policies. We also of course regard critique as part of our collective responsibility and hence value large-scale data which can provide evidence on which rigorous critique of policy can be based (e.g. Brown *et al.*, 1998). Large-scale research can also often provide a basis for and/or evaluation of design and development work, which is a third essential element of the collective responsibility of mathematics educators.

Our support for large-scale work is by no means in opposition to arguing the importance of case-studies; indeed we believe that insight gained from successful small-scale work is not only valuable in itself but also an essential preliminary to large-scale studies. We do not therefore have a problem in being associated with both types of work in a complementary way, as in the Leverhulme Numeracy Research Programme and in many other studies in which we have been involved.

In the next sections, we will try to answer the most relevant of Steve's questions to our project. Sometimes the order is changed to assist the narrative; in some cases the pressure of space has forced rather briefer answers than we would like.

THE LEVERHULME NUMERACY RESEARCH PROGRAMME

1. Origins and Purposes

The Research Programme is funded by the Leverhulme Trust, which is independent of Government and distributes the interest on profits from manufacturing to support charitable work. The research agenda was partly that agreed by the Trust, advised by academics; they announced in 1996 that they would donate £1 million to a study of low achievement in primary schools focusing on literacy and/or numeracy. The Core Project, one of six projects in the Programme, was proposed *'to obtain large-scale longitudinal value-added data on numeracy, in order: to inform knowledge about the progression in pupils' learning of numeracy throughout the primary years; and to assess relative contributions to gains in numeracy of the different factors investigated in the programme.'*

The sample would be all the children from two different cohorts in 40 schools, ten in each of four local education authorities in different and varied areas of England. One cohort would start in Reception (age 4 to 5) and progress to Year 4 (age 8 to 9); the other would start in Year 4 and progress to Year 8 (age 12 to 13), although test data would be collected only in Years 2 to 7.

2. Theoretical Background

In the full research team there are about 16 researchers, although only two work full-time on the programme. In addition to mathematics educationists the team includes a cognitive psychologist and a social anthropologist, since an aim of the research programme is to examine primary school numeracy from multiple perspectives. Most of the group working on the large-scale assessment project are educationists who espouse a broadly social constructivist view of learning, involving progressive generalisation from situated cognition sometimes experienced within overlapping communities of practice (Wenger, 1998), and a dialectic both between the social and the individual, and between the cognitive and the affective. However our separate theoretical positions vary along these continuums.

There are clearly problems about the concept of criterion-referencing which assumes an all-or-nothing capability. The tests are only loosely criterion-referenced in the sense that most items can be related to specific cognitively based skills. Justification for this lies in the fact that the original form of the tests was as part of a diagnostic interview which was developed iteratively as part of a study identifying important steps in learning (Denvir & Brown, 1986). Any inferences from success in items to specific understandings is highly dangerous, but the analysis of associated sets of items will involve some inferences about learning which we recognise as very speculative. But such speculation lays the foundations for science.

As part of the study we have observed teaching in all the classrooms and adapted an analytic framework first used by Saxe (1991) which relates to a broadly activity theory approach focusing on discourse, tools, tasks, relationships and norms.

3. Methods

In Steve's terms the assessment is probably intended to be actual not proximate, although evidence suggests that items sometimes scaffold learning. We would all accept the crudeness of the method of assessing children's progress which we use, i.e. tests with questions read out by the teacher and answered in printed test booklets. It is constrained by the large size of the sample (about 1700 children in over 70 classes in each of the two cohorts) and the consequent need for brevity, simplicity and uniformity since we rely on teachers to administer the tests.

We have evidence through case-studies that a single child's test performance on a particular occasion is not particularly reliable; however our aim is not accurate assessment of individuals. First, we will make inferences about progression in learning of English children at a particular point in time. This will be based on changes in the proportion of the cohort over a particular time span who are able to obtain correct answer to a linked set of items. Thus most idiosyncratic variation in individual children will cancel out but sources of systematic bias will remain. We will however present as much data as possible for others to critique our conclusions.

For example, after a practice item where children discuss with their teacher a quick way of finding 30+21 if they know 30+20=50, they are given on their booklets that 86+57=146 and asked quickly to use this to work out answers to different questions. Different forms of this item were on the tests used from Year 2 to Year 6 (each year was tested in October and June), and the percentages successful were:

Table 1: Percentages successful on linked items (Year 2 & 3 data relating to Cohort 1 in 1999-2001 is in italic font; Year 4, 5 & 6 data relates to Cohort 2 in 1997-2000)

Given	*Year 2*		*Year 3*		Year 4		Year 5		Year 6	
86+57=146	*Oct*	*Jun*	*Oct*	*Jun*	Oct	Jun	Oct	Jun	Oct	Jun
87+57	*35*	*60*								
86+56	*17*	*45*	*60*	*71*	69	79				
57+86			*57*	*67*	66	78	81	80	92	91
85+57					40	56	66	68		
143-86			*15*	*36*	22	36	46	57	71	80
86+86+57+57					21	36	41	54	68	79
860+570	*1*	*5*	*7*	*21*	19	31	39	55	69	76

This information illustrates the second purpose of the testing which is to examine the effect of different factors which may influence numeracy attainment. We have related the average gains each class makes between October and June each year to a variety of other data on pupils, teachers, teaching, and schools. To remove the important prior attainment factor, the tests have been designed so that the average gains over this period are independent of the initial scores. Examples of the results on this analysis are given in the plenary paper (Brown, 2002).

Concerning curricular validity, there is little agreement internationally or indeed even in the UK (Brown et al., 1998) on the definition of numeracy. The tests have in fact been developed and researched over a long time-span. The original diagnostic interview was designed in the light of the research literature on development of knowledge and skills in number which seemed relevant to the 1980s curriculum in most English schools; this evolved and was evaluated in a PhD study (Denvir & Brown, 1986). This instrument has been updated as part of another study (Askew, Bibby & Brown, 2000) and is now published for teachers' use (Bibby & Denvir, 2002). A further study entailed the adaptation of the test for class use and validated it against the individual interview assessment (Denvir & Brown, 1987). Finally the test has been modified and trialled for use in three different age groups (Askew *et al.*, 1997) before being adapted and trialled for 6 age groups in this study (Brown *et al.*, 1996). In each of these adaptations a check was made that the content matched the primary number/numeracy curriculum at that time in England.

There are for practical reasons no long items of a problem-solving or investigative nature, but these are currently little used in teaching in England. An investigative item was abandoned when the analysis of the data revealed large inter-class differences in strategy which suggested that teacher influence was overwhelming. In addition to the much researched and refined nature of the tests we have other reasons for having confidence in the comparative validity and reliability of the tests:

• we have compared performances on specific items on seven successive occasions for our case-study sample of 30 children per cohort: this suggests that the results are generally consistent. The results are also generally consistent with our observations of these children's cognitive behaviour in their normal lessons:

• the overall performance of a whole cohort on successive occasions is consistent;

• the reliability measure on a single test occasion is high (Wiliam *et al.*, 1998);

• we have observed classes and interviewed all teachers each year; this enables us to investigate any results which seem to be inconsistent with classroom performances;

• teachers have been very satisfied with the content of the tests, except that some are concerned that some questions are too difficult and go beyond what has been taught. (This is unfortunate and may affect results for some children, but it is a necessary part of the test design to measure progression for children across the whole range; teachers are asked to explain to children before they start that they are not expected to be able to do all the questions as some are designed for older children.)

4. Consultation, dissemination and application

Headteachers were asked to consult staff before agreeing to take part (and almost all did). Over five years of intense national change only 3 schools out of 40 have withdrawn, due to new headteachers. Permission was sought from all parents.

The study has involved 45-minute interviews with all teachers (about 600 over 5 years, with less than five refusals), annual interviews with headteachers and mathematics co-ordinators, and 45-minute interviews and many informal conversations with the case-study sample of about 60 pupils (Years 3 and 6). In all interviews participants have been asked about their views of the tests and other aspects of the study. There was limited scope for changing the tests for that year group as the model was already set, but tests for other year groups could be and were slightly affected, as they were developed and trialled one year at a time.

Payment for teacher cover during the interviews has been provided so that schools and teachers did not suffer. Annual feedback on test results is provided to local authorities, headteachers and teachers, but names of schools, classes and children are not given in accordance with promises of confidentiality. At the same time we have provided some summary of aspects of the interim project findings. The majority of schools and teachers have voluntarily expressed their regret that the project is now finishing, so presumably they have found participation useful; several have cited specific uses of the feedback e.g. in self-evaluation or inspection.

We have presented papers for discussion and critique at both national and international research meetings, at the stage of design (e.g. Brown *et al.*, 1996) as well as at the results (Wiliam *et al.*,1998; Rhodes *et al.*,1998, Brown *et al.*, 2000; Brown, 2000). However we have generally avoided wide scale national publication of, and publicity for, interim results beyond the research community, as we have preferred to wait until possible effects were confirmed or contradicted in the data for successive years. We are contracted to Kluwer to write a series of four books.

In terms of our aims, the main output is in knowledge and understanding about primary numeracy in order to improve attainment; hence the main beneficiaries should be future generations of pupils and teachers. Specifically from the longitudinal study this relates to progression in learning and factors which affect it. On the basis of reactions to presentations we have given and articles we have published in professional journals (e.g. Brown, 1999) we believe that the large-scale results, illustrated by specific examples, will be of use and interest to individual teachers as well as to researchers, developers and policymakers. Apart from our own uses, both to achieve our aims and to exercise our right of policy critique, we will of course have little control over how the data is eventually used once it is in the public domain. But that is a necessary part of an open society.

References

Askew, M., Brown, M., Rhodes, V., Wiliam, D. & Johnson, D. (1997) *Effective Teachers of Numeracy*. London: King's College London, for the Teacher Training Agency.

Askew, M., Bibby, T. & Brown, M. (2000) *Raising Attainment in Primary Number Sense: from counting to strategy. (Research Paper RESO 1).* London: BEAM Educational.

Beaton, A.E., Mullis, I.V.S., Martin, M.O., Gonzalez, E.J., Kelly, D.L. and Smith, T.A. (1996) *Mathematics Achievement in the Middle School Years: IEA's Third Mathematics and Science Study*. Chestnut Hill, Massachusetts: Boston College.

Brown, M (Ed.) (1992) *Graded Assessment in Mathematics (GAIM).* Walton-on-Thames: Thomas Nelson.

Brown, M., Askew, M., Rhodes, V., Wiliam, D. & Johnson, D. (1996) Measuring standards of numeracy in England. Paper presented at the AERA Conference, New York, April, 1996.

Brown, M., Askew, M., Baker, D., Denvir, H., Millett, A. (1998) Is the National Numeracy Strategy research-based? *British Journal of Educational Studies, 46*(4), 362-385.

Brown, M. (1999) Is whole class teaching the answer? *Mathematics Teaching, 169,* 5-7.

Brown, M., Denvir, H., Rhodes, V., Askew, M., Wiliam, D. & Ranson, E. (2000) The effect of some classroom factors on Grade 3 pupil gains in the Leverhulme Numeracy Research Programme. In T. Nakahara & M. Koyama (Eds.) *Proceedings of the 24th Conference of the International Group for the Psychology of Mathematics Education*, 2, 121-128. Japan: Hiroshima University.

Brown, M. (2000) What kinds of teaching and what other factors accelerate primary pupils' progress in acquiring numeracy*? Proceedings of the Research Conference 2000, Brisbane*, 3-5. Melbourne: Australian Council for Educational Research.

Brown, M. (2002) Researching primary numeracy. In *Proceedings of the 26th Conference of the International Group for the Psychology of Mathematics Education.* University of East Anglia.

Denvir, B. & Brown, M. (1986) Understanding of number concepts in low attaining 7-9 year olds: parts 1 & 2. *Educational Studies in Mathematics, 17*, 15-36 and 143-164.

Denvir, B., & Brown, M. (1987) The feasibility of class-administered diagnostic assessment in primary mathematics. *Educational Research, 29(2),* 95-107.

Denvir, H. & Bibby, T. (2002) *Diagnostic Interviews in Number Sense.* London: BEAM Publications.

Hart, K. (Ed.) (1981) *Children's Understanding of Mathematics: 11-16.* London: John Murray.

Mullis, I.V.S., Martin, M.O., Beaton, A.E., Gonzalez, E.J., Kelly, D.L. & Smith, T.A. (1997) *Mathematics Achievement in the Primary School Years: IEA's Third Mathematics and Science Study.* Chestnut Hill, Massachusetts: Boston College.

Reynolds, D. & Farrell, S. (1996) *Worlds Apart? A review of international surveys of educational achievement involving England (Ofsted Reviews of Research Series).* London: HMSO.

Robitaille , D.F. and Garden, R.A. (1989)*The IEA Study of Mathematics II: contexts and outcomes of school mathematics.* Oxford: Pergamon Press.

Rhodes, V., Wiliam, D., Brown, M., Denvir, H. & Askew, M. (1998) Designing valid numeracy tests for the primary age range.

Saxe, G.B. (1991) *Culture and Cognitive Development: studies in mathematical understanding.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Wenger, E. (1998) *Communities of Practice: learning, meaning and identity.* Cambridge: Cambridge University Press.

Wiliam, D., Rhodes, V., Denvir, H., Brown, M., Ranson, E . & Askew, M. (1998) Analysing numeracy attainment and progress in a Year 4 cohort. Paper presented at the British Educational Research Association Conference, 1998. BERA EducatiON-line.

MEASURING AND DESCRIBING LEARNING: THE EARLY NUMERACY RESEARCH PROJECT

Barbara Clarke, Monash University, Australia
Andrea McDonough, Australian Catholic University
Peter Sullivan, La Trobe University, Australia

SYSTEMIC MOTIVATION

In the late 1980s, both levels of government in Australia collaborated in specifying the outcomes that could define learning (see Australian Education Council, 1991). While this type of curriculum specification and measurement was comprehensibly criticised (e.g., Ellerton & Clements, 1994) as constraining teaching and restricting achievement of students, it has also been argued that defining of curriculum goals in terms of standards and targets creates pressure on governments to provide resources for initiatives to ensure that the targets are met (e.g., Willis, 1995). It seems that both predictions have come true, although we focus on the latter here.

Governments have taken the standards seriously and initially invested heavily in literacy achievement for early years children with broad community support. Numeracy, which for all purposes at this level is used interchangeably with mathematics, received slightly lesser but still enthusiastic government and community support. Arising from this interest the relevant Victorian government department invited submissions for a large-scale research project, which was won by a team including these three authors. For this forum, there are two relevant motivations to consider: the motivation of the government department in soliciting the tenders and the motivation of the team who won the tender in pursuing their research goals.

For the government's part, they were keen to gather data to allow them to argue in competition against other sections of government for funding support for on-going professional development, school-based grants for co-ordination and resources, curriculum initiatives, and other support for teachers. Recognising that essentially different arms of government both within and outside education compete with each other for initiatives and support, the gathering of data which could justify support in one area as against another is central to the working of the bureaucracy. The government department was also keen to identify the content of professional development and resources which could be produced to support teaching.

From the researchers' perspectives, while the tender required us to serve the above goals, we had other aspirations. We sought insights that could describe student learning over time, perhaps identifying discontinuities that could be the focus of particular resource support, and articulating reasonable expectations for students in schools. We were also concerned about teachers and strategies for developing their confidence in teaching mathematics.

EARLY NUMERACY RESEARCH PROJECT

The Early Numeracy Research Project was initiated in Victoria following the success of the Early Literacy Research Project. The Early Literacy Research Project (Hill & Crévola, 1999) worked with 27 disadvantaged Victorian primary schools to bring about substantial improvements in early literacy outcomes. Part of this research involved the development of models and guidelines for teaching, assessment and additional support for young children learning to read. As a result of the research, Hill and Crévola offered a "general design for improving learning outcomes" (p. 122), which they believed had application in literacy, numeracy, and other curriculum areas.

The Early Numeracy Research Project (ENRP) was established in 1999 by the (then) Victorian Department of Education, with similar aims to those of the Early Literacy Research Project, but with a mathematics focus in the first three years of schooling.

The 35 trial schools were selected from approximately 400 schools who applied to participate. The trial schools were chosen to represent the range of Victorian schools, in terms of geographical location, socio-economic status, language background, school size and indigenous population. The 35 reference schools were matched from remaining Victorian primary schools, seeking the closest match taking into account the above variables.

The stated aims of the Early Numeracy Research Project were

- to assist schools to implement the key design elements as part of the school's numeracy program;
- to challenge teachers to explore their beliefs and understandings about how children develop their understanding of numeracy, and how this can be supported through the teaching program; and
- to evaluate the effect of the key design elements and the professional development program on student numeracy outcomes.

The central part of the data collection was an individual interview addressing nine domains of mathematics. In each domain, previous research was surveyed to identify key issues associated with developmental learning within those domains. These "growth points" were the basis of the quantitative scale used (Horne & Rowley, 2001). While a requirement of the data collection process, these growth points served a function of describing student thinking. It was intention of the research team to focus on student mathematical thinking and strategies. The nature of mathematical understanding was clearly implicit in the nature and form of the growth points (for more detailed discussion see Clarke, 2001). Essentially we believed that teachers can focus on a small number of big ideas in each of the domains and this can guide teaching, learning and assessment in way that can facilitate student learning. Ultimately we hoped that the description of the student achievement at this level would inform curriculum, both in scope and in sequence (see, for example, Sullivan, Clarke, Cheeseman, & Mulligan, 2001). The interview was used on 34 398 occasions with children during the ENRP, with 11 421 children being interviewed at least once. In addition 1152 and 868 interviews were conducted with Grade 3 and 4 children, giving a total of 36 426 interviews in the period 1999-2001.

The key recommendations from the student data overall were:

- The data considered individually and collectively highlighted the independence of the respective mathematics domains, and all domains should be part of the curriculum at each Grade level in the early years of schooling.
- The growth points seem to provide clear indications of standards for learning and targets for teaching at these levels, and teachers could use these to inform planning, teaching and assessment.
- By using the assessment interview, teachers can gain important insights into the learning of their individual children. The growth points then provide important information for teachers to allow them to adjust classroom experiences to address the particular learning needs of each student.
- In a number of the domains, barriers to student learning were identified. These were growth points that it seemed to take students significant time to achieve. Particular experiences should be planned by teachers to address the elements of the barrier explicitly.
- In each domain specific recommendations related to changes in the emphasis and specification of the Victorian were made. In some places serious deficiencies were identified.

The interview, of course, was only one part of a project. Indeed it could be argued that the most informative parts of the project are the intensive case studies conducted with both schools and individual teachers who appeared to have higher achievement of their students over the period of the research. Nevertheless the data from the interviews revealed important information about the sequence and rate at which young children learn those elements of numeracy that were part of the framework, and the results can be used to inform curriculum and resource development, and pre- and in-service teacher education.

We also conducted extensive professional development and gained significant insights into the way that such professional development changes teachers' beliefs and orientation to teaching, as discussed in the next section.

The impact of measuring and describing learning on the participating teachers

A range of data both qualitative and quantitative data was collected from the teachers. There were approximately 250 teachers involved in the professional development program throughout the three years of the project. Support provided to the teachers within the project can be considered as three "opportunities":

1. Opportunity to develop *knowledge of children's learning of mathematics* through a framework of key growth points and a task-based interview that provides insights into the range of strategies children understand and use, as well as an understanding of more sophisticated strategies that the children can be encouraged to work towards.
2. Opportunity to develop *knowledge of students* through one-to-one, mathematically-focused interaction during the interview.
3. Opportunities to explore a range of classroom strategies, in an environment of collegiality, to build upon children's mathematical understanding, through purposeful and informed teaching.

The ENRP growth points were designed to provide teachers with a framework to describe their children's learning in mathematics and the interview measured this learning. The perspective of the research team was of the teachers as co-researchers and teachers as having responsibility for the making of curriculum (Clarke, Clarke, & Sullivan, 1996).

It seems that the growth points provided not only a way to discuss what the children already know but the direction to move. Teachers in the ENRP had a clearer picture of the typical trajectories of student learning (Carpenter & Lehrer, 1999), and can recognise landmarks of understanding in individuals. Such a picture guided the decisions they make, in planning and in classroom interactions, as their knowledge of the understanding of individuals informs their practice. This increased knowledge in turn seemed to build the confidence of the teachers:

> I used to avoid maths when something had to be missed, now I can't find the time to do enough! I'm more confident and I enjoy it more. I'm more flexible and more responsive to the children. I have a better understanding of how children think and reason, due to the assessment. This has impacted on what/how I teach.
>
> The main thing that has changed is my confidence in my maths teaching, because I have more knowledge of how children learn maths, what they should know and some ideas of how to get them there. My lessons are more varied and fun now.
>
> It has given me a greater understanding of why and how I teach maths, therefore increasing my confidence in my maths teaching.

While ultimately the recommendations of the project reflect the statement of the growth points which in turn informed the assessment items, the findings can inform curriculum development, teacher education, and further research. The data were collected from large numbers of students over three years with a carefully selected, stratified sample, to ensure all sections of the community were represented, and can be taken as an accurate representation of the potential of all students to answer the questions asked at these levels. The teachers have been empowered through increased knowledge and a belief in their own teaching and we would argue –the freedom to make their own professional decisions. The project sees the provision of similar professional development for other teachers as a key recommendation for government action.

References:

Australian Education Council (1991). *National Statement on Mathematics for Australian Schools*, Curriculum Corporation, Victoria.

Carpenter, T., & Lehrer, R. (1999). Teaching and learning mathematics with understanding. In E. Fennema, & T. Romberg (Eds.), *Mathematics classrooms that promote understanding,* Lawrence Erlbaum, New Jersey.

Clarke, B., Clarke, D., & Sullivan, P. (1996). The mathematics teacher and curriculum development. In Alan Bishop et al. (Eds.) *International Handbook of Mathematics Education* (pp.1207-1235), Kluwer, Dordrecht.

Clarke, D. (2001). Understanding, Assessing and Developing Young Children's Mathematical Thinking: Research as a Powerful Tool for Professional Growth. 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*, Mathematics Education Research Group of Australasia, ???

Ellerton, N. F., & Clements, M. A. (1994). *The National Curriculum Debacle,* Meridian Press, Perth.

Hill, P. W., & Crévola, C. A. (1999). The role of standards in educational reform for the 21st century. In D. D. Marsh (Ed.), *Preparing our schools for the 21st century* (Association of Supervision and Curriculum Development Yearbook, pp. 117-142). Alexandria, VA: ASCD.

Horne, M, & Rowley, G. (2001). Measuring growth in early numeracy: Creation of interval scales to monitor development. In M. van den Heuvel-Panhuizen (Ed.), *Proceedings of the 25th conference of the International Group for the Psychology of Mathematics Education* (pp. 3-161 – 3-169), Utrecht, July.

Sullivan, P., Clarke, D., Cheeseman, J., & Mulligan, J. (2001). Moving beyond physical models in learning multiplicative reasoning. In M. van den Heuvel-Panhuizen (Ed.), *Proceedings of the 25th conference of the International Group for the Psychology of Mathematics Education* (pp. 4.233-241), Utrecht, July.

Willis, S. (1995). Social Justice and the Mathematics Profile. In C. Collins (Ed.) *Curriculum Stocktake, Evaluating Curriculum Change*, Australian College of Education, ACT.

LARGE-SCALE STUDIES IN MATHEMATICS EDUCATION: A SEARCH FOR IDENTITY

Zemira R. Mevarech

School of Education, Bar-Ilan University, Israel

International studies, such as TIMSS and PISA, consider how well different countries educate their students. These large-scale studies also highlight associations between student background characteristics and student performance. The important question of these studies is not the rating of the countries. The important questions had to do with attempting to discern what it is about the instructional practices, school policy, the curriculum, and the psycho-sociological environment in certain countries that result in higher levels of student achievement (Willms, 2001).

Along with international studies, many countries conduct also national surveys on students' mathematical achievements. The main purpose of these surveys is to provide insights into the factors that influence the development of mathematical knowledge, skills, and competencies, how these factors interact, and the implications of the findings for policy making. As part of this process, researchers seek to measure and describe the relationships between the "inputs" and the "outputs" as well as between the intended, implemented, and attained curriculum.

Who are the audience of large-scale studies? Whose interests they intend to serve, and how? Whose interests are indeed served, and how? By supporting a shift in policy focus from educational inputs to learning outcomes, large-scale studies aim to provide insights into the factors that contribute to the development of competencies and into how these factors operate. Undoubtedly, such studies seek to combine the P and the E in PME. By that, these studies intend to assist policy makers, as well as mathematics educators and the community in large.

Yet, the ambitious scope of these studies and their tremendous costs raise the issue of the extent to which large scale projects indeed play a part in the improvement of students' learning of mathematics. Repeatedly, PME researchers raise the question of the extent to which such projects can assist educational systems in developing students' mathematical competencies: How can the findings serve as a basis for changing classroom practice? How can they assist in developing mathematics curriculum? Or, how can they set up the infrastructure for effective professional development? The present study addresses some of these questions.

Do large scale projects play a part in the improvement of students' learning of mathematics?

Many think that research has no impact on policy: policy makers do not get assistance from researchers, and if they get it, it is often either too late, or too "theoretical", or too vague. Researchers, on the other hand, have the feelings that policy is not based on empirical findings. The frustration is expressed by many researchers. Husen (1968), for example, claimed that "in retrospective, I am completely aware of the failure of research to influence reforms in schools. One may ask if the policy would be different if the research would not be conducted on the first place." Ten years later, Malmquist and Grundin (in Mevarech & Blass, 1999) argued that "those who are responsible of the educational policy do not expect research to give answers to the ongoing problems that the educational system faces ... policy makers tend to consider research as a marginal activity." Recently, Robitaille et al (2000) summarize the impact of TIMSS by saying: "the results (of TIMSS) seem, at least until now, to have made a greater impact on the general public and in political circles through the mass media than they have on educators."

The impact of national and international studies on the mathematics education research community is also far from being significant. So far, large-scale studies have contributed very little to the development of new theories or to clarifying the "how's" questions regarding teaching, learning, or curriculum development. Gravemeijer (in Arcavi, 2000) explained: "the question with which the researchers are struggling is... how to design instructional activities that (a) link up the informal situated knowledge of the students and (b) enable them to develop more sophisticated, abstract, formal knowledge, while (c) complying with the basic principle of intellectual autonomy" (pp. 278-279). Maybe, large-scale studies are not appropriate for addressing such questions. Probably, studies that are based on qualitative methods or small scales are better able to answer this kind of questions than large-scale studies such as TIMSS, PISA, or national surveys. Why such ambitious studies have (so far) only small impact on policy, even though such studies are frequently cited in the media and often funded by institutions that make the policy? To answer this question we have to consider another series of questions: Whose agenda is the project? Who benefits from the project, and how: students, parents, teachers, society, Government, the mathematics education community? Are there any losers? Unfortunately, in many countries, Israel included, TIMSS findings are not on the agenda of the educational system. Even though the Ministry of Education funded the project, the educational system does not consider itself obliged to the findings. Several reasons may explain it. First, those who made the decisions felt threaten by the findings and thus defend themselves by ignoring it. Second, the immediate implications of the findings are not very clear – the studies did not identify the reasons for the low achievements, nor did they describe the factors that contribute to the good performance in high achieving countries. Third, ironic as it sounds, researchers may also be blamed (explicitly or implicitly) for the outcomes particularly in those countries where researchers are involved in curriculum development. Finally, given the assumption that there is a relationship between the intended, implemented, and attained curriculum, even teachers and principals may be criticized (implicitly or explicitly) for not being effective. Thus, even though students, parent, teachers, and the society in large can benefit from the study, the list of losers is not shorter.

There is, however, another option. The usefulness of large scale studies lies in identifying the factors that contribute to mathematics education. As indicated by Robittaille (2000): "educational systems can learn from one another. They can learn that different approaches to common concerns -- such as the streaming of students by some measure of ability are taken in different countries. They can study the relative success and efficacy of those different approaches and then make decisions about what might work in their setting." (p. 169). The following is an example that shows the usefulness of international studies beyond what can be achieved by any other kind of research.

Mathematics Education, Resource Allocation, and Equity Policy

Educational inequality remains a major challenge to policy makers as well as to researchers. There is a long debate in the educational community regarding the duality between excellency and equality in education. Whereas many believe that the role of society is to reduce inequality even if it may come on the expanse of excellency, others argue that the role of schools is to enable each student to fulfil his/her potential and attain high levels of achievements. To address the issue of the duality between excellency and equality, we investigated the relationship between mean achievement in mathematics and measures of inequality in mathematics performance across countries by using TIMSS-1999 data. Inequality in mathematics performance was defined as the gap in achievement between the 5th and 95th percentiles within each country. The analysis indicated that the correlation between mean achievement and inequality is -.75. In other words, countries in which mean performance is high the level of inequality is low and vice versa. Further analyses showed that countries that improved performance between 1995 and 1999 reduced inequality. It is interesting to note that analyses of economical data from 67 countries reveal similar findings regarding the strong negative relationships between growth and inequality. Additional analyses indicate that the strong negative correlation between achievement and inequality is not a function of investment. These findings led us to examine the contributions of social capital or human capital to excellency and inequality. (More information will be reported at the conference.)

This finding along with additional analyses explaining the factors that contribute to the high correlation reported above has clear indications for policy makers and for mathematics education community: schools can make a difference, curriculum can make a difference, and resource allocation can make a difference. This finding calls researchers in mathematics education to provide policy makers models that enhance mathematical performance along with reducing inequality. Examples of such models are those which utilize metacognitive instruction in mathematics education (e.g., Mevarech & Kramarski, 1997). To my knowledge, several countries have started to work in this direction.

Large-Scale Projects Looking for Identity

Arcavi (2000) distinguishes between two kinds of research in mathematics education: theory-driven and problem-driven. Within problem-driven research, he identifies three kinds of research: interesting/puzzling behaviors, a curriculum or a practice, and didactic opportunities. Large-scale projects do not fit any of these categories. It seems that a new category is suitable for large-scale studies. I suggest "checkup-driven research". (I prefer the term knowledge-driven research, but tend to believe that all research is knowledge driven.) By checkup-driven research I refer to a systematic inquiry into a subject, in our case -- mathematics education, designed to uncover pertinent information about mathematics education and the educational system in large. Checkup-driven research is more than comparing countries' curricula, or analyzing cost-effectiveness between and within countries. It is also more than explaining students' puzzling behaviors, didactic opportunities, or practices. Checkup-driven research has to provide analytical descriptions of the system, usually with respect to certain possible causal connections. [So far, TIMSS, PISA and national surveys provide reliable descriptive information, but their causality part is weak.] Conceiving large-scale studies as a checkup-driven research enables us to develop appropriate expectations of such studies: their main purpose is to provide ongoing evidence on the development of mathematics competencies, and to suggest follow-up research whenever a problem is discovered.

Like medical check-ups, also educational checkups have unique characteristics and methodologies that distinguish them from theory-driven or problem-driven research. Of these characteristics, perhaps the most important ones are those in which evaluation is viewed as a process of identifying and collecting information to assist decision-makers in planning knowledge-based policy, and researchers in understanding the macro picture of mathematics education. Since large-scale research is a relatively new kind of research, its main concern is now setting up an original framework, and developing its own concepts and methods that would satisfy three criteria: relevance to the observable phenomena, exhausitivity with our understanding of mathematical thinking, learning, and teaching, and consistency of the concepts and methodology within the theoretical framework. Large-scale research should include, therefore, the methodological requirements of sharing knowledge, reflections, and curriculum design. These comprise methodological and ethical questions and call for larger involvement of the mathematics education research community, and other academic communities at all stages of the research. Prior arrangements such as these may eventually lead to knowledge-based policy making and may serve as a springboard for changing curriculum, classroom practice, and even national policy in attempts to enhance mathematical education. These issues will be further elaborated in the forum.

References

Arcavi, A. (2000). Problem-driven research in mathematics education. *Journal of Mathematical Behavior, 19*, 141-173.

Husen, T. (1968). Educational research and the state. In: Educational Research and Policy Making. NFER in England and Wales. A special paper read on the occasion of the 21st annual general meeting of NFER.

Mevarech. Z. R., & Blass, N. (1999). Science, research, and policy making in educational systems. In E. Peled (Ed.). *Fifty Years to Israel Educational System* (Hebrew) (pp. 441-453).

Mevarech, Z. R. & Kramarski, B. (1997). IMPROVE: A multidimensional method for teaching mathematics in heterogeneous classrooms. *American Educational Research Journal, 34,* 365-394.

Willms, D. (2001). What makes a difference to PISA results: some indications for policy. In: *OECD: Knowledge and skills for life: First results from the OECD programme for international student assessment* (PISA) 2000.

FROM CORE GOALS TO LEARNING-TEACHING TRAJECTORIES AS A GUIDE FOR TEACHING PRIMARY-SCHOOL MATHEMATICS IN THE NETHERLANDS

Marja van den Heuvel-Panhuizen

Freudenthal Institute, Utrecht University, The Netherlands

This paper is dedicated to the Dutch TAL project. In this project learning-teaching trajectories for primary school mathematics are being developed. Regarding the title of this Research Forum, the focus of this paper is rather on "describing goals" than on "measuring learning". Different from an approach in which large-scale assessment projects are carried out to provide evidence for policy on developing mathematical competence, the TAL project can be characterized as a didactical-phenomenological approach in which the nature of mathematical learning content, its sequencing over the grades and the way it is taught in primary school, are reviewed and overhauled.

INTRODUCTION

The nineties can be seen as the decade of standards. In many countries, at governmental level, decisions are made about what schools should teach their students. Examples of these standards are the NCTM Standards in the United States, the National Curriculum and Numeracy Project in the United Kingdom, and the Learning Framework in Number in Australia. This paper deals with the Dutch chapter of this trend.

Compared to many other countries, the Netherlands do not have centralized decision making for the primary school mathematics curriculum (see Mullis et al., 1997). Nevertheless — or probably one should say thanks to this — the mathematical content taught in primary schools does not differ much between schools. In general, all teachers follow roughly the same curriculum. A very important reason for this is the existence of a qualitatively high-standing base of commercially published mathematics teaching methods. Even the implementation of the Dutch reform of mathematics education has in fact been achieved voluntarily through these methods. This rather informal implementation mechanism is more or less the result of the existence a strong community of researchers and developers of mathematics education in the Netherlands. This provides a strong driving and regulating force for reform.

THE CORE GOALS FOR PRIMARY-SCHOOL MATHEMATICS CURRICULUM

Until recently, there was no real interference from the Dutch government regarding the content of educational programs. There only was a general law containing a list of subjects to be taught. In 1993, however, government policy changed and the

Ministry of Education came up with a list of 23 attainment targets for the end of primary school[1], called "core goals". The goals are split into six domains, including general abilities, written algorithms, ratio and percentage, fractions, measurement, and geometry. Table 1 shows the goals for the first two domains.

Table 1: Part of the core goals for Dutch primary school students in mathematics

By the end of primary school, the students (at age twelve) ...		
General abilities	1	Can count forward and backward with changing units
	2	Can do addition tables and multiplication tables up to ten
	3	Can do easy mental-arithmetic problems in a quick way with insight in the operations
	4	Can estimate by determining the answer globally, also with fractions and decimals
	5	Have insight into the structure of whole numbers and the place-value system of decimals
	6	Can use the calculator with insight
	7	Can convert simple problems which are not presented in a mathematical way into a mathematical problem
Written algorithms	8	Can apply the standard algorithms, or variations of these, to the basic operations, of addition, subtraction, multiplication and division in simple context situations

Compared to goal descriptions and programs from other countries this list is a very simple one. It means that there is a lot of freedom in interpreting the goals. At the same time, however, such a list does not give much support for educational decision making.

In the years since 1993, there have been discussions about these core goals (see De Wit, 1997). Almost everybody agreed that they can never be sufficient to support improvements in classroom practice or to control the outcome of education.

LEARNING-TEACHING TRAJECTORIES — A NEW FACTOR FOR MACRO-DIDACTIC TRACKING

For several years it was unclear which direction would be chosen for improving the core goals: either providing a more detailed list of goals for each grade, expressed in operationalized terms, or a description which supports teaching rather than pure testing. In 1997, the Dutch Ministry of Education tentatively opted for the latter and asked the Freudenthal Institute to work out the description for mathematics. In September 1997, this decision resulted in the start of the TAL Project[2].

Aim of the TAL Project

The aim of the TAL Project is to develop learning-teaching trajectories for all domains of the primary-school mathematics curriculum. In total three learning-teaching trajectories will be developed: a trajectory for whole number calculation, one for measurement and geometry, and one for fractions, decimals and percentages.

The project started with the development of a learning-teaching trajectory for whole-number calculation. This first trajectory description for the lower grades (including K1, K2, and grades 1 and 2) was published in November 1998. The definitive version was released a year later. The following year the whole-number trajectory for the higher grades of primary school (including grades 3 through 6) was published. In 2001, both learning-teaching trajectories were translated in English and published together in one book (Van den Heuvel-Panhuizen, 2001).

In 1999, a start was made on the development of a learning-teaching trajectory for measurement and geometry. This will be finished in the end of 2002.

What is meant by a learning-teaching trajectory?

Giving the teachers a pointed overview of how children's mathematical understanding can develop from K1 through grade 6 and of how education can contribute to this development is the main purpose of this alternative to the traditional focus on strictly operationalized goals as the most powerful engine for enhancing classroom practice. In no way, however, is the trajectory meant as a recipe book. It is, rather, intended to provide teachers with a "*mental educational map*" which can help them, if necessary, to make adjustments to the textbook.

Although a learning-teaching trajectory puts the learning process in line, it should not be seen as a linear and singular step-by-step regime in which each step is necessarily and inexorably followed by the next one. A learning-teaching trajectory should be seen as being broader than a single track and should have a particular bandwidth. It should do justice to differences in learning processes between individual students and to the different levels at which children master particular skills and concepts.

A new educational phenomenon

Compared to the goal descriptions that were traditionally supposed to guide education and support educational decision making, the learning-teaching trajectory as it is worked out in the TAL Project has some new elements that makes it a new educational phenomenon.

First of all, the trajectory is more than an assembled collection of the attainment targets of all the different grades. Instead of a checklist of isolated abilities, the trajectory makes clear how the abilities are built up in connection with each other. It shows what is coming earlier and what is coming later. In other words, the most important characteristic of the learning-teaching trajectory is its *longitudinal perspective* which has a long history in the Dutch didactical (subject-matter connected) approach to mathematics education.

A second characteristic is its *double perspective of attainment targets and teaching methods*. The learning-teaching trajectory does not only describe the landmarks in student learning that can be recognized en route, but it also portrays the key activities in teaching that lead to these landmarks.

The third feature is its *inherent coherence, based on the distinction of levels*. The description makes it clear that what is learned in one stage, is understood and performed on a higher level in a following stage. A recurring pattern of interlocking transitions to a higher level forms the connecting element in the trajectory. It is this level characteristic of learning processes, which is also a constitutive element of the Dutch approach to mathematics education, that brings longitudinal coherence into the learning-teaching trajectory. Another crucial implication of this level characteristic is that students can understand something on different levels. In other words, they can work on the same problems without being on the same level of understanding. The distinction of levels in understanding, which can have different appearances for different sub-domains within the whole number strand, is very fruitful for working on the progress of children's understanding. It offers footholds for stimulating this progress.

The fourth attribute of the TAL learning-teaching trajectory is the new description format that has been chosen for it. The description is not a simple list of skills and insights to be achieved, nor a strict formulation of behavioral parameters that can be tested directly. Instead, *a sketchy and narrative description, completed with many examples*, of the continued development that takes place in the teaching-learning process is given.

Development of the TAL learning-teaching trajectories

In the development of the TAL learning-teaching trajectories "didactical phenomenological analyses" – as Freudenthal (1983) called them – play a crucial role. These analyses reveal what kind of mathematics is worthwhile to learn and which actual phenomena can offer possibilities to develop the intended mathematical knowledge and understanding. Important is that one tries to discover how students can come into contact with these phenomena, and how they appear to the students. This means that problems and problem situations that give students opportunities to develop insight in mathematical concepts and strategies must be identified. Therefore a team, containing all kinds of specialisms in primary school mathematics, has been formed. The group contains experience in research and development of mathematics education, assessment, teacher educating, teacher advice, and teaching mathematics in primary school. The core of the work is formed by the (almost) weekly discussions in the project team, for which input comes from a variety of sources: analyses of textbook series, analyses of research literature, investigations in classrooms, and extensive consultations of experts in mathematics education. An earlier example of such an approach, but aimed at finding the long term learning process for the domain of ratio, can be found in Streefland (1984/1985).

THE TAL TRAJECTORY FOR WHOLE NUMBER CALCULATION

In the TAL trajectory for calculation with whole numbers (see Figure 1), calculation is interpreted in a broad sense, including number knowledge, number sense, mental arithmetic, estimation and algorithms. The trajectory description gives an overview of how all these number elements are related to each other.

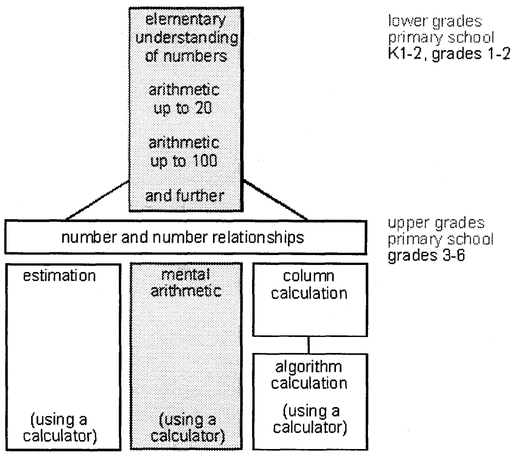

Figure 1: The TAL learning-teaching trajectory for whole number calculation in primary school.

The scheme reflects that the students gradually come from a non-differentiated way of counting-and-calculating to calculations in more specialized formats that fit particular kinds of problems in a particular number domain. Mental arithmetic is considered to play a central role in whole number calculation. It is seen as an elaboration of the arithmetic work that is rooted in the lower grades and forms the backbone in the upper grades.

ESTIMATION AS AN EXAMPLE

New in this trajectory is also the proposed didactics for estimation. Although estimation is now widely acknowledged as an important goal of mathematics education, in most textbooks a framework for how to learn to estimate is lacking. The textbooks at most only contain some problems on estimation, but doing some estimation problems from time to time is not enough to develop real understanding in how an estimation works, and it is not sufficient to comprehend what is 'allowed' and what is not when estimating.

At the Research Forum an overview will be presented of the TAL trajectory that has been developed for the domain of estimation. The proposed sequenced teaching structure with the intermediate attainment targets will be compared with the findings of the PPON, a large-scale Dutch assessment of students' performance in school subjects (Janssen et al., 1999).

Mathematics education researchers should play a part in the improvement of students' learning of mathematics. The question however is how? Providing empirical evidence about student performance is certainly important input, but will students' scores be sufficient to answer the core question about the goals te be achieved?

Notes

1. In the Netherlands, primary school is meant for students of ages 4 to 12 and includes eight grade classes. The first two are kindergarten classes.
2. The TAL Project is carried out by the Freudenthal Institute and the SLO (the Dutch Institute for Curriculum Development), in collaboration with CED (school advisory center for the city of Rotterdam). TAL is a Dutch abbreviation and stands for Intermediate Goals Annex Learning-Teaching Trajectories. Since the beginning of the project the following people have contributed to the development of the learning-teaching trajectory: Joop Bokhove (FI), Jan van den Brink (FI), Arlette Buter (FI/CED), Kees Buys (SLO), Nico Eigenhuis (CED), Erica de Goeij (FI), Marja van den Heuvel-Panhuizen (co-ordinator) (FI), Jan Hochstenbach (FI), Christien Janssen (FI), Julie Menne (FI), Ed de Moor (FI), Jo Nelissen (FI), Anneke Noteboom (FI), Markus Nijmeijer (FI), Adri Treffers (FI), Ans Veltman (FI), Jantina Verwaal (FI). In total the size of the TAL Team was equivalent to three fully employed persons.

References

De Wit, C.N.M. (1997). *Over tussendoelen gesproken. Tussendoelen als component van leerlijnen.* 's-Hertogenbosch: KPC Onderwijs Innovatie Centrum.

Freudenthal, H. (1983). *Didactical fenomenology of mathematical structures.* Dordrecht: Reidel Publishing Company.

Janssen, J., van der Schoot, F., Hemker, B. & Verhelst, N. (1999). *Balans van het reken-wiskundeonderwijs aan het einde van de basisschool 3.* Arnhem: Cito.

Mullis, I.V.S., Martin, M.O., Beaton, A.E., Gonzalez, E.J., Kelly, D.L., and Smith, T.A. (1997). *Mathematics achievement in the primary school years: IEA's Third International Mathematics and Science Study (TIMSS).* Chestnut Hill, MA: Boston College.

Streefland, L. (1984/1985). Search for the roots of ratio: Some thoughts on the long term learning process (Towards... a theory). *Educational Studies in Mathematics, 15*, 327-348 (Part I), *16*, 75-94 (Part II).

Van den Heuvel-Panhuizen, M. (Ed.) (2001). *Children learn mathematics.* Utrecht/Enschede: Freudenthal Institute, Utrecht University / SLO.

USING A LEARNING FRAMEWORK TO DOCUMENT STUDENTS' PROGRESS IN MATHEMATICS IN A LARGE SCHOOL SYSTEM

Robert (Bob) Wright, Southern Cross University, NSW, Australia
Peter Gould, New South Wales Department of Education and Training

The large school system referred to in this paper is the government school system in the state of New South Wales (NSW), Australia. The school system serves a population of approximately seven million people. In administrative terms, the school system is organised into 40 districts of roughly equal student populations. Approximately half of these districts serve part of the metropolitan area of Sydney and the other half serve regional and rural areas of the state. Almost all aspects of school administration and governance occur at the state rather than the district level.

The school system has approximately 1,700 schools with classes in the K-6 range. The age range of K-6 students is about 4 years 9 months to 12 years plus. In 1989, the government introduced a basic skills testing program (BSTP) in literacy and numeracy. All students in government primary schools in NSW in Years 3 and 5 are required to take these tests. As well, the tests are used by many of the non-government schools in NSW. The stated purpose of these tests is diagnostic. The tests are intended to provide information to teachers and schools about weaknesses in knowledge at individual, class or school levels. These tests are developed using Item Response Theory (IRT), have a predominantly multiple choice format and are computer-scored.

Assessment can have many different purposes and a range of audiences. The audience for the BSTP is parents, who receive individual reports of their child's performance, and teachers, who have access to the diagnostic information from the test — a pencil and paper, single-occasion assessment. The mandatory testing system just described is not the main focus of this paper. The description of the testing system is given to provide contextual background to the main focus of the paper, that is, a novel approach to student assessment that has been adopted by the school system. This approach involves interview-based assessments undertaken by class teachers, for the purpose of documenting students' knowledge and learning.

THE NEED FOR A SYSTEMIC INITIATIVE IN EARLY NUMBER

In 1995, the NSW school system recognised a need to undertake a systemic initiative in mathematics in the early years (Years K, 1 & 2). This initiative was intended to address several issues: (a) There were vast differences in the levels of mathematical knowledge of students at the Year 3 level, as indicated by the state-wide basic skills testing program. (b) Approaches to assessing students' mathematical knowledge in the K-2 range were problematic. (c) Some students (K-2) were not being sufficiently challenged in mathematics. (d) Teaching programs were not catering sufficiently for the range of student knowledge in mathematics. (e) By and large, teaching programs did not reflect particularly strong or useful theories of young children's mathematics learning or teaching. In the early 1990s, school systems in NSW and other states had undertaken significant initiatives in early literacy learning. These initiatives seemed to address issues in literacy, many of which are similar to those listed above for early mathematics. Thus at this time, the NSW school system was ready to focus attention on reviewing and overhauling the teaching of mathematics in the early years.

BACKGROUND AND OVERVIEW OF CMIT

Development of the CMIT initiative involved from the outset, collaboration between system leaders in mathematics on one hand, and university-based researchers in mathematics education on the other hand. An interview-based assessment had already been developed by researchers and had been used for example, to document the mathematical knowledge of school entrants (Wright, 1991) and the development of mathematical knowledge during the first two years of school (Wright, 1994a). As well, during 1992-5, this approach to assessment had been the basis of a major research program, the focus of which was teacher development in assessing and teaching mathematics in the early years of school (Wright, 1994b; 2000; Wright et al., 1996). Thus the initial development of CMIT involved adapting several key aspects of the earlier research program. This included (a) the assessment tasks; (b) the approach to assessment; (c) the theoretical orientation to learning; (d) explanatory models of student learning; and (e) approaches to teachers' learning (Wright et al., 2000; 2002).

The goal of the CMIT initiative is to develop teachers' professional knowledge of teaching and learning mathematics in the early years, and through this to increase their students' learning of mathematics (NSW Department of Education and Training, 2000). Within CMIT assessment provides the central impetus for teacher change. It is for this reason that the main audience for the assessment data is the classroom teacher. The assessment plays a pivotal role in providing information on what each student can do. Assessment in CMIT is viewed as the process of gathering information to provide direction for teaching. That is, assessment is not an endpoint but a starting point to step into the cycle of teaching and monitoring. If the purpose of the assessment is to provide a lens for teachers' practice then the teacher must trust and own the data. The one-to-one assessment interview is used not only to obtain information as to what students can do. Rather, the teacher carries out assessment in order to learn more about their students' thinking. In this way teachers develop a very strong sense of ownership of the assessment data.

The approach to teachers' learning is school-based, problem-based and team-based. In each of the 40 districts a position of K-8 numeracy coordinator was created. One of the key responsibilities of the numeracy coordinator is the support of CMIT in the schools of their district. Typically this involves supporting an implementation of CMIT in 8-10 of the districts schools, each year. The introduction of CMIT was both an outside-in and inside-out process. As Fullan (1999, p.62) notes, 'Two-way inside-outside reciprocity is the elusive key to large-scale reform'. The numeracy coordinator provides 'at the elbow' consultancy support in classrooms working as a learning partner to teachers in the school teams. The classroom becomes the place of learning for both students and teachers. Teachers learn more about how students think and use that information to adjust the planned instruction for students. Often, the desire to participate in the program originates in the school. Over a period of 6-8 years, almost all schools will implement the initiative.

The resource materials for CMIT took the form of a professional development package which provides a basis for the teachers' learning. Key elements of the professional development package are the Learning Framework in Number (LFIN) and the Schedule for Early Number Assessment (SENA). These are described in the next two sections. Development of these resources drew on current research and in particular the constructivist teaching experiment methodology and results of Steffe and colleagues (Steffe & Cobb, 1988) and investigations by Wright (1994a, 2000).

THE LEARING FRAMEWORK IN NUMBER (LFIN)

The LFIN (Wright, 1998; Wright et al., 2000) provides several key ingredients for the CMIT initiative. First, it provides a general orientation to young children's learning of number. This includes highlighting the significant mathematics content to be learned, describing instructional contexts and tasks for assessment and teaching, and detailing the strategies students use to solve tasks which are problematic for them. The LFIN describes a range of significant aspects of students' learning of early number. These include strategies students use in counting, additive and subtractive situations, students' facility with number words and numerals, students' knowledge of simple addition and subtraction combinations (number facts, number bonds), early multiplication and division strategies (Mulligan & Wright, 2000), ability to think about multi-digit numbers in terms of tens and ones, and use of finger patterns and spatial patterns in numerical contexts. As well, the LFIN provides an indication of directionality in student learning. The teacher using the LFIN should be able to determine the extent of the students' learning in various aspects of LFIN, likely progressions in the students' learning, and instructional situations based on the framework which are likely to lead to those progressions. LFIN can be regarded as a rich description of children's early number knowledge. The term 'knowledge' is used here in an all-encompassing sense — everything the child knows about early number, and includes but is not limited to the child's arithmetical strategies.

Another key feature of the LFIN is that it enables profiling of students' knowledge across various aspects of early number learning. This feature involves writing summary descriptions of aspects of early number learning in tabular forms. Aspects of early number learning for which tabular forms are used include early arithmetical strategies, that is strategies for counting, adding and subtracting, early multiplication and division strategies, knowledge of forward and backward number word sequences, numeral identification, and knowledge of tens and ones. These so-called models of aspects of early number learning can be used to document in summary form, students' current knowledge and progressions of knowledge over time of large numbers of students (Wright et al., 2000; 2002).

THE APPROACH TO ASSESSMENT

In CMIT, individualised student assessment is regarded as a key initial part of teachers' learning (Bobis & Gould, 1998). Teachers learn to administer a schedule of assessment tasks, that is the Schedule of Early Number Assessment (SENA). This involves a one-to-one interview during which the teachers' goal is to elicit as much information as possible about the students' arithmetical knowledge and strategies. Reviewing of the interviews involves using a standard procedure to complete an assessment record sheet, which details the child's responses and strategies, and leads to determination of the student's levels. The assessment process results in three levels of assessment information relating to the students' early number knowledge. First, and most detailed is the videotape record of the assessment interview; second is the completed assessment record sheet which contains extensive written information about the child's responses and strategies. Third, is a table showing the levels of the student's knowledge in terms of four or five of the key aspects of LFIN. The latter constitutes a summarising profile of the students' knowledge and is particularly useful when documenting the knowledge of large numbers of students and making comparisons in students' knowledge over time (Bobis & Gould, 1999).

The CMIT assessment procedure recognises that children frequently use strategies that are less sophisticated than those of which they are capable. This may happen for one or more of several reasons, eg (a) it may be easier and although it may take more time, this may not be of concern to the child; or (b) some feature of the child's thinking immediately prior to solving the current task may focus the child's attention on a less sophisticated strategy. Thus an important challenge for the teacher as observer and diagnostician is to attempt to elicit the child's most sophisticated strategy. This is crucial to gaining a powerful understanding of the leading edge of the child's current knowledge. Viewing assessment and teaching through the lens of the learning framework helps to mediate its adoption by classroom teachers. 'Learning about the Learning framework: I can see...understand where individuals fit in, and therefore what I need to teach them in more detail and more specifically than I've learnt from using the curriculum' (Kindergarten teacher, 16-20 years experience).

DISCUSSION

The approach to documenting student learning that is described above can and should be considered in its broader context. CMIT is a program of teacher development and systemic change that seems to be effective. The program is well-regarded by participating teachers and schools, and also by external observers (Bobis & Gould, 1998; 1999). As well, there are good indications that the program has positively influenced the number learning of vast numbers of students (White & Mitchelmore, 2002).

The novel approach to assessment described above is one of the key elements of the initiative. Assessment in CMIT focuses on more than simply 'right or wrong'. The psychological models that are integral to the learning framework determine the nature of the assessment. The methods of solution children use are important to determine the child's location within the models. The nature of the models also provides a sense of direction for instruction. We agree with Simon (1995) in that our conceptual framework with respect to mathematics teaching includes an emphasis on teachers' inquiry into the nature of students' understandings so they can pose tasks in which those understandings might be challenged and extended.

CMIT does not involve prescribing teaching programs and lessons for teachers. Rather, it provides a school-based context for teachers to review their professional practice. When a team of teachers in a given school participate in the processes associated with CMIT, in a very significant way, they invent or construct the program. CMIT recognises that there is not a single mathematics curriculum (designed, delivered, attained, assessed). Rather, the program works on aligning the assessed curriculum, the implemented curriculum, and the attained curriculum. In overall terms, CMIT is a learning community that: (a) includes students, teachers, numeracy coordinators, system leaders and researchers; and (b) incorporates Senge's (1990) five basic learning disciplines — systems thinking, personal mastery, mental models, building a shared vision and team learning.

References

Bobis, J., & Gould, P. (1998). The impact of an early number project on the professional development of teachers. In C. Kanes, M. Goos & E. Warren (Eds.), *Proceedings of the 21st Annual Conference of the Mathematics Education Research Group of Australasia, Vol. 2, pp. 107-113.* Brisbane: Griffith University.

Bobis, J., & Gould, P. (1999). The mathematical achievement of children in the Count Me In Too Program. In Truran, J. & Truran, K (Eds.), *Proceedings of the 22nd Annual Conference of the Mathematics Education Research Group of Australasia, (pp. 84-90).* Sydney: MERGA.

Fullan, M. (1999). *Change forces: The sequel.* London: Falmer.

Mulligan, J.T. & Wright, R.J (2000). An assessment schedule for multiplication and division. In T. Nakahara & M. Koyama (Eds.), *Proceedings of the 24th Annual Conference of the International Group for the Psychology of Mathematics Education. Vol 4 (pp. 17-25).* University of Hiroshima, Japan: Program Committee.

NSW Department of Education and Training. (2000*). Count Me In Too Professional Development Package.* Sydney: Author.

Senge, P. M. (1990). *The fifth discipline: The art and practice of the learning organization.* New York: Currency Doubleday.

Simon, M. A. (1995). Reconstructing mathematics pedagogy from a constructivist perspective. *Journal for Research in Mathematics Education, 26, 114-145.*

Steffe, L. P., & Cobb, P. (with E. von Glasersfeld). (1988). Construction of arithmetical meanings and strategies. New York: Springer-Verlag.

Thomas, G. & Ward, J. (2001). *An evaluation of the Count Me In Too Pilot Project.* Wellington, New Zealand: Ministry of Education, [Available from Learning Media Customer Services, Box 3293 Wellington, New Zealand (quote item #10211)]

White, P. & Mitchelmore, M. (2002). *The impact of Count Me In Too on Year 3 Basic Skills Test numeracy scores.* Sydney: NSW Department of Education and Training.

Wright, R. J. (1991). What number knowledge is possessed by children entering the kindergarten year of school? *The Mathematics Education Research Journal, 3 (1), 1-16.*

Wright, R. J. (1994a). A Study of the Numerical Development of 5-year-olds and 6-year-olds. *Educational Studies in Mathematics, 26, 25-44.*

Wright, R. J. (1994b). Working with teachers to advance the arithmetical knowledge of low-attaining 6- and 7-year-olds: First year results. In J. Ponte & J. Matos (Eds.),*Proceedings of the Eighteenth International Conference for the Psychology of Mathematics Education, Vol. 4, pp. 377-384.* Portugal: University of Lisbon.

Wright, R. J. (1998). An overview of a research-based framework for assessment and teaching early number. In C. Kanes, M. Goos, & E. Warren (Eds.), *Teaching mathematics in new times: Proceedings of the 21st Annual Conference of the Mathematics Education Research Group of Australasia, (pp. 701-708).* Brisbane: MERGA.

Wright, R. J. (2000). Professional development in recovery education. In L. P. Steffe & P. W. Thompson (Eds.), *Radical constructivism in action: Building on the pioneering work of Ernst von Glasersfeld (pp. 135-151).* London: Falmer.

Wright, R. J., Martland, J., & Stafford, A. (2000). *Early numeracy: Assessment for teaching and intervention.* London: Paul Chapman Publications /Sage.

Wright, R. J., Martland, J., Stafford, A, & Stanger, G. (2002). *Teaching Number: Advancing skills and strategies.* London: Paul Chapman Publications /Sage.

Wright, R. J., Stanger, G., Cowper, M. & Dyson, R. (1996). First-graders' progress in an experimental mathematics recovery program. In J. Mulligan & M. Mitchelmore, *Research in early number learning: An Australian Perspective (pp. 55-72).* Adelaide: AAMT.