Constructing scales for reporting growth in numeracy: the ACER Longitudinal Literacy and Numeracy Study

Andrew Stephanou, Marion Meiers & Margaret Forster
Australian Council for Educational Research

Andrew Stephanou is a researcher at the Australian Council for Educational Research. He has taught physics for 17 years at secondary and tertiary levels. He has extensive experience in large scale testing programs (HSC/VCE Physics and literacy/numeracy statewide testing in Australia) and research in higher education (at the University of Melbourne). Currently his main interest is in educational assessment and psychometrics, more precisely using Rasch Measurement in the construction and description of educational variables. His work includes the construction of phenomenographic categories calibrated with Rasch Measurement on a single scale for the preparation of an instrument for measuring conceptual understanding in physics.

Marion Meiers came to ACER in 1998 with wide-ranging professional experience at all levels of education, particularly in the professional fields of literacy and the teaching of English. She has played leadership roles in the contexts of teaching, professional development, curriculum and assessment, and policy implementation. At ACER she is project director of the ACER Longitudinal Literacy and Numeracy Study. She also directed the DETYA-funded cross–sectoral project, the *Successful Interventions Literacy Research.* She has played a major role in coordinating the 1999 and 2000 ACER Research Conferences, *Improving Literacy Learning* (1999) and *Improving Numeracy Learning,* (2000). Prior to joining the ACER, she was consultant to the Literacy Section in DEETYA, Canberra, in 1997; and Executive Liaison Officer of the Australian Literacy Federation from 1993 - 1997.She was a member of the Expert Group for the development of the national literacy benchmarks. She has a particular interest in curriculum development and assessment of Year 12 English, and chaired the accreditation panel for the new VCE English Language Study. She has been State Reviewer for VCE English since 1992. She also teaches the third year B Ed program in Secondary English Method at RMIT University.

Margaret Forster is a Senior Research Fellow at ACER. Prior to joining ACER, while undertaking post graduate studies in special education, she worked as a special resource teacher. Since 1989 she has been a member of the humanities item writing team, working on a number of testing programs including the NSW Basic Skills Testing Program, the Queensland Core Skills Testing Program, the Victorian General Achievement Test and the Western Australian Monitoring Standards in Education Program. Ms Forster conceptualised and co-authored the first *Developmental Assessment Resource for Teachers – DART English* (upper primary), and supervised the development of the DART middle primary English and DART upper primary mathematics kits. She also supervised the development of materials for use in the 1996 National School English Literacy Survey (NSELS). Margaret is co-author of the NSELS report *Mapping Literacy Achievement* and of the *Assessment Resource Kit* (ARK) materials. From 1995 – March 2000 she led the Humanities (Primary) test development team. She directs the research area, Assessment and Reporting, within the ACER core program. At present on study leave in the USA, Margaret contributed to the initial work on this paper.

Assessing and reporting students' numeracy growth

The ACER Longitudinal Literacy and Numeracy Study is set within the conceptual framework of developmental assessment. The feature of developmental assessment which distinguishes it from other forms of assessment is:

> that the intention of developmental assessment is to obtain an estimate of a student's current location on a progress map as a guide to the kinds of learning experiences likely to be most useful at that stage in the student's learning and as a basis for monitoring growth over time (Masters & Forster 1997, pp. 1–2).

In a paper presented to the 1999 ACER Research Conference, *Improving Literacy Learning,* the early stages of work on the development of a literacy scale were described (Meiers & Forster 1999). The present paper will describe the work which has been done since then on the development of a numeracy scale, and will demonstrate how the progressive achievement of the LLANS cohort of students can be reported on that scale and subscales.

Central to developmental assessment is the use of progress maps, or continua describing increasing levels of achievement. These progress maps or continua provide frames of reference for monitoring the development of individuals or groups. At different points in time, estimates can be made of a student's location on the continuum, and changes in location provide measures of growth over time.

The LLANS will provide the empirical evidence for constructing a numeracy scale, that is:

> a scale based on a measure of each participating student's achievement. Each student's responses to LLANS tasks [will be] used to construct the scale so that the location of [numeracy] skills [will be] based on students' observed performances on the [numeracy] tasks. The method used to construct the scale allows achievement measures to be interpreted in terms of the skills typical of students at various levels of achievement (Meiers & Forster 1999).

Once the LLANS numeracy scales have been developed, it will then be possible to map the achievement of the whole cohort of participating students, subgroups, or individuals at several points in time.

The Longitudinal Literacy and Numeracy Study

The key research question to be investigated in the LLANS is: What is the nature of literacy and numeracy

development amongst Australian school children? One way of responding to this question is to develop a set of scales describing growth in literacy and numeracy. As a national longitudinal study, the LLANS creates an opportunity to develop achievement scales which will describe growth from the first year of schooling through to the stage when students make the transition to secondary school.

A national sample of students was selected from an Australia-wide sample of 100 schools. Ten students were randomly selected from class lists provided at the beginning of the 1999 school year by the 100 schools in the project, creating a total initial sample of 1000 students. As far as possible, students who have changed schools have been retained in the study. Where students have moved to other schools, their continuation in the study has been negotiated with the principal of the new school. Around 900 students from some 140 schools are now participating in the study.

Comprehensive data on the literacy and numeracy growth of the students is being collected each year from two sources: common tasks developed at ACER and work samples selected from the students' normal classroom work. A range of background data on the school, teachers and student variables is also collected annually from a set of questionnaires. This will enable analyses to be made in relation to various subgroups including gender, ESL learners, language background other than English, time spent reading, watching television and using computers at home, and so on.

An important aspect of the methodology is the role of teachers as partners in the study. In these first years of the study, the students' own teachers have worked in a one-to-one interview-like context to administer and record students' oral responses to the common tasks. The item writers faced the challenge of designing tasks to be administered easily by teachers working one-to-one with students. Each set of instructions was carefully worded to make the requirements absolutely clear to both teacher and student, thus ensuring the reliability of the assessments. In designing the common tasks it was also essential to take account of time, acknowledging the practicalities of managing one-to-one assessments in the classroom. Therefore, the focus was on essential aspects of emerging literacy and numeracy.

The LLANS numeracy assessment tasks

Three sets of common tasks have now been completed by students at three key stages of schooling: in the first and final terms of the first year at school (1999) and in the first term of their second year of school (2000). A fourth set of common tasks will be administered in the last term of the 2000 school year.

Four broad aspects of numeracy have been investigated in each of the sets of common tasks:

- number
- space
- measurement
- chance and data

The tasks designed for the first two years of primary school have required students to answer questions orally, while teachers assess and record their responses. Wherever possible, hands-on aids such as rods, counters, shapes, coloured stars, pipe cleaners and match sticks have been provided to support students in responding to the tasks.

The item writers have designed tasks of various difficulties in each of the four broad aspects. For example, in the space strand, the tasks included:

Start of school (Term 1, 1999)

- placing an object upside down, on top of, in front of, behind, under etc;
- identifying shared and different attributes of shapes;
- identifying shapes with same colour and different attributes; and
- naming a geometric figure (square, circle, triangle).

End of first year at school (Term 4, 1999)

- dentifying the first, third and last object in a line;
- naming the fourth position and placing an object between the second and third position;
- counting objects; and
- naming a rectangle.

Start of second year at school (Term 1, 2000)

- reading a simple map to locate adjacent animals and counting animals on the map;
- using arrows to determine direction of a path;
- identifying the first stopping point on a path and identifying animals missed by path;
- using tiles to copy map onto a grid;
- making a square (using match sticks) then making a larger square;
- making a triangle then making a larger triangle; and
- indicating where to cut shapes in half (kite, fish shapes).

The marking guide (categorisation of children's responses) is included with the tasks, and the teacher judges the child's responses against the marking guide. Precise instructions have been provided for teachers to follow so that the tasks are, as far as possible, administered under standard conditions.

Counting (end of first year at school)

Equipment	Instructions and questions	Marking guide	Record
18 counters	Tip out all counters in front of child. Count these and tell me how many there are.	Eighteen Seventeen or nineteen Any other number No attempt	

Pet Shop (start of second year at school)

Equipment	Instructions and questions	Marking guide	Record
Pet shop map and yellow tiles	Tip the tiles out of the bag near the child. One of these tiles fits exactly on a small cage, put a tile on the dog cage. How many tiles do you need to fit exactly on the snake's cage? Don't do it yet. What do you think? *Tell the child to use the tiles if they don't know or guess incorrectly.*	Two, answered without putting tiles on grid. Two, answered after putting tiles on grid. Other No attempt	

The marking guide is coded for data entry, and the entered data is then analysed.

Constructing the LLANS scales: the process

Rasch measurement allows us to display the performance of children and the difficulty of tasks on the same interval scale, in the same units of measurement. High on the scale we see the best performances and the most difficult tasks. Low on the scale we find the poor performances and the easiest tasks.

The three LLANS surveys completed so far contain common items that allow the calibration of all tasks used so far to be displayed on the same scale. The difficulty of a task for which responses have been marked either correct or incorrect is represented by the position of its threshold on the scale. Children above the threshold are more likely to be correct and children below are more likely to be incorrect. A similar explanation is given for tasks rated in more than two categories.

The calibration of the tasks on the scale is followed by an analysis of fit to check the extent to which these tasks target the same latent trait. Misfits in Rasch measurement are a source of information on the performance of children. All misfits are considered and explanations sought. In examining the result of the analysis of fit, some collapsing of the categories in which children's responses had been assigned becomes necessary either because there is insufficient data available for a sufficiently accurate calibration, or because adjacent categories are not clearly and meaningfully discerned. For example, two categories may be too close along the continuum and the location of their thresholds overlap considering the error of measurement.

In figure 2, four of the LLANS early numeracy scale items are used to illustrate our methodology for describing the variable constructed with the tasks from the first three surveys. The relative positions of the thresholds on the interval scale are those calibrated with the data. Item 2PPC is more difficult than item 1PT3, as indicated by their threshold values. Children's responses for item 3RP2, and also item 1OU, have been classified into three categories, therefore their difficulty is shown by two thresholds. The calibration takes into account the different abilities of the children to whom the items have been administered. If, for example, an item has been answered correctly by 60% of the children at Survey 1 and another item by 60% of the children at Survey 3, then the second item must be more difficult because the same percentage of better performing children can answer it correctly. The relative difficulty of the items applies to children anywhere on the scale.

It can be seen that children find more difficulty in recognising that there are three more fish than butterflies in a set of six fish and three butterflies (2PPC) than in creating their own pattern after having been shown an example of a repetitive pattern (1PT3).

The description of the measured variable is a lengthy process in which common features in the categories of items belonging to the same part of the scale are identified. Regions of the scale, partly overlapping, with qualitatively different and meaningful description are formed. The description of these regions constitutes the description of the measured variable. The process would be similar to a verbal description of the temperature scale. The region around 0ºC would be described differently from the region around 15ºC degrees, etc.

Reporting growth on the LLANS numeracy scales

The construction and description of suitable variables for showing the variation in the skills children develop during their early years at school makes it possible to show the rate at which children develop various skills.

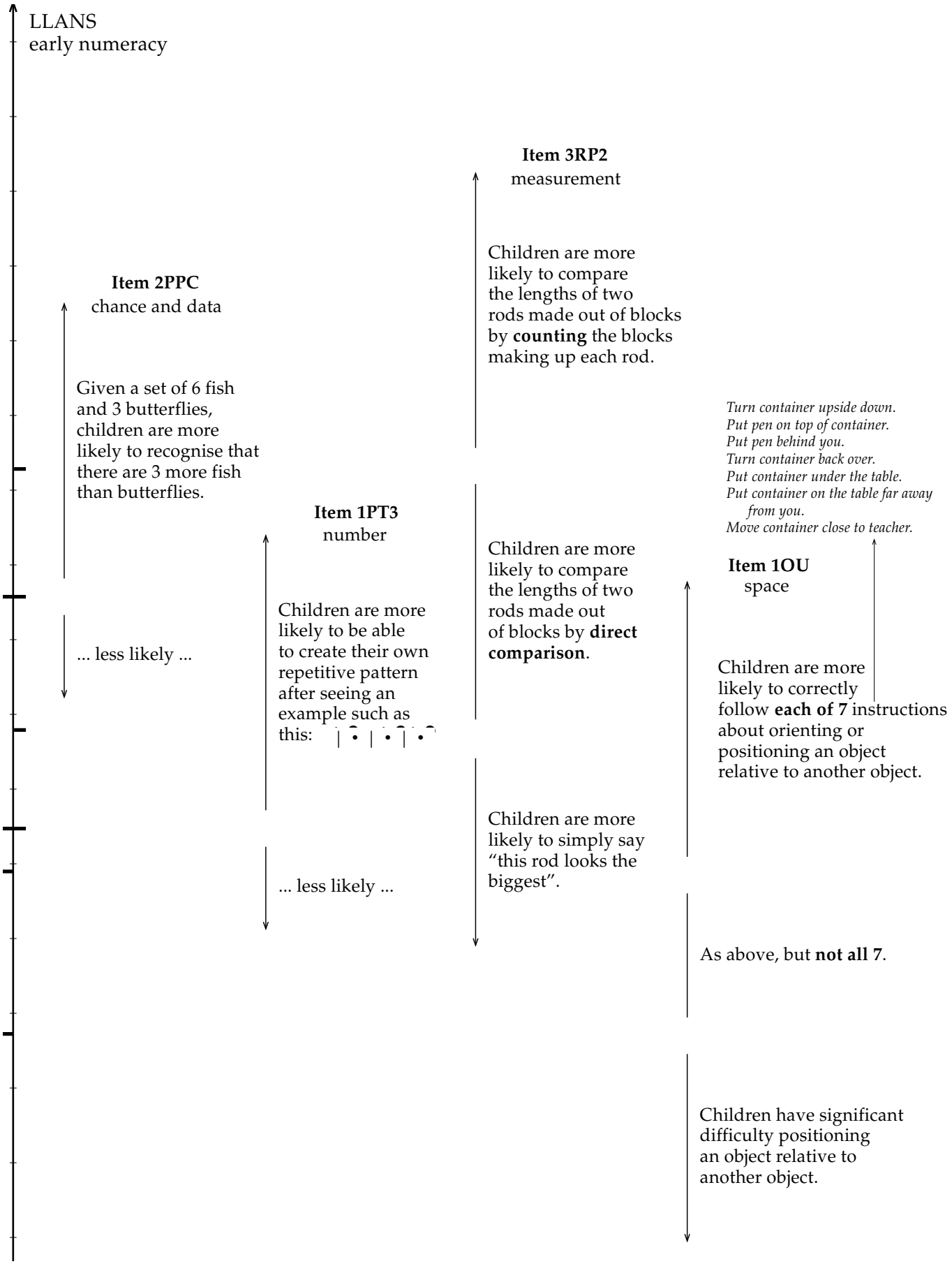

Figure 1. The LLANS early numeracy scale and four of the tasks used to describe it. The two items on the left are dichotomous and the two on the right are polytomous (three categories).

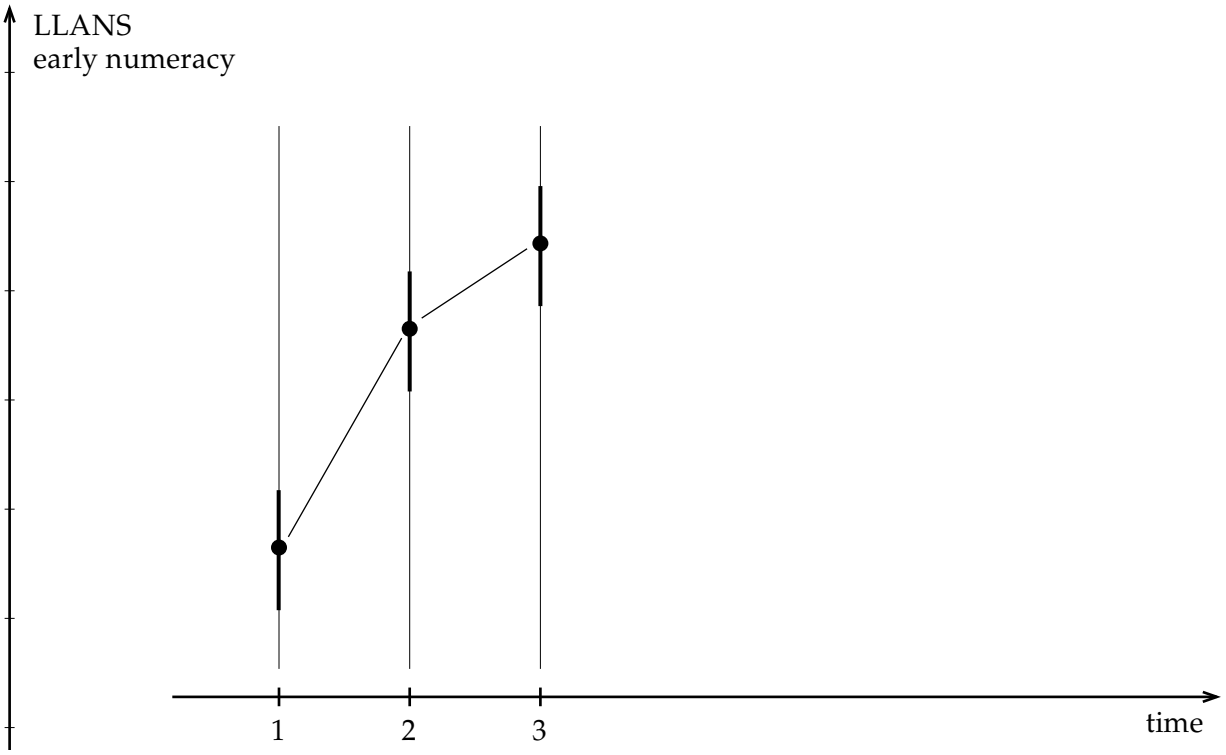

Figure 2. The LLANS Early Numeracy scale and an example of a growth path

Figure 2 shows how rate of growth over time can be reported at the individual child level and for selected groups of children. It can be shown how rate of growth depends on prior achievement through real examples.

Multilevel modelling of the measures of children's performance on the LLANS scales will allow the identification of factors affecting growth.

Acknowledgements

The development of the numeracy items was carried out by Prue Anderson, Wendy Bodey, Lynne Darkin and Eve Recht from the Australian Council for Educational Research.

References

Masters, G. & Forster, M. (1997). *Developmental Assessment, ACER Assessment Resource Kit (ARK),* Australian Council for Educational Research, Camberwell.

Meiers, M. & Forster, M. (1999). *The ACER Longitudinal Literacy and Numeracy Study (LLANS),* paper presented to ACER Research Conference: Improving Literacy learning – What does the research tell us?, Adelaide, 18–19 October 1999.