Appendix 8 Curriculum, Evaluation and Management (CEM) Centre-Consistency in the learning difficulty Scale for numerals as an example of potential tools to support teachers.

Tymms & Wylde (2003), Tymms, Merrell & Jones (2004), and Merrell & Tymms (2007) published data on cross-cultural difficulty patterns for sets of Mathematics, Reading and Vocabulary test items in teacher administered, computer-adaptive tests using the Performance Indicators in Primary Schools (PIPS) On-Entry Baseline Assessment. The PIPS On-Entry Baseline Assessment, developed and managed by the CEM centre, combines objective assessment and teacher ratings to provide information about each child as they enter their first year in full time education. At the core of the PIPS On-Entry Baseline is an assessment of early reading, early mathematics, phonological awareness and short-term memory. The assessment is completed by an adult (usually but not necessarily the teacher) working with each child on a one-to-one basis at a computer screen and takes about 20 minutes.

As a result of the very comprehensive record storage process, including the individual responses at each assessment session, a database has been developed covering a large number of cases. The database has records across countries including England, Scotland, the Netherlands, Germany, Hong Kong, Australia and New Zealand.

Merrell & Tymms (2007) reported student responses from England, Scotland, Western Australia and New Zealand. Although English students could also be tested in Bengali, Cantonese or Urdu only the English language tested data subset was used (Tymms, personal correspondence, 2008). Other publications (Tymms & Wylde, 2003; Tymms et al., 2004) include students from the Netherlands (in Dutch), Western Australian Indigenous communities and English hearing –impaired students. Tymms has also supplied the detailed item difficulties for some of the samples involved, to the author (Tymms, personal communication, 2006). All these sources have been drawn upon to develop a conceptual argument concerning the possibility of invariance of item parameters across cultures in the learning of numbers in English.

If certain numerals appear to be recognised before others, this phenomenon allows the observer to monitor learning progress as more difficult numerals are recognised. More importantly, if the ‘distance’ from demonstrating the ability to recognise a particular numeral is consistently the same learning distance (in terms of differences of item difficulty) from another specified numeral, there are strong hints that there is a scale for the learning of numerals. The conditions of order and consistent intervals, the prerequisites for measurement, are met.

The Tymms et al. data provide convincing examples of consistency of item order across cultures, strongly suggesting in numeral development at least, there is a natural approximate order in which students master the naming and recall of numerals 0 to 9, that is in their development of a language of number words. Extending this further to naming 2 and 3 digit numerals and computations, the consistent item difficulties (and inter-item interval distances) obtained, are shown to be independent of the English-speaking culture from which a student is derived.

More broadly Merrell & Tymms (2007) reported across a wider set of items covering mathematics, reading and vocabulary. The strongest correlations between item difficulties across countries were in mathematics. Correlations of the difficulties of the items in the four countries (England, Scotland, New Zealand and (Western) Australia) were all 0.99. “This is so high that no further preliminary action was needed before making comparisons. The difficulties of the reading items were also strongly related but not quite so strongly.” (Merrell & Tymms, 2007, p. 127.). Table 8.1 reproduces Tables 6 to 8 from Merrell & Tymms (2007) showing the inter-item difficulty correlations across countries. The reading correlations (Table 6) ranged from 0.99 to 0.94, vocabulary items (Table 7) ranged from 0.99 to 0.95 and phonological awareness items from 0.91 to 0.98.

Table A8.1 Correlations between item difficulties, by country, reprinted from Merrell & Tymms (2007)

table 6 correlation between difficulties of 56 reading items[a]

	WA	NZ	England
NZ	0.99		
England	0.97	0.96	
Scotland	0.96	0.94	0.99

table 7 correlation between difficulties of 17 vocabulary items[a]

	WA	NZ	England
NZ	0.96		
England	0.97	0.96	
Scotland	0.95	0.96	0.99

table 8 correlation between difficulties of 17 phonological awareness items[a]

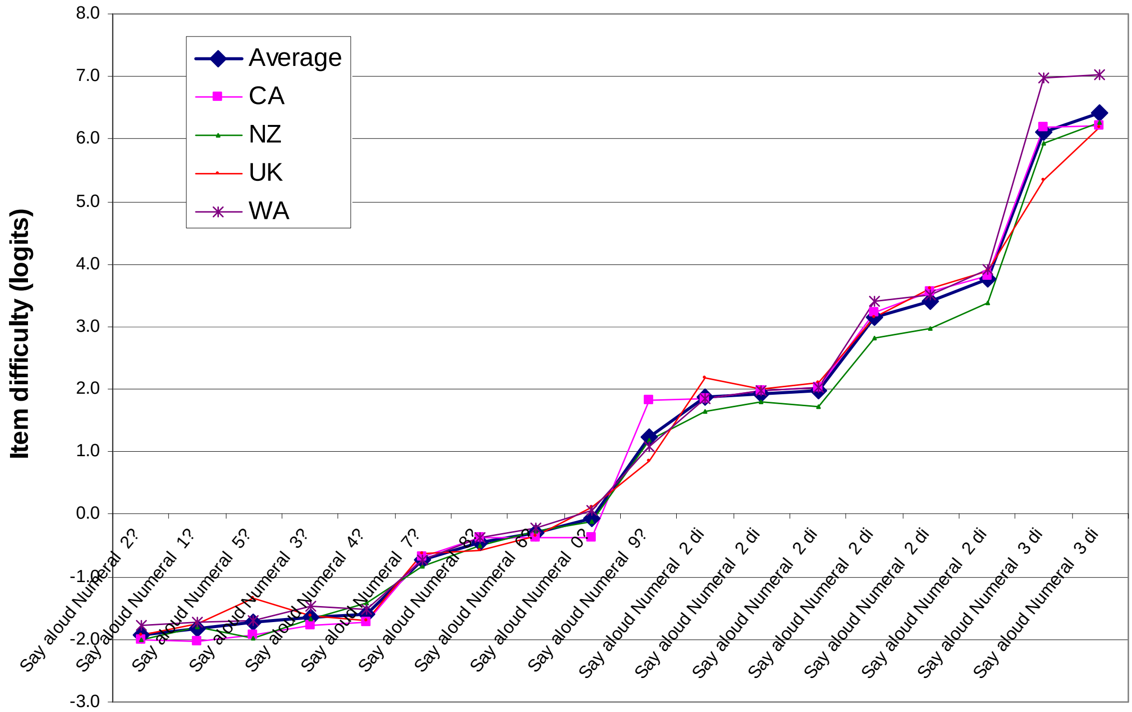
	WA	NZ	England
NZ	0.98		
England	0.96	0.98	
Scotland	0.91	0.91	0.95

[a] A few infrequently presented items with large errors were omitted.

Figure A8.1, based on the data supplied by Tymms (personal communication, 2006), illustrates that there were minimal variations in individual item difficulties across cultures in saying aloud numerals. English speaking Cantonese (CA) show a slight advantage (easier to say the number name), consistent with Miller, Smith, Zhu & Zhang (1995), where the simplicity of Chinese word names provide an advantage to Chinese speakers. This advantage would appear to flow over to English language names.

The line of mean item difficulties (designated as 'average' in the figure) shows items in increasing difficulty order; the individual sample lines illustrate the small variation in difficulty across cultures. New Zealand for example appears to show a slight advantage (i.e. easier) for naming two digit numerals. Cantonese English speakers learning English numerals found 9 harder to master than did other language/cultural groups. Three digit numbers were harder to master in Western Australia than elsewhere. The pattern of similarity across cultures is, however, remarkably consistent.

Figure A8.1 Item Difficulties over Four Culturally Different Samples (Tymms, personal communication, 2006)

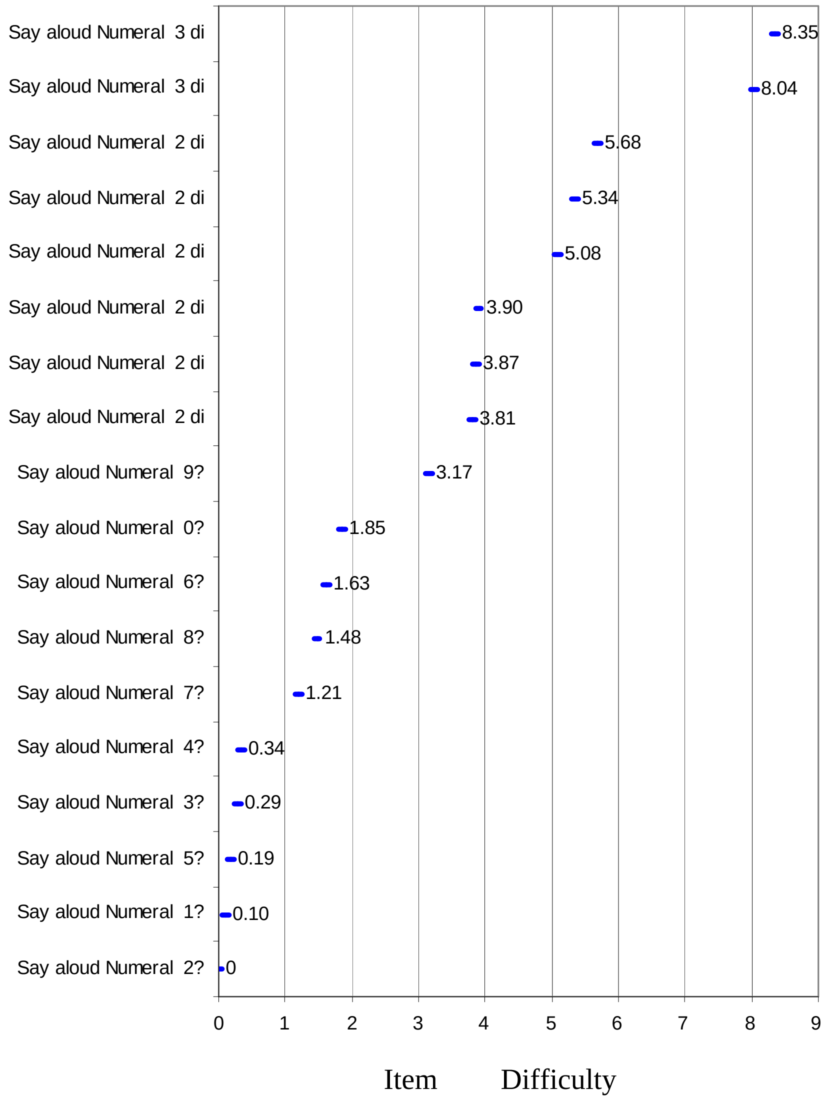


Tymms et al. (2004) and Merrell & Tymms (2007) applied differential item function (DIF) analyses to show that there was a no difference (i.e., lack of bias) in numerical recognition items, for any of the cultural groups. Some DIF was found for a small number of vocabulary items and explained as culturally related (`wasp' and `pigeon' were more difficult in Australia). These estimates of student performance also show a strong pattern by age, similar to the NAPLAN data. .

In more detail, Figure A8.2 presents a map of the item difficulties, anchored to the difficulty of learning to recognise the numeral 2 (say aloud the number word, the marginally easiest numeral to identify over all cases[36]).

[36] Later in this Appendix, a similar analysis, but with more cases, reverses the position of 1 and 2.

Figure A8.2 Map of Item Difficulties anchored to Recognise/Say Aloud 2 =0



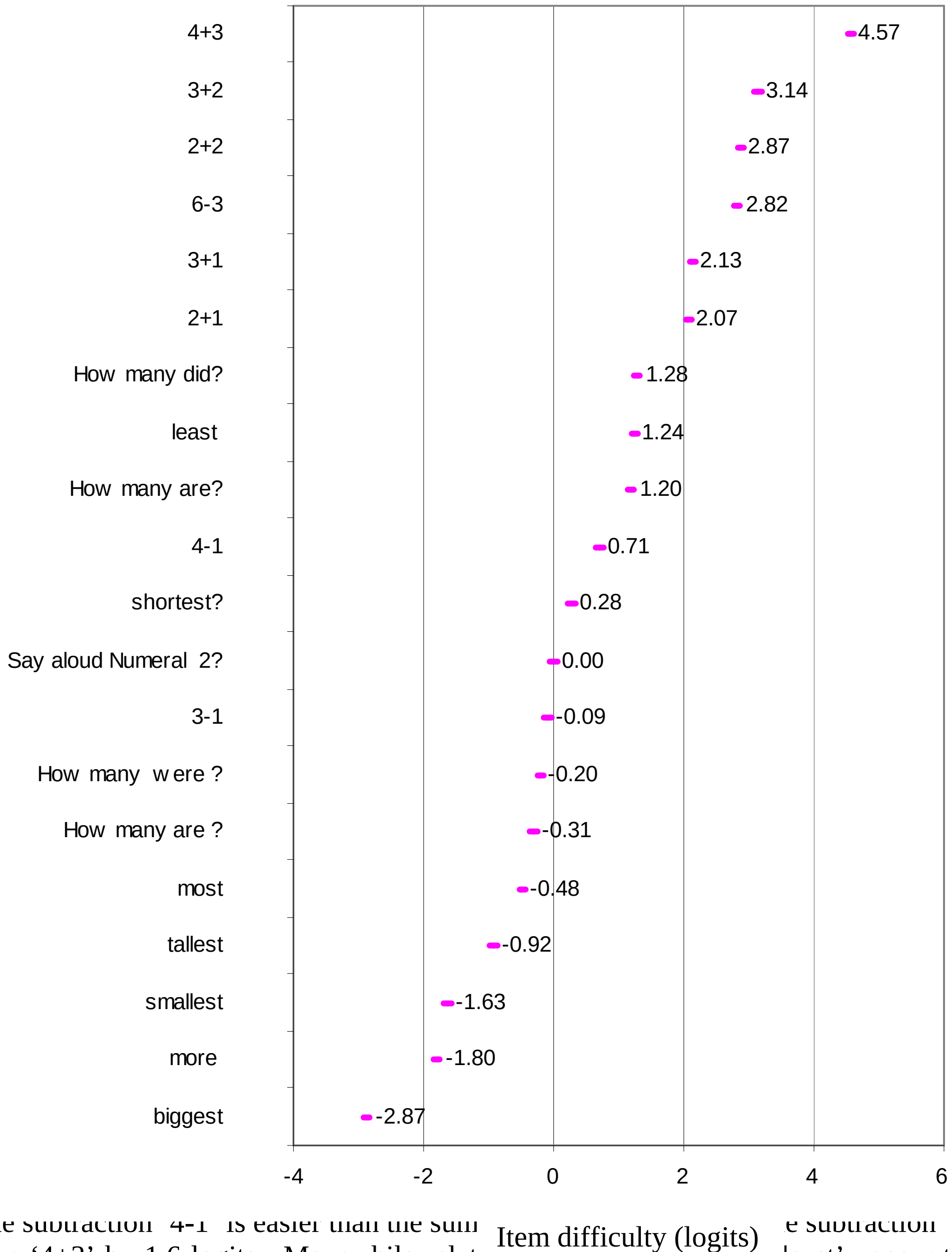
The CEM data set is expanding annually and internationally and thus more data might revise the order. Based on the data to date there is an order in numeral recognition which is plausible (that is logically consistent with experience) but at a level of refinement apparently not appreciated by most observers. It appears 2 is recognised marginally ahead of 1, that is it is slightly easier to recognise. The order of numeral recognition appears to be 2,1,5,3,4,7,8,6,0,9. The increase in difficulty from recognising 4 to recognising 7 is almost one logit, the distance from 2 to 9 being over 3 logits. The challenge to young children in building numeral recognition skills (as a small example of the complexity of all the early mathematics and language skills) is great. The change in difficulty levels of average Year 3 students progressing to Year 5 in mathematics or literacy is approximately 1 logit, although the logits may not be directly comparable. The high rate of change of learning development implied in early number recognition is however consistent with the diminishing learning rate with time described earlier in Chapter 5.

From the map in Figure A8.2 saying aloud 3 digit numerals, by implication, the recognition of place value using appropriate descriptions is more difficult than recognising numeral 2 by more than 8 logits, a large increase in difficulty. Average performance of students measured in logits (though not exactly comparable) in progressing from Year 2 to Year 12 is estimated to be about 6 logits.

Figure A8.3 maps the difficulties of small calculations and the recognition of relativities (larger, smaller, most), once again anchored to a difficulty of recognising 2 given a value of 0. Identifying which is the biggest of three items (cats) is almost 3 logits easier than recognising 2. Counting 4 items is about 0.3 logits (i.e., just measurably) easier than recognising the numeral 2.

Counter-intuitively, but not overly surprising, subtracting '1' seems to be easier than adding 1. The calculation '3-1' in the form of 'Here are three beach balls. If you took one away how many would be left?' has a difficulty of –0.9 (i.e., it is easier than recognising the numeral 2) while the sum '3+1' in the form, 'Here are three bikes. If you put one more bike in the picture how many would there be?' has a relative difficulty of 2.13, over 2 logits more difficult than the subtraction equivalent.

Figure A8.3 Map of Difficulties for Relative and Computational Items (anchored to 'Say aloud 2' =0 logits)



The subtraction '4-1' is easier than the sum [illegible] e subtraction '6-3' easier than '4+3' by 1.6 logits. Meanwhile relat[illegible] shortest' and 'least' appear to be much harder than their opposites, 'tallest' and 'most'.

Additional data (Merrell and Tymms, 2007), not detailed here, illustrate similar regularities in the areas of reading, phonological awareness and vocabulary. The full suite of school entry assessment items covers writing, vocabulary, ideas about reading, repeating words (assessment of phonological awareness), rhyming words, letter identification, word recognition and reading, ideas about mathematics, counting and numeracy, sums (addition and subtraction problems presented without symbols), shape identification, digit (numeral) identification (single, two and three digit numerals), and mathematics problems (including calculations with symbols).

The purpose of this analysis is to illustrate that developmental maps, based on empirical student-derived data could provide teachers with some of the tools to note and understand each student's progress. This is a key element of the general thesis, that teachers with the right conceptual tools (likely-order maps say) could observe, understand and record student progress. This proposition is not particularly earth shattering. As described in the main text, Masters and Forster (1996) and others have been proposing similar approaches for almost two decades. The advantage brought by the CEM data is the confidence it should give to teachers in classrooms and others, that the patterns of order of skill development are genuine and consistent across cultures and thus reflect some significant underlying characteristic of learning. Particular skills and developing abilities appear, unsurprisingly, to be dependent on earlier skills and abilities and progress along the developmental pathway is not necessarily smooth or easily achieved, based on the estimated 'difficulty measures' of learning later skills/tasks relative to earlier ones.

Item difficulty can be varied somewhat by the design of the item. A trivial and obvious example is the contrast of the difficulties of requiring the student to read the item (when the child cannot yet read) versus a teacher mediated computer presentation of the same item as CEM provides. In one form the computational or recognition skill alone is observed, in the other the computational skill and the ability to read are combined producing an item of much higher difficulty. Clearly establishing the relative difficulties of particular skills will require dissection of the contributors to item difficulty but the CEM data sets have already shown that the difficulty patterns are likely to be consistent over cultures and are thus most likely related to inherent properties of the particular cognitive skills relative to other skills.

Tymms et al.(2004) make a strong case for the usefulness of their assessment approach in cross-cultural studies to better understand cognitive development.

> The analyses presented … have explored the possibility that a baseline assessment (the PIPS On-entry Baseline) could be used to make comparisons of pupils starting school in different countries and cultures. The evidence suggests that this is indeed possible. The assessment behaved well across the groups that were studied and the general developmental patterns also appeared to be very similar across the groups. Clearly, some of the analyses indicate that more work is needed on the assessment items but that is to be expected in a pilot. The way is now open for a serious international study of the cognitive developmental levels of children starting school. (Tymms et al., 2004, p. 688)

A significant benefit of the CEM approach, through promotion of potential item maps derived from their data, would be to help teachers understand cognitive development. The cross cultural consistency should help teachers believe such skill development orders might be genuine and based on a roughly predictable model of learning development. The maps would become reference frames for understanding and documenting where each student is in real time.

CEM Number word development updated

More recent data (Jones, 2008, personal communication) is recorded in Figure A8.4. This new analysis determined the difficulty of every number presented from 0 to 999. Figure A8.4

is based on unpublished data from CEM. The difficulty scale is based on at least three separate analyses. To very approximately equate the scale for Figure A8.4 to be similar to Figure A8.2, the scale for Figure 8.4 was rescaled to make the difficulty difference from 1 to 9 to be 3.1 logits, the same as in Figure A8.2. The numerals follow approximately the same order but Figure A8.4 has identified each number sequence from 1 to 999, providing much more detail about the likely order of numeral learning, although for numerals higher up the order the differences in relative difficulty are too small to be meaningful. To illustrate the general trends, most 2 digit numbers in the thirties are about 1 unit more difficult than repeat digits (22, 44, 88), which are learnt it seems approximately in the order of the initial difficulty of their single digit components (66 and 99 exceptions). Saying aloud 100 is 2 units less difficult than 200, which is equivalent in difficulty to 101 the next hardest 3 digit number. The round hundreds are easier than most other three-digit combinations. The last numerals identified are 556, 716, 701, 770 and 917. The differences are small for a large portion of the difficulty sequence and the error of measurement is of the order of 0.75 to 1.0 units (a consequence of small numbers of cases combined with large individual differences in the difficulties of digits at the 'hard' end of the scale).

The link between developing a vocabulary of number words and the development of counting (which comes first?) has been investigated recently by Condry & Spelke, 2008. In a series of experiments with young children learning number words and counting they report their work

> provides evidence that natural number concepts emerge in children along with or after, rather than prior to, the acquisition of language. These concepts likely emerge, in part, as a consequence of children's efforts to make sense of number words and to learn to use the counting routine to represent number: achievements that the children in the present experiments had not yet attained. (Condry & Spelke, 2008, p. 35)

Surprisingly this is quite recent research. This understanding, combined with a number word-learning map of the sort derivable from CEM research and confirmable from other sources, would provide a teacher with an observation framework for number word development. A logit related scale would provide a basis for easy recording of learning status (and linking to actual testing at the next CEM assessment for clients of that system).

The CEM data are presented to illustrate why knowing the relative difficulties of learning to say aloud a number (recognise the digits and verbalise the name) might help a teacher. A student who can say aloud a single digit but not two digit number, can be recorded as having developed to position x at time t_1. Specific 2 digit numbers indicate position y at time t_2 and 3 digit numbers indicate position z at time t_3. The specific numbers verbalised are indicators of generally where a student is in their recognition of numbers. Assuming a more refined analysis than Figure A8.4, which has high errors of measurement for 3 digit numbers as the sample sizes per case are of the order of 50, an indicator of progress is provided. The assessments can be observational, unobtrusive, and recorded in each case on the last recorded numeral said aloud, reported as a common vertical scale value.

Figure A8.4 Numbers in Estimated Order of Difficulty to Say Aloud-all numbers to 20, samples from thereon (Difficulties relative to '1')