Evaluating an adaptive computer system for teaching about decimals: Two case studies.

Kaye STACEY and Jennifer FLYNN[1]
Department of Science and Mathematics Education,
University of Melbourne, Victoria 3010 Australia

Abstract. This paper reports an evaluation of an adaptive computer system designed to diagnose and improve students' conceptions of decimal numbers. The system provides some simple computer games and a test for students aged about 10 to 14. A Bayesian net is used for diagnosis, prediction and sensitivity analysis. Field testing with 25 students who had substantial misconceptions after classroom instruction demonstrated very good results, with 40% attaining expertise and 28% showing some improvement. Two case studies demonstrate the complexity behind these findings and ways in which the system could be improved. One student was mis-classified by the system, but learned very successfully. The second case study demonstrates successful classification but with poor learning. Controlling the presentation of a range of tasks which highlight many possible misunderstandings seems to be the most critical attribute of the system.

1. Introduction

This paper reports an evaluation of an adaptive computer system designed to diagnose and improve students' conceptions of decimal numbers. The system provides students with four simple computer games and a test. It uses students' responses to diagnose their (mis)-understandings and provides a basic level of teaching through visual scaffolding and feedback on answers. The computer system is said to be adaptive because it uses the current and previous responses of the individual student to select which items within the games are subsequently presented. The system uses a popular artificial intelligence technique for reasoning under uncertainty called a Bayesian network [1]. In this domain, there is uncertainty about what a student "really thinks" about decimal numbers and also whether the answers that they give to any item "really" reflect their understandings, or whether they are due to extraneous factors such as misreading, hitting a wrong key, being disturbed or distracted when answering or simply guessing. A computerised system which sets out to diagnose and teach students has to take these uncertainties into account. The Bayesian network manages uncertainty by assigning probabilities to classifications and predictions.

The paper presents the results of an evaluation and two case studies of students using the system. The first sections give a very brief overview of the domain and the system, which has been described more fully in [2] and [3]. Evaluation of the system has given promising results, indicating that the system can provide a much-needed tool to supplement normal teaching. The case studies serve to illustrate the mechanisms responsible for the learning improvements and their complexity and to raise issues about the design and implementation of adaptive learning systems such as this.

2. Brief overview of system

The purpose of the adaptive system is to provide a tool which can be used in classrooms to improve students' understanding of the meaning of decimal notation. Our estimate [4] is that only about two thirds of students aged 15 fully understand the size of a number such as 8.1234 and the value of its various components, which is a serious disadvantage in a world of metric measurement and digital displays. A major obstacle is that this is a field where misconceptions abound. Students are confident of incorrect interpretations which produce correct answers in restricted circumstances. As a consequence, many students hold erroneous ideas about decimal notation for years through normal class instruction. We therefore designed a computer system which could be used by individuals and which could address their misconceptions principally by provoking cognitive conflict, upsetting students' false confidence and making them receptive to new ideas. The computer system consists of four computer games, each with minimal teaching and help, and a test, linked by a Bayesian net.

The system models students primarily by the misconceptions which they hold. These are detailed in [5] and elsewhere in the mathematics education literature. Most misconceptions are based on false analogies, often embellished by isolated facts such as "zero doesn't matter" (so 3.01 = 3.1). The most common decimal misconception amongst young students is what we call LWH, where students interpret the decimal part of a number as indicating a number of parts but of unspecified size. For example, they will see 9.123 as 9 ones and 123 parts and 9.023 as 9 and 23 parts and so will think these are both larger than 9.8 (9 ones and 8 parts). Alternatively, this basic misconception may be embellished by the isolated fact that "zero makes things small", so that 9.023 may be regarded as smaller than 9.8, even though 9.123 is regarded as larger. This second misconception is called LZE. The misconception SDF is less common, affecting about 5% of students throughout most of the school years. These students interpret 9.12 as 9 ones and 12 thousandths (correctly) but assume that since thousandths are smaller than hundredths, 9.123 would be smaller than 9.12. These students therefore effectively select shorter decimals as larger numbers. Some students (AMO) interpret decimals by analogy with money. They correctly interpret one and two decimal place decimals, but have no meaning for the decimal places beyond the second. They may, for example, be unsure whether 2.3456 is smaller or greater or equal to 2.34. There are many other misconceptions, which can be diagnosed by looking at patterns of answers to a 24–item "Decimal Comparison Test" (DCT) where students are asked to select the larger decimal from pairs of decimal numbers (e.g. 4.8 and 4.63). Students who can perform this task correctly are given the classification ATE (apparent task expert). This name reflects our observation that expertise on our aspect of decimal knowledge does not necessarily imply expertise on others. Students' knowledge is often fragmentary and highly disconnected, a theme to which we will return in the second case study.

The Bayesian net uses responses to an on-line DCT (many items in common with the above) to diagnose students' misconceptions. It also contains nodes which record and track students' responses to a variety of other tasks, which arise during the games. The essence of the game *Flying Photographer* is to identify the position of a number on an unmarked number line from 0 to 1 (these endpoints can be changed at higher levels). The game aims to encourage estimation, rather than accuracy, so that the position of the required point needs to be indicated by 'taking a photo' when an 'aeroplane' flies over the required point. The teacher can adjust the speed of the aeroplane and the required accuracy. For *Flying Photographer*, information is held in the Bayesian net about the ability of the student to place long numbers which are large (e.g. 0.989999) or small (e.g. 0.11586), numbers with zero in the tenths place, and short numbers which are large (e.g. 0.9) or small (e.g. 0.l2). These attributes are therefore part of the student model. In playing the game *Hidden Numbers*, students have to guess which of two hidden decimal numbers is the larger, by gradually opening doors to reveal the digits one at a time. There are penalties for opening unnecessarily many doors and rewards for guess correctly. If all the doors are open, playing *Hidden Numbers* is the same as doing the DCT, but the method of playing is revealing. Therefore the

Bayesian net records information about the efficiency of the order of opening the doors and these nodes too become part of the student model. Space precludes description of the two other games *Decimaliens* and *Number Between*, but they also contribute to the student model and are linked into the Bayesian net.

The high level system architecture is shown in Figure 1. The adaptive Bayesian net receives information from the population (prior probabilities) and can be tailored with information about the student, such as age, misconception, prior experience with system. The on-line DCT provides the first source of information for the BN and progressively results from the games are used as evidence to update the probability estimates for diagnosis of misconceptions. This information is fed to the system controller module, together with the specified sequencing tactics, to select items to present to the student, to decide whether help should be presented to the students and possibly to decide whether the student should move to another game. The arrows into the student area in Figure 1 indicate help, feedback and instruction. Because the games contain only minimal instruction and feedback, in keeping with the computer game genre, they are intended to be used within a normal classroom context and so the teacher and classroom instruction are featured here as part of the system architecture. The teacher model, consisting of decisions of when to offer help, the nature of that help and the sequencing of the items which are presented to students, is discussed in [2].

Figure 2 shows a schematic illustration of the BN representation of the student model. The central node represents the misconception. Within the sets of nodes for each game there is little connection. For the *Hidden Numbers* game, there are nodes which represent each type of comparison (for example 9.8 / 9.123 is of a different type to 9.8/9.023 since different groups of students are right and wrong). These nodes are shared with the DCT and are represented at the base of Figure 2. Additionally, there are nodes which receive evidence of the number and the efficiency with which the doors are opened. A root node reflects a player's overall game ability. Further details are given in [2].

Given the model and given evidence about answers to one or more items, the Bayesian belief updating algorithm performs diagnosis by calculating the probabilities that a student has each of the misconceptions. Changes in these beliefs are propagated within the network to perform prediction and the updating algorithm calculates new probabilities that the student will get items of other types right or wrong. After each evidence is added, the network is progressively updated, allowing changes in students' ideas to be tracked.

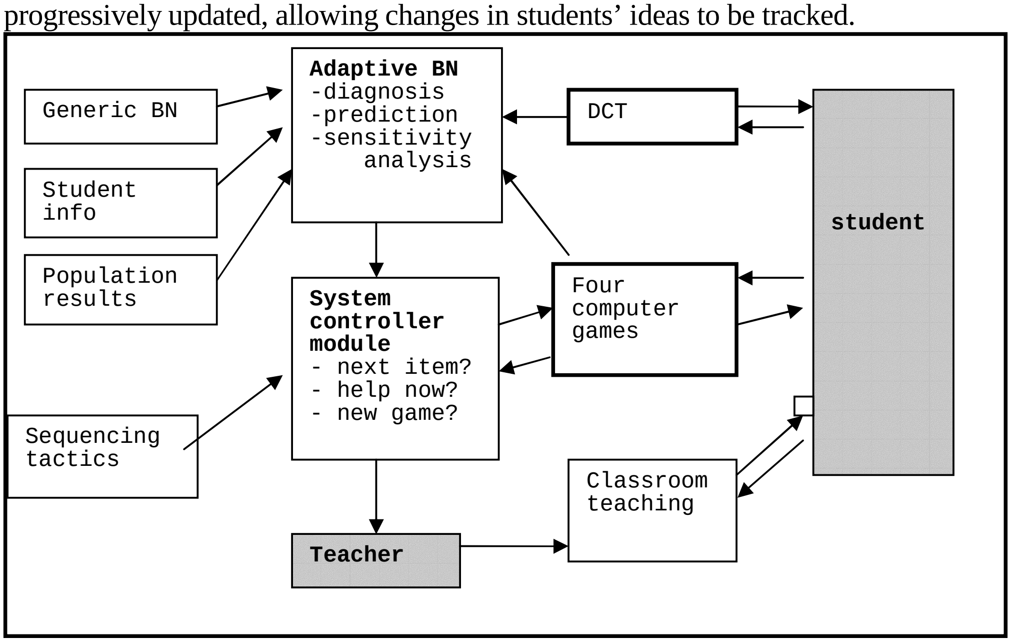

Figure 1. Schematic Depiction of System Architecture

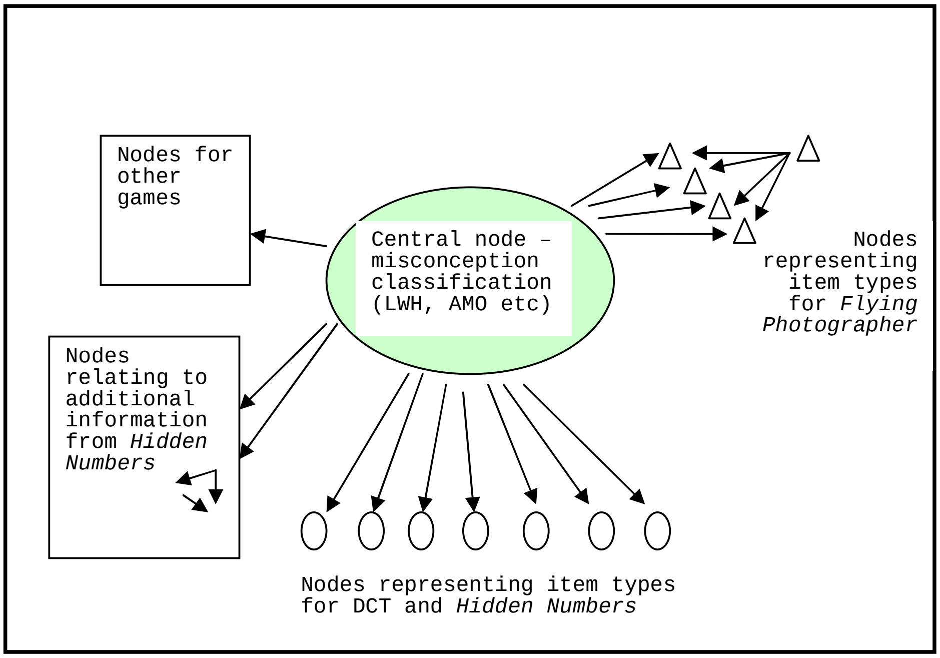

Figure 2. BN representation of the student model

3. Case studies

The complete system has been field tested in two schools with 25 students in Grades 5 and 6 (ages approx 10-11 years) who were selected because they had misconceptions persisting after normal school instruction. Nearly all students worked with a partner, where possible one with a similar misconception about decimals. Sessions lasted for up to 30 minutes, with an adult observer, normally making no intervention although occasionally asking a question. The observer recorded conversations, which were linked to the on-line records and subsequently analysed to see where students learned and where they missed learning opportunities. Long term conceptual change was measured with a delayed post-test after approximately 6 weeks. Of the 25 students, 10 tested as experts on the delayed post-test, which indicated significant progress. Seven students demonstrated improvement, without attaining expertise and eight retained their original misconception. These results are pleasing in the light of the findings [4] that in normal classroom settings, students often persist with misconceptions over some years. The transcripts revealed that some students learned from the visual scaffolding that was provided by the help screens, but active teacher intervention seems required for most students to benefit fully from these. Very frequently, students learned by observing their partners, but working with a partner was not always beneficial, as in some instances one partner dominated and took away learning opportunities from the other. Two cases which reveal the working of the BN in interesting ways are now presented.

3.1 Case study 1: Effective learning in the context of mis-diagnosis

Mick's case illustrates effective learning in the context of mis-classification by the system. In both the May and August paper 24-item DCT, Mick's answers fitted the SDF pattern (see above) but in October, one month after his sessions with the computer, all the DCT items were correct. At the first session with the computer in early September, Mick demonstrated again his SDF misconception that shorter decimals are larger numbers. On beginning the computer test, he explained to the observer that 5.671 is larger than 5.2378 "*because the* *smaller behind the decimal point the bigger, yeah, the less amount [un]til the whole number*" and "*because it takes a minimum amount of numbers to get to the next hundred (thousand)*". This and other explanations lead us to interpret Mick's thinking thus: he knows that the 671 refers to thousandths and hence that 671 needs to be built up to 1000 to get to the next whole number. He also knows that 2378 needs to be built up to 10000 to get to the next whole number. Since more needs to be added to 2378 to get to the next whole number than needs to be added to 617, 5.2367 is less than 5.671. Mick is therefore conscious of the place value to the extent that he sees 671 as being thousandths in some sense and 2378 as being ten-thousandths, which is typical of SDF thinkers. (Note that whilst Mick's answer for the comparison of 5.671/5.2378 was correct, his reasoning was not.) In all but one item in May, August and on the all but one item on the shorter computer test, Mick's answers were consistent with this reasoning. At this stage, Mick does not understand the meaning of decimal notation, although he has some of the components of this understanding in place.

Initially, the Bayesian net assigned to Mick the *a priori* probabilities of a student of his grade level having each of the tracked misconceptions (43% ATE, 3% AMO, 3% SDF etc) and after the 14-item computer test, these changed to 39% ATE, 26% AMO and 3% SDF. The researchers believe that Mick should be diagnosed here with a high probability of being SDF, rather than AMO or ATE. However, in the computer test, one of the critical items for SDF involved only short decimals (compare 4.6 and 4.84) and Mick applied unconnected information to this item (perhaps coming from money knowledge). He commented, "*If there's just two decimal places you add a zero and 84 is bigger than 60*". Whereas the paper test had 5 items of this type, the computer test had only two (one short –correct - and one longer - incorrect), the net did not increase the estimated probability of SDF above the *a priori* level.

For the rest of this session, Mick chose to play the game *Decimaliens*, which tests verbal and fraction knowledge of place value. Evidence of this basic knowledge of place value was accumulated in a node and the probability that Mick understood these aspects of place value rose from the prior of 87% to 100% over seven games. These observations are inputted to a separate section of the Bayesian net and do not affect beliefs about his SDF misconception. The observer saw that Mick played this game with confidence and judged it reminded him of the need to consider place value when interpreting decimal notation.

Two weeks later, Mick began a second game, *Flying Photographer*, which proved to be a powerful and exciting learning tool. He played 37 times. The first item selected was to place the number 0.5101. Consistent with his "real" SDF misconception, Mick regarded 0.5101 as a small number (because of its length) and hence guessed near zero (0.354, then 0.046). The program then provided feedback, showing sad-faced triangles at 0.354 and 0.046 and a happy-faced triangle at 0.5101 along with a scale marking the points 0.1, 0.2, 0.3 … 0.9 superimposed on the number line. The observer noted that Mick observed this carefully and thereafter placed numbers correctly according to the tenths digit whenever the help (i.e. the scale marked in tenths) was showing. This seems to have been the significant moment of learning for Mick. Later, he commented to the observer that *Flying Photographer* had taught him "where decimals would be on the number line". This learning stayed with Mick throughout the rest of the study, and he applied the strategy of comparing tenths consistently and successfully in other games (*Hidden Numbers*) and at the delayed post-test.

There was immediately a second significant learning moment for Mick when he was presented with the item 0.0052. This item was presented with the tenths scale visible on the screen, since the decision had been made to provide help for the item immediately following any item which was answered incorrectly. This decision followed experiments [6] which showed that students were reluctant to request help but missed learning opportunities when they did not have it. Mick looked carefully at the tenths scale and took his photo at 0.534. This was consistent with Mick's belief that he should ignore the initial zeros, expressed to the observer in two items in the initial computer test. Mick had decided that 1.067 = 1.67, adding the comment "*because in 1.067 you just take away the zero*" and deciding that 1.518 > 1.0631 "*because 631 is bigger than 518 . . . and the zero is nothing*". Mick had interpreted 0.0052 as being the same as 0.52. This possible misconception about decimals had been missed by the initial analysis and so is not identified by the DCT and has not been built into the Bayesian net. Subsequent analysis showed that it was held by half of the students in this sample and therefore should be added to the system. When the game displayed the correct answer, Mick again very carefully examined the visual feedback, and thereafter always placed decimals with initial zeros below 0.1. His belief that "*zero is nothing*" (and hence should be ignored) changed so that at the delayed post-test, he explained to the observer that 8.69 is larger than 8.069 "*because there is no tenths*", not because zero was nothing. This incident underlines the importance of using the widest variety of item types.

In summary, Mick was mis-diagnosed by the system and therefore items that were expected to be easy were hard for him. However, his thoughtful reaction to the cognitive conflict provided by the feedback enabled him to learn effectively and to retain what he learned. The diagnosis could be improved by including additional misconceptions and by adding more items to the computer DCT. The sequencing of the items did not matter for Mick, but it was important that he be given items of a type which challenged him.

3.2 Correct diagnosis, without learning

Jill and Jack worked together on the computer games. They were matched according to identified misconception, both being classified as LWH (see above) on the May and August paper DCT. Jack's answers also identical on the October paper DCT, whereas Jill's October test showed that she had learned that a decimal with 0 in the tenths column was small (i.e. she continued to choose 4.8 < 4.63 but she now knew that 3.073 < 3.72). Thus, according to the by-hand testing, Jill and Jack made very little progress, through both the classroom instruction and their one computer session. The session began with the 14 item computer DCT, after which the Bayesian net classified them as LWH with 94% probability. They then began to play *Flying Photographer*, and the net provided them 13 items of which 7 were correct on the first trial (some of these with visual scaffolding of the tenths scale). They needed three trials on three item types, including both the ones where the net predicted the greatest and least probability of success for LWH thinkers. At the end of this game, the net beliefs of classification were LWH (56%) and ATE (13%).

Jill and Jack then played 13 games of *Hidden Numbers* and were highly successful from the start, opening doors for tenths first. At the end of this game, the net beliefs of classification were LWH (0%) and ATE (72%). They then played *Number Between* successfully, resulting in net beliefs of classifications of LWH (0%) and ATE (94%). They then immediately took the computer DCT again, and completed it exactly according to the LWH pattern, and as noted above, this was repeated (with only a minor variation from Jill) in October.

In the case of Jack and Jill, the net performed well, classifying them in the same way as an observer would and recording increasing confidence in their expertise as they demonstrated ability to perform on a variety of tasks. However, the delayed post-test indicates that they did not change their ideas. Certainly, the observer noted that when they made mistakes they just tried again, rather than looking carefully at the feedback provided and there was no discussion to report. Possibly Jill and Jack have highly compartmentalised knowledge, not making connections from one part of their understandings to another. Perhaps they hit upon (or overheard?) successful strategies which they followed as rules, rather than analysing the meanings. The case illustrates the uncertainties involved in inferring ability to perform one task from ability to perform another, but without cross-inferences such as this being built into the BN, surely information is wasted. The case also illustrates the general difficulties of inferring understanding from behaviour that beset human observers and computers.

4. Conclusions

The overall results for the field trials are very pleasing, since students as a whole made good progress with generally independent work. In both case studies, the performance of the BN was explicable. Mick was wrongly diagnosed, but improvements can be made to avoid this in the future. Jack and Jill were diagnosed sensibly and a human observer would also have been "fooled". Neither of the case studies supports the notion that it is important to use the BN to carefully sequence the items presented. Mick's learning occurred with items presented in an "inappropriate" order because of the wrong diagnosis and Jack and Jill did not seem to learn, despite the order being as prescribed. Evidence from cases not presented here indicates that the sequencing may be more effective if items which are expected to be hard are given more frequently. Mick's case, however, emphasises the importance of presenting students with a wide variety of different item types and immediate feedback. Even with the seemingly trivial task of comparing two decimal numbers, there are many cognitively different types of items and different items will uncover different parts of the knowledge of decimal notation. Here is a clear benefit of the computer. Students' understanding is tested and extended in a way which teachers cannot do, without extremely detailed planning and extreme mental effort. Making mistakes in the games does often provoke cognitive conflict in students, which can sometimes be resolved with minimal feedback. The system, however, does not provide the thorough instruction in place value that is necessary in some cases to produce real understanding for young students. This has to come from a teacher.

References

[1] Pearl, J.: *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann, San Mateo Ca, (1988)

[2] Stacey, K., Sonenberg, E., Nicholson, A., Boneh, T., Steinle, V.: A teaching model exploiting cognitive conflict driven by a Bayesian Network" . Paper presented at UM2003 Ninth International Conference on User Modelling (UM2003) University of Pittsburg, Johnstown, USA. June 22 – 26, 2003.

[3] Nicholson, A., Boneh, T., Wilkin, T., Stacey, K. Sonenberg, E., Steinle, V.: A case study in knowledge discovery and elicitation in an intelligent tutoring application. In J. Breese & D. Koller (Eds) *Int Proc of the 17th Conf. On Uncertainty in AI,* Seattle: Morgan Kaufmann (2001) 386 – 384.

[4] Steinle, V. and Stacey, K.: Understanding Decimals: The Path to Expertise. *Proceedings of the Twenty second Annual Conference of MERGA,* Sydney: MERGA. (1999), 446 – 453.

[5] Stacey, K., Steinle, V.: Refining the classification of students' interpretations of decimal notation. *Hiroshima Journal of Mathematics Education* 6 (1998) 49 – 69.

[6] McIntosh, J., Stacey, K., Tromp, S. & Lightfoot, D.: Designing constructivist computer games for teaching about decimal numbers. In J. Bana & A. Chapman (Eds) Proc. 23rd conf Mathematics Education Research Group of Australasia, Fremantle (2000) , 409 - 416.

---

[1] ACKNOWLEGEMENT

The adaptive computer system described in this paper was created by Liz Sonenberg, Kaye Stacey (Melbourne) and Ann Nicholson and Tal Boneh (Monash University, Australia). Vicki Steinle and Sally Helm (Melbourne) worked on the students' conceptions of decimals. Smaller contributions from numerous colleagues and student are also acknowledged. The research has been partially funded by the Australian Research Council and the University of Melbourne.