TRAVELLING THE ROAD TO EXPERTISE: A LONGITUDINAL STUDY OF LEARNING

Kaye Stacey

University of Melbourne, Australia

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

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

Our travelling companions: the students

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

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

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

The transport: their teaching

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

The destination: understanding decimal notation

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

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

Early explorers mapping the territory

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

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

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

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

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

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

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

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

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

The territory and the mapping tool

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

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

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

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

Table 2: Matching of codes to the ways of thinking

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

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

THE JOURNEYS

Some sample journeys

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

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

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

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

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

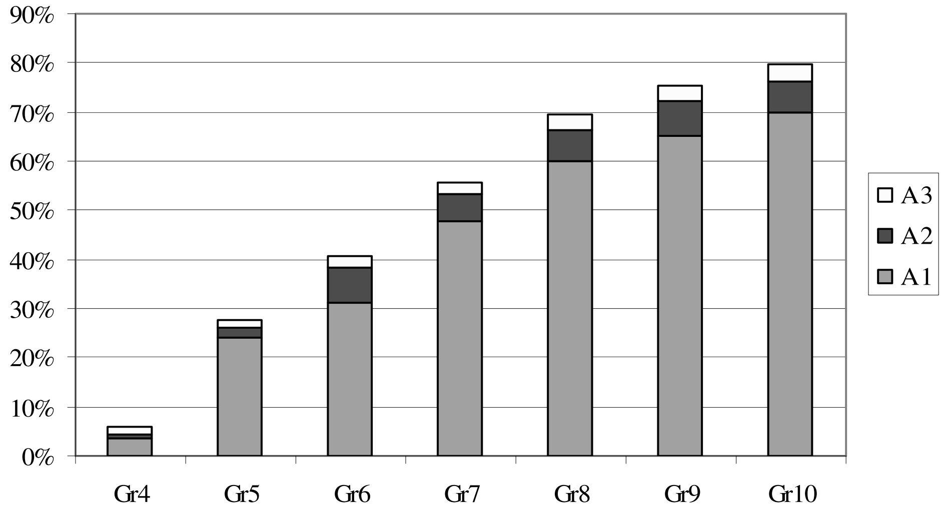


Figure 2: Best estimate of the prevalence of A codes by grade
(from Figure 9.3, Steinle, 2004)

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

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

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

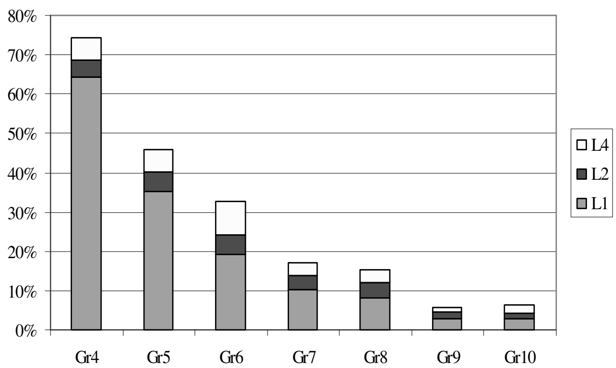


Figure 3a: Prevalence of L codes by grade (from Figure 9.7, Steinle, 2004)

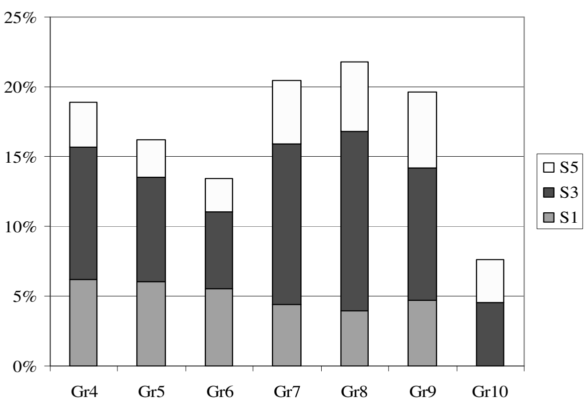


Figure 3b: Prevalence of S codes by grade (from Figure 9.10, Steinle, 2004)

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

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

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

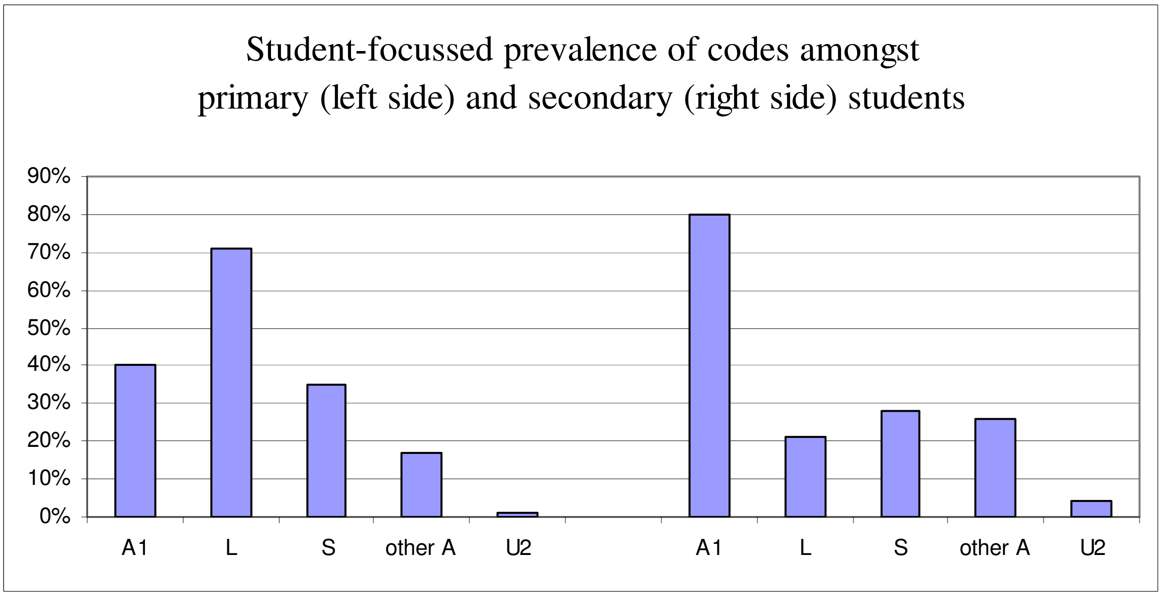


Figure 4: The percentage of students who test in given codes at some stage in primary and secondary school (derived from Steinle, 2004, Ch 9).

Persistence: which towns are hard to leave?

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

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

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

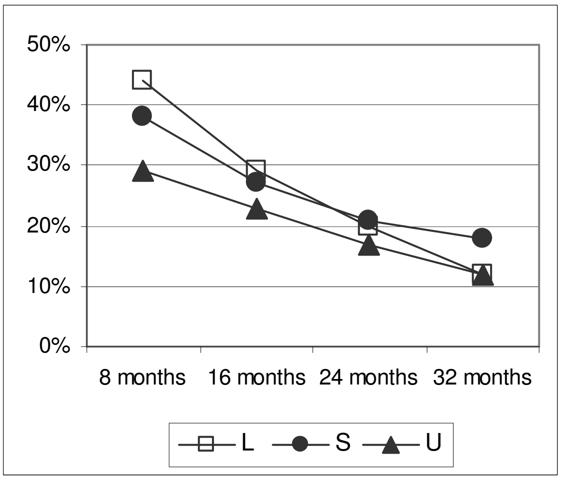


Figure 5a: Persistence in L, S and U codes after 1, 2, 3 or 4 semesters (adapted from Steinle, 2004, Fig. 6.5)

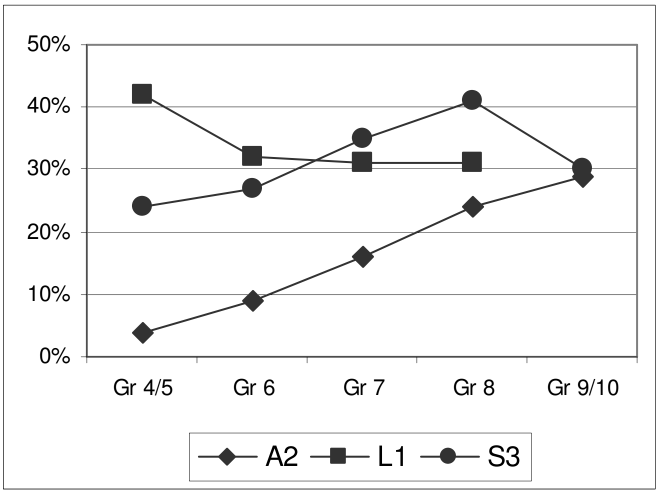
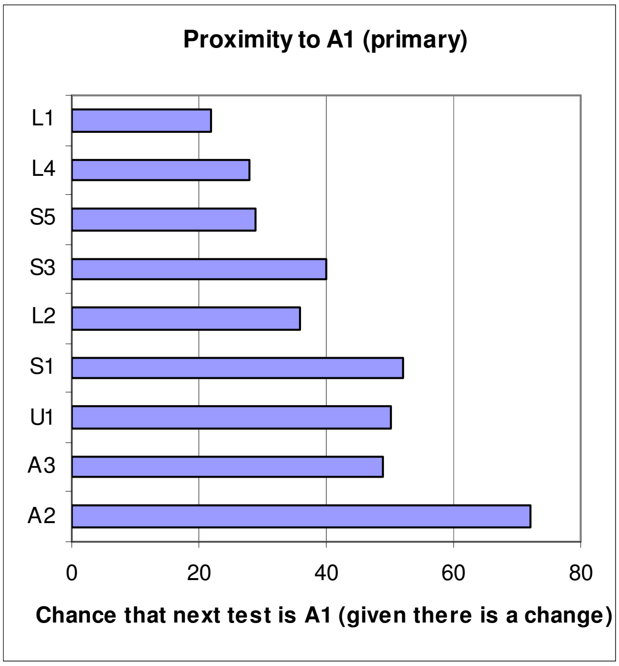


Figure 5b: Persistence in A2, L1, S3 and S5 over one semester by grade of current test (adapted from Steinle, 2004, Fig. 6.1)

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

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



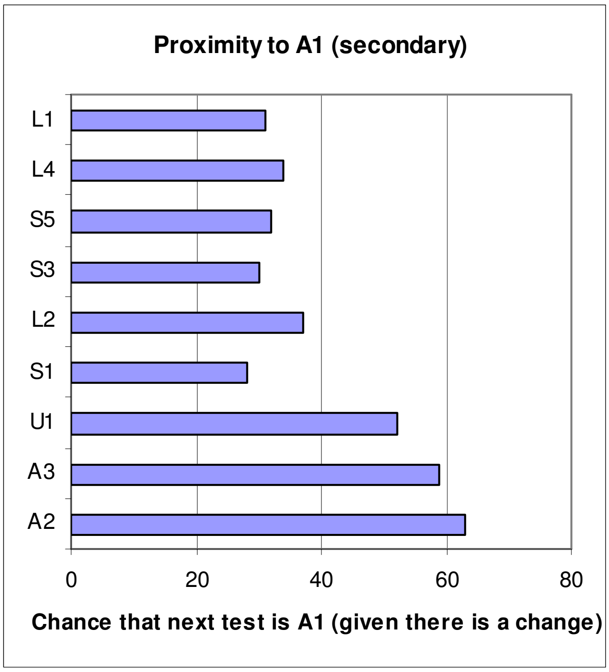


Figure 6: Chance that the next test is A1, given there is a change of code, for primary and secondary cohorts. (Codes ordered according to combined cohort proximity.)

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

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

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

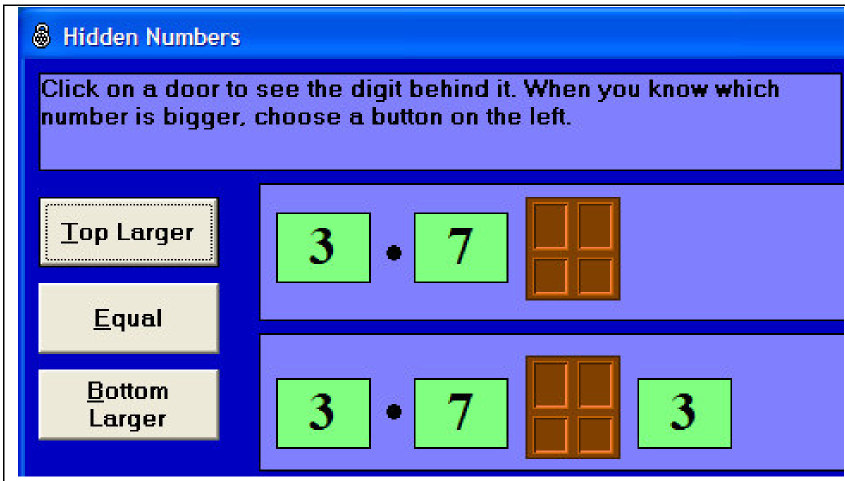
The study of student's thinking especially in the A and S codes has highlighted difficulties associated with zero, both as a *number* and as a *digit*, that need attention throughout schooling (Steinle & Stacey, 2001). Zeros can be visible or invisible and represent the number between positive and negative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

HOW IS A DETAILED MAP OF LEARNING USEFUL?

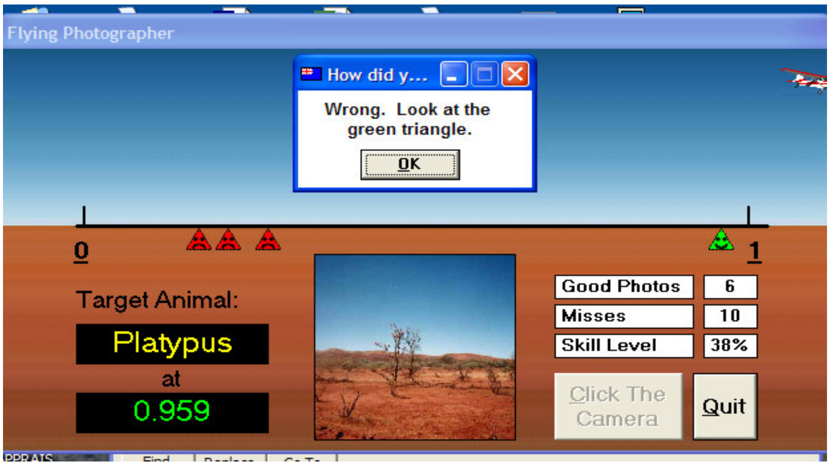
The research work in the 1980s using comparison of decimals identified three "errorful rules". The map of the territory of learning decimals at that stage therefore divided it into four regions (expertise and three others). DCT2 can diagnose students into 12 groups (the 11 of the longitudinal study and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

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

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



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



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

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

LESSONS ABOUT LEARNING

An overview of the journey

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

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

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

How the learning environment affects the paths students take

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

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

Learning principles or collecting facts

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

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

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

Acknowledgements

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

REFERENCES

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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