The Psychology of a 'near-miss' in the 4-Digit Lottery: A Spreadsheet Simulation

Tay Eng Guan
egtay@nie.edu.sg

Mathematics and Mathematics Education Academic Group
National Institute of Education, Nanyang Technological University
1 Nanyang Walk, Singapore 637616

Abstract. For low cost and more flexibility and choice, instructional materials can be prepared by using a spreadsheet. This paper shows how the software *Near Miss* written on a spreadsheet is used to simulate the popular gambling game 4-D so that students may investigate probability and a little psychology.

1 Introduction

The lack of affordable customised software frustrates the implementation of IT-based instruction in the schools. Teachers continue to rely heavily on mathematics worksheets prepared using a word processor, printed out on paper, worked on using pencils and scored with red ink. Low, Tay and Chen [5, 7] proposed that the many features of the ubiquitous spreadsheet be exploited for the benefit of both teachers and students. Their papers describe the developmental stages achieved and the subsequent progress to the production of an interactive game *Football Fractions* [6] on CD.

This paper introduces another software *Near Miss* [1] which was written using Visual Basic on Microsoft Excel. While *Football Fractions* utilised the graphic functions of the spreadsheet to animate the concepts of a fraction as 'part of a whole' and as a ratio for the primary school level, *Near Miss* leverages on the random function to introduce probability at the secondary level.

2 Motivation for using a spreadsheet

The following discusses the relative advantage of using spreadsheets to plan mathematics worksheets for students over using other computer-based tools. It further asserts and demonstrates how a spreadsheet can become an interactive environment in which students can learn the basics of mathematics. The motivation of using a spreadsheet is presented in five questions which represent our thinking process and indicate how the decision of using a spreadsheet was derived.

Why not use word processors? Our underlying assumption and philosophy is the focus of a contructivist approach to learning (see Davis [3], Glasersfeld [4]). Word processed worksheets are generally teacher-centred and inflexible in engaging students in the process of constructing knowledge, whereas the use of spreadsheet allows teachers to create a learning environment that is interactive and student-centred.

Why not use other available educational CDs? One of the challenges today is to empower teachers to produce interactive instructional materials painlessly for classroom use. This is different from using commercially produced educational CDs, which by virtue of being a commercial product, is primarily for student use. As such, commercially produced CDs cannot fit the dynamic demands of teaching in the actual classroom. Each CD can only fit a restricted number of modes and situations. The CD programme cannot be altered to suit the need of a particular class at a specific instant. They can only be played by their own rules.

Why not develop the concept but farm out the production to commercial companies? Prices quoted were prohibitively high – at least $10,000 for a simple production.

Why not use specialised mathematics software? Commercial software for producing customised materials is expensive and not readily available. Besides, very few teachers are familiar with these packages. Although a spreadsheet is not superior to more specialised software, its power for large scale IT based teaching lies in its availability and simplicity.

Why not use presentation software? Presentation software such as PowerPoint is more suited for dissemination of information. However, it has no feature that allows the user to create an interactive platform. On the other hand, the spreadsheet has many of such features. In particular, the logical functions (IF, OR, AND, etc.) and macros can be utilised to create interactive materials. There is much user-control as the student can manipulate the items by keying in his input or clicking on some buttons. The animation that results helps in visualisation. The interactive nature of the materials also lends itself well to the constructivist approach. The teacher is not constrained by the availability of commercial software as he is able to 'program' his own worksheets according to the topic and level of difficulty that he desires. With the addition of Visual Basic programming, interactive educational games can be developed within the spreadsheet. When a teacher has mastered the art of using spreadsheets for classroom teaching, he or she has acquired an integral part of teaching skill, which is more powerful than the traditional chalk and board, and more flexible than the excessive use of worksheets.

3 The 4-Digit lottery and the probability of a 'near-miss'

The 4-Digit lottery, also known as 4-D, is a popular gambling game in Singapore. Punters can purchase in Big Bets or Small Bets. Each costs $1 per 4-digit sequence from 0000 to 9999. In Big Bets, the 4-digit sequence purchased is matched against 23 winning numbers (1st, 2nd, 3rd Prizes, 10 Starters, and 10 Consolations). In Small Bets, the 4-digit sequence is matched only against the top three winning numbers (1st, 2nd, and 3rd Prizes). For purchase of a $1 bet, a Big Bet win would be $2000 ($1^{st}$), $1000 ($2^{nd}$), $490 ($3^{rd}$), $250 each (Starter) and $60 each (Consolation). For purchase of a $1 bet, a Small Bet win would be $3,000 ($1^{st}$), $2,000 ($2^{nd}$), $800 ($3^{rd}$) with no Starter or Consolation prizes.

As a matter of interest, the first three prizes and the Starter prizes used to be based on a selected horse race which usually featured 13 horses. At first, 13 winning numbers were drawn either electronically or mechanically. Then, each of these numbers was assigned to a horse in the race. At the end of the race, the number assigned to the winning horse would be awarded the first prize, with the second and third awarded accordingly. The remaining 10 numbers were awarded Starter prizes which alluded to the rest of the horses which 'started' the race. A consolation prize is, as the name suggests, a small reward to salve the pain of missing the big prizes (and, I think, to encourage further attempts at the lottery).

To be aware of the evils of gambling requires a certain amount of numeracy and the secondary mathematics class is a good place to educate young minds. One can work with the class to calculate the expectation of a Big Bet (\$0.659) and a Small Bet (\$0.58) and thus show how heavily biased the lottery is.

However, gambling is more than knowing the probabilities of getting a prize; it is also the perception that one is inherently 'lucky' and likely to beat the odds if one were bolder in making a bet. Often, this perception of being successful if one were bolder is reinforced by the built-in aspects of the gambling game one plays. For example, in horse-racing, if one's second choice horse came in first (which is quite likely), one would think that if only a bet had been made on more than one horse, there would be some money to win. In the case, of the 4-D lottery, the perception of being successful if one were bolder is reinforced by one's chosen numbers almost matching the winning numbers. For example, if one were to buy the number 1234 (that being the combination of the number of horses one sees in a dream with the date of the dream 12th of March, say) and the first prize was 1235, one would scream, "Just missed!"

We shall estimate the probability of such an occurrence for a Big Bet. We first define a 'near-miss' as a 4-digit sequence that differs from one's chosen number by one digit or is a permutation of that number. Thus, 1235, 1334 and 4321 are examples of near-misses of 1234. An upper bound for the probability of a near-miss can be calculated if we assume that all the 23 winning numbers have all 4 digits different and that a near-miss of one number is not a near-miss of any other. Thus, each winning number has $9 \times 4 = 36$ near-misses that differ by one digit and $4! - 1 = 23$ near-misses that are permutations of it, the total number of near-misses being $36 + 23 = 59$. Hence, the number of near-misses for all 23 numbers would be $23 \times 59 = 1357$. Thus, an upper bound for the probability of a near-miss = $\frac{1357}{10000} = 0.1357$, which is slightly less than one in seven! In other words, in such an upper-bound situation, about one in seven persons will have a near-miss or a person buying six unrelated numbers will almost certainly (> 99% chance) have a near-miss. It is my hypothesis that this occurrence fuels the gambling impulse because it is thought to be uncommon when it is actually not so.

A lower bound for the probability of a near-miss can be calculated if we assume that all the 23 winning numbers are exactly the same and is of the form *xxxx*. Then there are exactly 36 near-misses for the draw. Thus, a lower bound for the probability of a near-miss = $\frac{36}{10000} = 0.0036$. This situation however is very unlikely to happen. On the other hand, the upper bound situation described earlier is much more likely to occur plus a few overlaps in near-misses. Thus, it is plausible that the probability of a near-miss is about 0.1.

4 Simulating the 4-Digit lottery and the probability of a 'near-miss'

The game is presented on a Microsoft Excel worksheet. There is a cell which records the amount of money one has in the 'bank'. Underneath this is a table in which the punter may record up to twelve 4-digit sequences and the wager for each of them. The main table shows the 23 prize winning numbers. The situation after a fourth draw is captured in the screenshot (Figure 1) below.

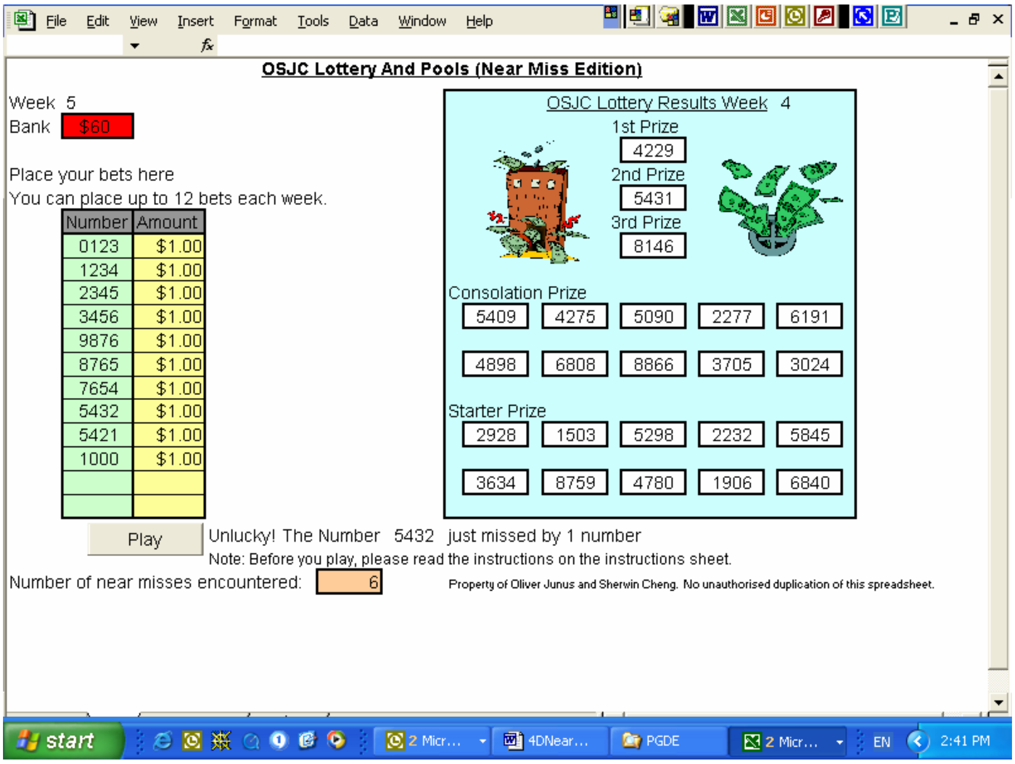

Figure 1

After placing the bets, the punter clicks on the button 'Play'. The macro will then do the following:

1. Generate 23 random 4-digit sequences for the 23 winning numbers.
2. Check if the amount in the 'bank' is enough for the wager. If yes, deduct the total wager from the amount in the bank, otherwise a message box will appear informing that the amount in the bank is insufficient.
3. Match the wagered numbers against the winning numbers, declare any winning numbers and add the prize money to the amount in the bank.
4. Match the wagered numbers against the 'near-misses' of the winning numbers according to a miss by being different in one digit and a miss by permutation. Declare any near-miss and count up the number of near-misses so far.

The simulation may be used for a teacher-centred discussion or it may be used as a student-centred discovery project using a suitable worksheet to guide the student. A simulation of 10 numbers on 170 draws resulted in 241 near-misses giving an estimation of the probability of a near-miss as 0.14.

The author recently conducted a lesson on the Teaching of Probability for a class of trainee teachers where the software *Near Miss* was used. The discussion was steered towards gambling and the teacher asked the trainees how many of them actually bought the 4-D lottery. Most of them answered in the negative but as these were adults with some of them in their thirties and beginning a second career, it wasn't surprising that some said that they did. Almost everyone agreed that they knew someone who was a regular punter.

The prize structure of the 4-D lottery was then described and a discussion began on which was better to buy - Big or Small. The trainees talked about the need to use probability to determine the answer but were a bit rusty until someone suggested 'expectation'. The expectations were duly computed and all could clearly see that Big was better. "And yet some people still want to buy Small," someone commented and another replied that this was because they were willing to risk a small amount of money for a chance to win big (since Small gave out bigger prizes though fewer). This led on to the calculation that the expectation of a payout in either version of the lottery was negative and the same observation that people were still willing to risk a small amount of money for a chance to win big. Some called this buying 'hope'.

An earlier version of the software which did not have the near-miss function, *4-D* [2], was then used to run a simulation of the 4-D lottery in the class. Trainees were given $100 in 'imaginary money' to wager. All they had to do was to write down on a piece of paper their chosen numbers with their bets beside them. Then the simulation was run, numbers checked, and profits and losses recorded. As it turned out, in the class of 25 trainees, most lost all of their 'money' after just 3 rounds and none showed a profit. We discussed again why anyone would want to continue playing such a game when one is almost sure to lose. In one of the rounds, I had asked them to check if they had any near-misses. Now, we discussed if this could be a reason why people continued to press on with the game. The response was that it was possible.

NearMiss was then run. This time, everyone contributed numbers, from dates to car licence numbers, and wagers in a common pool against the computer. We struck second prize on a car licence number but eventually went bankrupt. The count of near-misses was impressive and the realisation that it was so common registered clearly in the trainees 'Oh's and 'Ah's.

The class then went into a calculation of the probability of a near-miss in any 4-D draw with results as described earlier. The 'second career' trainee who had earlier replied that he frequently bought 4-D numbers was then asked if he had known earlier that the probability was so great. He replied that he did not with a look that suggested that this was something to think about more deeply. When asked if their friends who gambled on the 4-D would change their habit if they were told of the high probability of a near-miss, most laughed and said that they would not understand and would still carry on!

5 Conclusion

The potential for the production of software using the spreadsheet at the school or cluster level is immense. The teachers have the content knowledge. It used to be that commercial programmers had most of the IT savvy needed to convert concepts to software. Now, interactive instructional materials may be produced not on any high-powered software but on a simple spreadsheet by a suitably motivated teacher or team of teachers. The software *Near-Miss* was in fact produced by two trainee teachers in the National Institute of Education. All they needed was a mathematical concept, an interesting real-life observation and a basic lesson on producing interactive teaching materials on a spreadsheet.

References

[1] Cheng, S.S. and Junus, O.O., Near Miss [Computer software] (2000).

[2] Cheng, S.S. and Junus, O.O., 4-D [Computer software] (2000).

[3] Davis, A., Contructivism. In Davis, A. and Pettitt, D. (Ed.) Developing Understanding in Primary Mathematics, The Falmer Press (1994).

[4] Glasersfeld E.V., A Radical Constructivist View of Basic Mathematical Concepts. In Ernest, P. (Ed.), Constructing Mathematical Knowledge: Epistemology and Mathematics Education, The Falmer Press (1994).

[5] Low, S.K., Tay E.G. and Chen, D.T., If you can't find it, make it yourself!: Producing interactive instructional materials using a spreadsheet. In Margit Waas (Ed.), Enhancing Learning: Challenge of integrating thinking and information technology into the curriculum, Educational Research Association: Singapore (1999) 210-215.

[6] Specialised Pupil Programmes Branch, Ministry of Education, Singapore, Football Fractions [Computer software] (1999).

[7] Tay E.G., Low, S.K. and Chen, D.T., 4C software for the schools: Constructivist, Customised, CD and Cheap, Proceedings of 4th Global Chinese Conference on Computers in Education (2000).