A Longitudinal Study of Students 'Understanding of Decimal Notation: An Overview and Refined Results.

Vicki Steinle
University of Melbourne
<v.steinle@unimelb.edu.au>

Kaye Stacey
University of Melbourne
< k.stacey@unimelb.edu.au>

This paper provides an overview of the major results of a large-scale longitudinal study of students' misconceptions of decimal notation, drawing them together and presenting refined results. Best estimates of the prevalence of various misconceptions about decimal numbers from both cross-sectional and longitudinal perspectives are provided, as well as some estimates of persistence. Strengths, limitations and suggestions for improvements to the Decimal Comparison Test as well as major implications for teaching are discussed.

This paper aims to provide an overview and discussion of the results of a longitudinal study of students' understanding of decimal notation, and to present some refinements to previously published results. As far as we know, this is the first study to track the misconceptions of a large number of students longitudinally and hence for the first time to describe the paths that students take through misconceptions on their way to expertise. The study revealed new misconceptions of decimals, found how common they are across the middle years of schooling, explored which misconceptions were "better" or "worse" to have, and showed underlying links between them. A range of associated studies not reported here and using different samples of students, investigated effective teaching.

Twelve volunteer schools from high, medium and low socio-economic areas of Melbourne were involved over a 4-year period (mid-1995 to mid-1999). Over 3000 students from Grades 4 to 10 completed nearly 10 000 tests. Classes were tested twice each year; due to occasional late class testing and absences of individual students on the designated days, the average inter-test time was 8.3 months. The number of tests completed by individual students ranged from one to seven. The project did not suggest to the schools that any special teaching of decimal notation should be undertaken: some teachers provided special teaching while others did not. Schools were selected as a representative, but not random, sample but a comparison with a TIMSS-R (1999) item shows that the results may be a reasonably good indicator of the general Australian situation.

A one-page Decimal Comparison Test was used to collect data. The particular version used is called DCT2, which consists of 30 pairs of decimals with the instruction: *For each pair of decimal numbers circle the one which is LARGER.* This test diagnoses a student's misconception about decimal numbers by the precise pattern of responses made by that student on the test items. Patterns of responses are first classified as exhibiting one of four behaviours (and allocated a coarse code A, L, S, U) and then more detailed analysis identifies one of 12 fine codes, which each link to one or two ways of thinking (expert or a misconception). Full details of the allocation of a code to a student's test are provided in Steinle and Stacey (2003a), which also provides a full description of the ways of thinking which gives rise to each code. Table 1 summarises the four coarse codes.

A discussion of the strengths, weaknesses and improvements to DCT2 is given in a later section. It is important to note, however, that the code A1, which indicates expertise (i.e. very few errors on the test) does not necessarily imply that a student is truly an expert with respect to decimal understanding. Students who can accurately follow correct (or nearly correct) procedures for comparing decimals will score highly, whether or not they understand why those procedures work. For these two reasons, all the estimates in this paper of the numbers of expert students are over-estimates.

Table 1
Description of Four Behaviours Represented by Coarse Codes A, L, S, U.

Behaviour		Description
A	apparent-expert	A collection of ways of thinking (A1, A2, A3) that generally lead to students choosing the correct decimal. Some, but not all, of these students are true experts.
L	longer-is-larger	A collection of ways of thinking (L1, L2, L4) that generally lead to students choosing the longer decimal (more digits) when asked to choose the larger number.
S	shorter-is-larger	A collection of ways of thinking (S1, S3, S5) that generally lead to students choosing the shorter decimal (fewer digits) when asked to choose the larger number.
U	unclassified	None of the above. (U1, U2)

How Prevalent are the Misconceptions and Expertise?

There are two ways of describing the prevalence of the misconceptions. First we examine the cross-sectional prevalence by grade, which gives the proportion of students in a given grade who are expected to have the misconception. Second, we estimate the proportion of students who hold a misconception at some stage of their schooling. Both of these measures are recommended for use in future research.

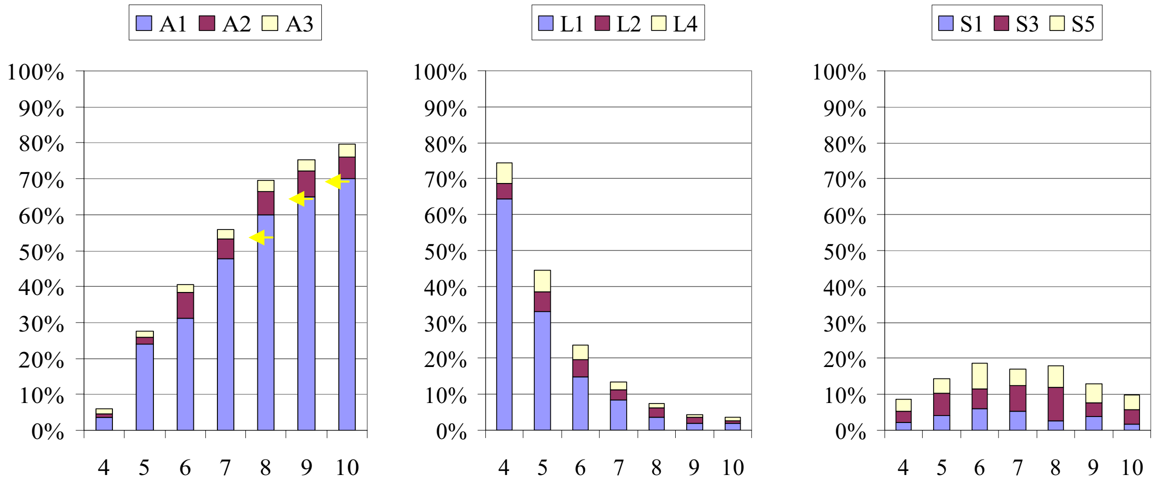
Cross-Sectional Prevalence by Grade

The cross-sectional prevalences are provided in Figure 1; the fine codes being grouped into the coarse codes A, L and S. The prevalence of the coarse codes were provided in Steinle and Stacey (2003b), but the prevalence of the fine codes is now included. The cross-sectional prevalence by grade in Figure 1 is calculated only on the first test that each student completed, as it was found that having been in the testing process resulted in an additional 10% of students being experts (i.e. code A1). This may have been due to teachers giving the topic additional attention, or due to raising the awareness of individual students that there was something that they had to learn. The restriction to student's first tests resulted in a non-representative sample in Grades 8, 9 and 10 as there were no students from the "highest achieving" school. While the actual results are indicated by an arrow in Figure 1a, the prevalence of A1 in Grades 8, 9 and 10 have been adjusted upward to compensate for this sampling issue, thereby providing a better estimate.

Figure 1a confirms that the prevalence of A1 (an expert on DCT2) increases with grade as one might expect, although reaching only 70% at Grade 10 is a major cause for concern. A further 10% of students at each grade in the secondary school are in codes A2 and A3. This means that they can only order straightforward decimals (e.g. those distinguishable from their first one or two decimal places and without zeros in key positions). These students may have very little understanding of the meaning of a decimal number, but their misunderstandings can be masked by ability to perform correctly in most circumstances.

Figure 1b indicates that the youngest students start with L behaviours, but that most leave these ideas behind. In contrast, Figure 1c indicates that the prevalence of S behaviour is more constant at 10%-20% throughout the secondary school. Stacey, Helme, Steinle, Baturo, Irwin and Bana (2001) found that between 3% and 8% of the pre-service teachers in their sample from four Australian universities exhibited S behaviour, indicating the importance of addressing these misconceptions at school and in teacher education. While not shown in this figure, approximately 10% to 15% of students in each grade completed a test coded as U (i.e. it did not meet the strict criteria to be coded as A, L or S).

A comparison of the Grade 8 results for the properly constituted random Australian sample on the TIMSS-R item B10 (TIMSS-R, 1999) indicates that the sample in this study is slightly above average in prevalence of expertise, but a comparison of the distribution of responses on each distractor (data held at ACER) indicates a very close match with the prevalence of L and S behaviours in Figure 1b and 1c. Research in other countries (e.g. Brekke, 1996; Peled, 2003) can identify similar misconceptions although the prevalence varies from one country to another reflecting local curricula.



a) A1*, A2 and A3 b) L1, L2 and L4 c) S1, S3 and S5

Figure 1. Cross-sectional prevalence of the coarse and fine codes by grade (*A1 adjusted).

Longitudinal Prevalence by School Level

An alternative view of prevalence of misconceptions is to find out how many students are affected during some stage of their schooling. Steinle and Stacey (2003b) proposed that the most useful measures supported by the data were for primary school and secondary school separately, and provided the results for the coarse codes. Here we give the results for the fine codes. The longitudinal prevalence of a code is the percentage of students who completed at least one test that was allocated the given code whilst they were in primary school (for this sample Grades 4 – 6) and in secondary school (Grades 7 – 10). The calculations for Table 2 were therefore based on two restricted samples; 333 students who completed at least four tests while in Grades 4 to 6 and 682 students who completed at least four tests in Grades 7 to 10. Note that the values in this table cannot be combined as most students contribute to more than one row. The restriction to students who completed at least four tests is necessary because “less-tested” students had less opportunity to show their thinking and changes in their thinking and so “dilute” the measures. More testing of each student causes these measures to increase, if they change their ways of thinking. There is one drawback of this procedure that requires a further refinement. Due to the effect of repeated testing mentioned above, students who have completed at least four tests are considerably more likely to be experts than students completing their first test. Hence, the best estimates for the longitudinal prevalence of A1 and A in Table 2 are based instead on the percentages of experts in students' first tests in Grade 6 and 10 given in Figure 1a.

Table 2
Longitudinal Prevalence of the Coarse and Fine Codes over Two School Levels

Coarse Codes	Fine Codes	School Levels Primary (n=333)		Secondary (n=682)	
A		40*		80*	
	A1		30*		70*
	A2/A3		17		26
L		71		21	
	L1		62		12
	L2		19		7
	L4		17		5
S		35		28	
	S1		17		10
	S3		16		17
	S5		9		10
U		44		28	
	U1		44		28
	U2		1		4

* adjustments down have been made to compensate for effect of repeat testing

Table 2 provides a different perspective on the importance of these codes to teachers and researchers. For example, while Figure 1b suggests that L behaviours are really only a concern for primary school teachers, Table 2 indicates that 21% of the secondary students tested as L at some time, so this remains an issue for the secondary school. Likewise, over a quarter of secondary students demonstrate between Grades 7 and 10 ways of thinking associated with non-expert A (A2/A3, 26%) and with S (28%). An example is that 17% of secondary students were involved in the code S3, where students chose incorrectly on equal length decimals (e.g. they say that 0.3 is larger than 0.4). This is a surprise to many people; early researchers (such as Resnick, Nesher, Leonard, Magone, Omanson, & Peled, 1989, and Sackur-Grisvard & Leonard, 1985) did not include such comparison items in their test, as it is hard to imagine any student who could answer this incorrectly. Stacey and Steinle (1998) provided interview evidence of two ways of thinking that result in these errors. Students using *reciprocal thinking* have made their choice by a loose analogy with the fact that 1/3 is greater than 1/4, while students using *negative thinking* have made their choice recalling -3 is greater than -4. While confusing decimals with negative numbers may seem unlikely, we believe that learning about negative numbers and index notation (e.g. 0.000003 = 3 x 10-6 which clearly links decimals with negative numbers) is a likely cause of the increased prevalence of S3 in Grade 8 (see Figure 1c). Stacey, Helme and Steinle (2001) explored underlying reasons, using the theory of embodied cognition. New learning therefore "attracts" students to S thinking, so they need an opportunity to clarify, contrast and sort out their ideas. Explicit discussion of the location of positive and negative decimals and fractions on the number line is a good start.

Other analyses (see Steinle & Stacey, submitted) show that students exhibiting S behaviour exhibit unexpected tendencies. As they get older, these students become more likely to persist with this behaviour; for example, 40% of S3 students in Grade 8 who complete a subsequent test, retest as S3. This is the most persistent code in secondary schools. Not only do the S3 students tend to persist in this code, even when they do move from S3, they are less likely to move to expertise than other students. S behaviour has therefore emerged as a major difficulty requiring attention in secondary schools.

The Paths in Student Learning

The longitudinal study has provided a unique data set on how individual students change and retain their ideas about decimals for up to three years. This information has supplemented interview evidence on the nature of the misconceptions and the reasons for the behaviours exhibited.

Above we commented on the high persistence of S3, especially around Year 8. In general, the misconceptions that are associated with interpreting the decimal portion as a whole number, referred to as *whole number thinking* (L1), and *reciprocal thinking* and *negative thinking*, (both coded as S3) are the most persistent (i.e. have the highest probability that a student will retest in the same code on their next test). Of the students who completed tests that indicated S3 or L1 thinking, about one in six responded to the test in the same way more than two years later.

The students who completed tests coded as U (unclassified) were more likely to become experts by their next test, compared with students who were exhibiting L or S behaviours. Such students are possibly using a combination of different ideas, and may feel confused as they complete the 30-item test; appreciating that they have something to learn is the best state of mind for a non-expert. Such students may be more receptive to teaching than those with strongly held beliefs.

The phenomenon of regression affected one in five students. In other words, of the students who completed one test as an expert (A1) and who completed additional tests, one in five students were unable to complete a later test as an expert. The over-representation of certain codes in regression (in particular A2 and A3) confirms that there are considerable numbers of students who are using incomplete algorithms to compare decimal numbers. Items such as the comparison of 0.45 and 0.453 (where one number is a truncation of the other) are the most likely to detect students who are using an incomplete algorithm. For example, students trying to use the *left-to-right digit comparison* algorithm will run out of digits to compare and do not know that the space at the end of the shorter number can be replaced by a zero for continuation of the algorithm. Similarly, students who compare by just looking at the initial digits to the right of the decimal point, or by rounding to two decimal places (as one would with dollars and cents) will not be able to order these numbers. When an algorithm fails to provide a definite solution, students might guess or they might resort to a *latent* misconception (i.e. L or S behaviours). This confirms that rather than students moving away from these misconceptions, they are often retained but hidden behind algorithms and procedures (which, to a casual observer, indicate understanding). In other words, some students are receiving teaching that is *covering over* rather than *overcoming* misconceptions. In fact, uncovering the extent of this problem has been a major contribution of this study to the decimal misconceptions literature.

Evaluating and Improving the Decimal Comparison Test

The longitudinal study has been based on data collected using DCT2, a 30-item version of a Decimal Comparison Test. This is a test with great strengths and some clear weaknesses. One strength is the ease of administration: about 10 minutes to administer to a whole class, with minimal reading demand and one page of paper per student. Coding at the coarse level (A, L, S and U) is easy and can provide a classroom teacher with adequate information to undertake remedial work. Another strength of the test is that it is soundly based on research, as there is a long history of the use of comparison items as being particularly revealing (e.g. Swan, 1983).

Research has enabled close definition of the actual items in each group, so that the 12 fine codes can be allocated to students' tests with confidence. For example, the test contains 5 Type 1 and 5 Type 2 items. Students with L behaviour are expected to choose incorrectly on the Type 1 items (4.8/4.63, 0.5/0.36, 0.8/0.75, 3.92/3.4813 and 0.37/0.216) and correctly on the Type 2 items (0.75/0.5, 7.942/7.63, 2.8325/2.516, 5.736/5.62 and 0.426/0.3). Students with S behaviour exhibit the opposite behaviour (correct on Type 1, incorrect on Type 2) and A students get both groups correct. Steinle and Stacey (2003b) analysed 3531 tests completed during 1997. The most common score on the Type 1 and Type 2 items was (5,5) and the next two most common scores were (0,5) and (5,0). The other 33 pairs of scores attracted only 28% of the tests. This striking clustering is one demonstration of the validity of the diagnoses that can be made.

There are clear limitations to DCT2. Firstly, the structure of the test; it is composed only of comparison items, and does not test the wider domain such as the ability to perform operations, solve contextualised problems or other probing tasks, such as insert a number between two given numbers. We know that it is possible to complete the test without an understanding of decimal place value, just by accurately following a rule; so all measures of expertise are over-estimates of "true understanding". The strength of the DCT is in identifying erroneous thinking and it is surprising that such a simple test can do this so effectively and identify so many misconceptions. Intriguingly, it does not identify them all well. For example, students with *reverse thinking,* a misconception where numbers are effectively read backwards (so that 0.123 is read as something like 321 thousandths) seem to be frequently found in interviews but were found rarely with this test. We explain this by proposing that the format of the test is not conducive to making this error; there are features of any context that lead students to make certain errors and to avoid others. DCT2 shares these limitations with most other tests to varying degrees: its simple structure simply makes the limitations more evident.

Secondly, the items within DCT2 could be expanded to enable diagnosis of further misconceptions and to separate the several ways of thinking that lie behind some of the 12 codes. Another Decimal Comparison Test (DCT0) has included items such as comparison of 0.6 with 0, which generated quite unexpected responses. Stacey, Helme, Steinle, Baturo, Irwin and Bana (2001) found that 13% of a sample of over 500 pre-service teachers from four Australian universities chose 0 as larger than 0.6. Many of these students only made errors on the three items involving a comparison with zero; in other words, they would have been classified as experts (A1) using DCT2. Hence, students who believe that 0.6 is less than 0 inflate the prevalence of expertise reported above.

Another improvement is to include the option of choosing equality. This allows additional types of items such as comparison of 0.8 with 0.80 and 3 with 3.0 (which are equal) as well as 0.7 with 0.07 (which some students consider to be equal), which in turn allows better diagnosis of thinking. Steinle and Stacey (2001) demonstrated that this is simple and effective and diagnoses other difficulties. The results of the study has emphasised that zero is both a very difficult number and a very difficult digit, so these items can be a valuable addition.

Finally, Steinle and Stacey (2003a) identified two factors in the comparison items (number of decimal places exceeding two, and whole number portion being zero) that had a secondary but measurable effect on students' choices. A better Decimal Comparison Test would be obtained by splitting the Type 1 items listed above into two groups, based on whether the pairs of decimals were greater or less than one, and ensuring that at least one number in every pair contains more than 2 places. The resulting increased homogeneity of items would be expected to decrease the number of unclassified students. All of these improvements relate to a fundamental observation about diagnosis. To discover how students think, one needs to ask the right, probing questions, but to find out what questions these are, one needs to know how students think. Good research in mathematics education breaks into this cycle.

Not only do most students answer consistently on the items in the types within the tests, this study has extensive evidence that (unfortunately!) they tend to answer consistently from one test to another. This confirms the reliability of this test. Future research should used improved versions of the test, but the use of DCT2 in the longitudinal study has provided valuable results. The simplicity of this test and its power to diagnose misconceptions has produced a very positive reaction from teachers who have used this test in their classroom. In summary, provided its limitations are understood, the Decimal Comparison Test is a very powerful instrument for researchers and teachers.

Implications for Teaching

Space does not permit a thorough discussion of the implications of this study for teaching practice. However, a few central points can be made. Of major concern are the low rates of expertise (especially as we note that they are inflated) at every grade up to Grade 10 in this study and probably also beyond it. Students who do not know that 0.453 is near 0.45 but a little bigger, or students who think that 0.2 is near 0.3 but a long way from 0.21345, cannot make sense of the mathematics they are being taught. Moreover, the study has demonstrated that many students persist with misconceptions as they move through school; see for example, Steinle and Stacey (2003b). Hence, it is clear that the normal teaching that most students receive is inadequate to remove such misconceptions. This is particularly disturbing when evidence seems to show that, for a reasonably sized group of students, just a little targeted teaching can make a big difference. This is seen in the longitudinal data, where having been tested before resulted in 10% more students testing as experts, as well as in our teaching studies (e.g. Helme & Stacey, 2000, and Stacey & Flynn, 2003). Peled (2003) demonstrated the increased effectiveness of teaching that takes students' misconceptions into account.

We commented above that much teaching seems to cover over misconceptions, rather than overcome them. There are many examples. One is the "careful" approach adopted in many textbook series of considering only tenths one year, hundredths the next and maybe thousandths the next, and never drawing them together. Another example is the use of rules to cover over misconceptions, and it is the unusual items that reveal when there is a lack of solid conceptions supporting them. For example, students who learn to compare 4.3 with 4.37 by "adding zeros and then comparing 30 with 37" might not be receiving any lasting teaching if they do not integrate this with the fact that 3 tenths is equal to 30 hundredths. Indeed, discussions of 30 and 37 may reinforce the misconceptions that involve students treating the decimal portion of a number as a whole number. This study found that these misconceptions are the hardest for students to leave. Teachers need to be aware that always rounding the result of a calculation to two decimal places can reinforce the belief that decimals form a discrete system and that there are no numbers between 4.31 and 4.32, for example.

The decimal misconceptions have a wide range of causes including inadequate instruction, deep interactions between the way the mind works and mathematical content, misremembering taught rules and drawing false analogies. The remedy for all of these misconceptions is basically the same: to expose the underlying place value structure of the number system and to make explicit connections between its many different facets. The benefit for the teacher of knowing about misconceptions in general and a given student's misconception in particular, however, is to be able to select items that will demonstrate to the student(s) that there is something they need to learn, and to identify the connections (and non-connections) between ideas that need to be made explicit to them.

References

Brekke, G. (1996). A decimal number is a pair of whole numbers. In L. Puig & A. Gutierrez (Eds.), *Proceedings of the 20th conference for the International Group for the Psychology of Mathematics Education* (Vol. 2, pp. 137-144). Valencia, Spain: PME.

Helme, S., & Stacey, K. (2000). Can minimal support for teachers make a difference to students' understanding of decimals? *Mathematics Teacher Education and Development, 2*, 105-120.

Peled, I. (2003) Improving decimal number conception by transfer from fractions to decimals. *Proceedings of the 27th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 1-6). Hawai'i: PME.

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. (2003). Evaluating an adaptive computer system for teaching about decimals: Two case studies. In V. Aleven et al (Eds.) *Artificial Intelligence in Education AIED2003 Supplementary Proceedings.* (pp. 454-460) Sydney: School of Information Technologies, University of Sydney.

Stacey, K., Helme, S., & Steinle, V. (2001). 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., Steinle, V., Baturo, A., Irwin, K., & Bana, J. (2001). Preservice teachers' knowledge of difficulties in decimal numeration. *Journal of Mathematics Teacher Education, 4*(3), 205-225.

Stacey, K., & Steinle, V. (1998). Refining the classification of students' interpretations of decimal Notation. *Hiroshima Journal of Mathematics Education, 6*, 49-69.

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, Sydney, Vol. 2, pp. 434-441). Sydney: MERGA.

Steinle, V., & Stacey, K. (2003a). Exploring the right, probing questions to uncover decimal misconceptions. In L. Bragg, Campbell, C., Herbert, G. & J. Mousley (Eds.), *Mathematics, education, research: Innovation, networking, opportunity* (Proceedings of the 26th annual conference of the Mathematics Education Research Group of Australasia, Geelong, Vol. 2, pp. 634-641). Sydney: MERGA.

Steinle, V., & Stacey, K. (2003b). Grade-related trends in the prevalence and persistence of decimal misconceptions, *Proceedings of the 27th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 259-266). Hawai'i: PME.

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 http://timss.bc.edu/timss1999i/pdf/t99math_items.pdf