Mathematics Teacher Education and Development 2000, Vol. 2, 121-135

Preservice Mathematics Teachers' Understanding of Sampling: Intuition or Mathematics

Jane Watson
University of Tasmania

This paper considers 33 preservice secondary mathematics teachers' solutions to a famous sampling problem, well known for confounding educated adults. Of particular interest is the use of intuition and/or formal mathematics in reaching a conclusion. The relationships of solution strategy to students' background in formal mathematics and to gender are also considered. Implications for teaching statistics at both the secondary and preservice teacher education levels are discussed briefly.

> A town has two hospitals. On the average, there are 45 babies delivered each day in the larger hospital. The smaller hospital has about 15 births each day. Fifty percent of all babies born in the town are boys. In one year each hospital recorded those days in which the number of boys born was 60% or more of the total deliveries for that day in that hospital.
>
> Do you think that it's more likely that the larger hospital recorded more such days, that the smaller hospital did, or that the two recorded roughly the same number of such days?

The hospital problem, as the above problem is usually named, began its life in the work of Kahneman and Tversky (1972) in their ground-breaking work on people's understanding of representativeness. They reported data from 95 undergraduate students with no background in probability or statistics, 50 of whom were asked the question in the above form and 45 of whom were asked which hospital had "less" days when 60% or more of the babies born were boys. For each form of the question the results were the same, with 20% of the students making the appropriate choice: the smaller hospital for "more" days with 60% or more boys and the larger hospital for "less" days with 60% or more boys. Fifty-six percent of each group said that the results would be "about the same for each hospital (i.e., within 5% of each other)" (p. 443). This exemplifies what Tversky and Kahneman (1971) described as the law of small numbers, that is, the belief that a small sample represents a population just as well as a large one. For the problem above this leads to the conclusion that the two hospitals recorded about the same number of days with 60% or more boys born.

Although there have been criticisms of the wording used by Kahneman and Tversky (1972) (e.g., Well, Pollatsek, & Boyce, 1990), this problem has been used repeatedly over the years to explore understanding of sampling and the variation associated with different sample sizes (cf. Well et al., 1990). These studies have been based mainly on tertiary students and school students have not often been asked the hospital question. Of 80 students in grades 5, 7, 9, and 11, and 18 prospective teachers specialising in mathematics, Fischbein and Schnarch (1997) found only one grade 9 who suggested that the smaller hospital would have the more extreme result. Their question included more explanation, for example the actual number of boys for each sample size that equalled 60% of births, but they observed the percent of students whose responses reflected the law of small numbers rise from 10% at grade 5 to 80% at grade 11, and 89% of the prospective teachers. Watson, Collis, and Moritz (1995), in a study of students' understanding of sampling in grades 3, 5, 7, and 9, found that of 12 students asked a similar question in an interview setting, none could argue from a correct basis. Watson and Moritz (2000) asked 59 students this problem in interviews and of these, only six recognised the smaller sample as more likely to have the extreme result and were able to give an adequate justification. The choice of the large or small sample often appeared random, with the reasons given for choosing the small sample also being given for choosing the large one. In a setting with the large sample from the city and the small sample from the country, students in grades 3, 5, and 9 chose the small sample "because there are more boys in the country" and students in grades 3 and 7 chose the large sample "because there are more boys in the city". Story-telling rather than an appreciation of sample size often dominated answers of younger students. Reasons given for saying the large and small samples had the same chance of having the extreme result included "because it's random" or "because percentages aren't definite numbers and so are the same for each sample".

Researchers studying people's appreciation of sample size in the context of the hospital problem, however, have not explored the understanding of people with a formal background in mathematics and the potential to call upon formal statistics to answer the hospital question in terms of the actual probabilities associated with the two samples having 60% or more boys. Fischbein and Schnarch (1997) did not document the mathematical background of the preservice teachers in their study but "none of the students had previously received any instruction in probability" (p. 97); hence it is likely this statement covers them. Anecdotal evidence (N. Windsor, personal communication, November 9, 1998) indicates that many senior secondary students, when given the hospital problem immediately after a unit on the normal approximation to the binomial distribution, do not recognise its applicability. Of those who do, many still have difficulty carrying out the process correctly to achieve the appropriate conclusion. A few, however, find it a straight-forward application.

Whereas perhaps not all preservice mathematics teachers would have been exposed to the theoretical mathematics necessary to calculate the probabilities in each hospital, at least it would be expected that they would have access to the basic mathematics to calculate percents, fractions, and basic probabilities based on equal likelihood (of births of boys and girls). For these people, it is hence of interest to observe what might be termed intuitive solution strategies versus strategies based on formal mathematical knowledge relevant to the problem. This distinction reflects the work of Fischbein and Grossman (1997), for whom "the term intuitive knowledge refers to a global, direct estimation, in contrast to a solution based on explicit computation" (p. 29). For the purposes of this study, the summary of Fischbein's (1987) comprehensive work on intuition presented by Fischbein and Schnarch (1997) highlights the essential features relevant to the hospital problem. "Intuition [is] a cognition that appears subjectively as self-evident, directly acceptable, holistic, coercive, and extrapolative... An intuitive cognition is distinguished from an analytically and logically based cognition by the feeling of obviousness, of intrinsic certainty" (p. 96). It is also of interest to see which mathematical skills are chosen by those with more formal training and whether this is backed up with any reference to intuition. Some people may have experienced sampling in their personal or professional lives that helps them intuit the likelihood of greater variation in the smaller sample and feel no need for further justification except for a statement. Others may have the ability to calculate a normal approximation to the binomial distribution without any appreciation for the meaning of the result.

The Current Study

The opportunity to explore further the use of formal mathematics as a solution strategy for the hospital problem arose in the context of a case study used in a preservice program for secondary mathematics teachers. The case study, *Chances Are*, was one of a series published by Harvard University to be used with high school mathematics teachers (Merseth & Karp, 1997). The pre-case worksheet for the case study presented the problem as given at the beginning of this paper and asked participants to provide a solution and list any assumptions made. The case study was based on a grade 11 class with a female teacher using this problem and group work to reinforce concepts related to probability and sample size that the class had been studying for the past few days. The main focus of the case study was on one group of five students who had difficulty with the problem and demonstrated many of the misconceptions found in earlier research. Social and pedagogical issues were also considered as part of the case study. The case study was used with three groups of preservice secondary mathematics teachers, students in the Bachelor of Teaching (BTeach) course at the University of Tasmania.

The objective in this small study is to add to the information available on strategies adults, in particular people who are planning to be high school mathematics teachers, will choose to answer the hospital problem. The aims are (a) to observe whether mathematics, intuition, or a combination of the two is most successful in achieving an acceptable solution, (b) to document the types of mathematics employed to support intuition or as the basis for a mathematical justification, (c) to suggest propensities for solution strategies of students with more or less formal mathematics training or by gender, (d) to report students' reactions to the problem, and (e) to suggest action to improve both intuitions and confirmatory skills in this area.

Subjects

The 33 preservice secondary mathematics teachers in this study were aged from 21 years, having just finished a university degree, to middle age, having worked in other careers before deciding to become mathematics teachers. The older students came from careers in geology, biology, engineering, meteorology, statistics, and taxi driving. The group varied widely in its previous exposure to formal mathematics courses. One was on leave from enrolment in a PhD program in mathematics, whereas others had done no formal mathematics at university but had convinced the head of school (not the author) that they had a mathematics background adequate for the course. Twenty-three BTeach students were classified as having at least a minor in mathematics (one-third of a second year university course) and 10 were classified as "other", that is they had fewer formal mathematics courses than required for a minor as part of their initial degree. There were 14 females and 19 males in the sample.

Procedure

The BTeach students were given the pre-case worksheet to complete overnight before the case study was handed out. They were told to answer the question by themselves and that although the solutions would be collected and read by the lecturer, no marks would be awarded, except for completing the exercise. The students knew that they would be participating in a case study based on the problem and appeared to be motivated to think seriously about it. Only one solution was very short and appeared to be by a student who had not taken the problem seriously or had little idea of how to solve it. All responses presented in the Results section are taken from the written responses of the BTeach students on the pre-case worksheet. The discussion of these solutions took place before the detailed discussion of the Harvard case study. Other student contributions, in particular the presentation of results from simulations, took place throughout the several sessions when the case study was the focus of discussion.

Analysis

The method of analysis employed was a clustering technique (Miles & Huberman, 1994), which looked at similarities in solutions to the problem. Of particular interest were those judged to be based on intuition in relation to sample size with no support in the way of mathematical calculations, those that displayed mathematical calculations but displayed no indication of intuition influencing the answer, and those that appeared to mix intuition with mathematical calculations. Further, when mathematics was employed in the solution, it was of interest whether it might be termed basic mathematics, such as percents and fractions, or theoretical mathematics based on the binomial and/or normal distributions. Finally it was of interest whether the conclusion reached was correct or incorrect. The analysis was based on "observed" outcomes, that is, only what was written on paper. If a student held an intuition that was not stated, obviously no credit was given. After the clustering analysis had taken place, note was made of gender and whether students were considered to have a strong background in mathematics.

Results

The results are summarised in Table 1 by whether the solution to the hospital problem was based on intuition as presented by the student, whether the presentation was based solely on a mathematical argument, or whether a mixture of approaches was stated. The results are also split by the correctness of the conclusion stated, except in one case. In that case a student stated a proposition of equal outcomes, attempted to calculate probabilities, made an error that indicated the small hospital had a very slightly higher probability, and chose that hospital. Given the numerical result the student should have retained the decision of "roughly equal" and was allocated as if that were the conclusion. Examples that illustrate the categories will be given according to the method of approach.

Table 1
Clusters of Responses to the Hospital Problem

Correctness of Conclusion	Strategy		
	Intuition	Mathematics	Intuition and Mathematics
		Formal maths	Formal maths
Correct	2 males	1 male	2 males
(n = 18)	5 females	1 female	1 female
		Basic maths	
		6 males	
		Formal maths	Basic maths
Incorrect	4 males	3 males	1 male
(n = 15)	4 females	1 female	2 females
	Total = 15	Total = 12	Total = 6

Intuition Only

Fifteen BTeach students gave responses deemed to be based on intuition only. Although some of these solutions contained numbers, these were not used with calculations to support the conclusion. Of these 15 solutions, 7 chose the smaller hospital based on the correct reasoning of sample size. These responses appeared to reflect previous experience of some sort involving variation in samples. Typical responses are the following, which, although including numbers, had no calculations.

> I can't work this out mathematically, but intuitively I would say that the smaller hospital is more likely to have more variation from the average of 50% boys. This is because even though the probability of a head (or tail) when tossing a coin is 1/2, this is only the case the more times we toss the coin (ie the head won't necessarily turn up once in every 2 tosses, but the more times the coin is tossed, the closer the heads get to a frequency of a 1/2). Similarly, the more babies there are, the more likely the frequency of boys is to converge to 50% of births.

> Hospital 1 has on average 3/4 of births for the town. Hospital 2 has on average 1/4 of births for the town. Hospital 1 has the greater % of babies being born – it is more representative of the whole town, so if 50% of the births were boys for the whole year, the likelihood of 60% or more of the births in one day being boys is less than that of Hospital 2. Because when dealing with larger numbers, the chance of a day's births being 60% or more boys is less than when dealing with smaller numbers, such as Hospital 2.

One person in this group mentioned that data would approximate a normal distribution but expanded only to say, "the degree of fit to the normal curve can be expected to be better for a larger sample size". The statement appears to reflect previous experience but is not backed up with any further mathematical argument.

Of the eight solutions that did not choose the small hospital, two did not reach a discernible conclusion. One of these stated two possibilities:

> Different sample sizes, 45 to 15 babies \ larger sample is more representative of the community as a whole. Considering this I would suggest that the larger hospital is more likely to have a birth rate closer to the norm. However, over a period of a year I would suggest that the 2 hospitals would have roughly the same number of such days.

The other only eliminated the possibility that the larger hospital would have more days with 60% boys born, but did not choose between the other two alternatives. One student stated the following as a final comment: "If more births occur at the larger hospital, there is a greater likelihood that more boys will be delivered at the larger hospital." Although this statement is technically correct and would not preclude a correct comment on the hospitals, no other comment was made. The other five students indicated that the hospitals would have roughly the same number of days with 60% or more boys born. Typical of these is the following.

> Percentages are proportions of original sizes. 60% is the same for both hospitals – only depends on original number to work it out. There is no reason why the larger hospital should have 60% boys more often than the smaller hospital. It's not asking which hospital turns out more boys because then the bigger hospital would do that.

Of the 15 BTeach students who used intuition only as a basis for their solutions, 8 were among the group with less theoretical mathematics in their undergraduate degrees. Of these 8, half reached the correct conclusion. Of the 15, 6 were male and 9 were female with the females being more successful (56%) than the males (33%).

Mathematics Only

Twelve students provided solutions based on mathematical arguments with no indication of previous experience or other intuition that would lead to belief in that solution at the start of the exercise. Eight of these twelve students reached the correct conclusion. Two of these students recalled from previous experience that the problem involved the binomial distribution and went to their old texts to find out how to carry out the calculations. One used the normal approximation to the binomial for both hospitals, whereas the other used binomial tables for the small hospital and the normal approximation for the large one as shown in the following argument.

> When I first read this problem I saw 2 normal distribution curves representing the two hospitals and the number of births each day, then as I read on this was replaced by another curve for the total population and the proportion/percentage of boys born each day (the hospitals then forming samples of the total population of the towns' births). When I came to the question I immediately thought – they must have both recorded roughly the same number of days where the number of boys born was 60% or more because they (the hospitals) sample the same population (the town).
>
> Assumptions: Each hospital is a random sample of the population, i.e., 50% of babies born in each hospital are boys.
>
> The sample size of smaller hospital is large enough to accurately represent the total population and hence no significant difference between the hospitals.

Reviewing the problem I decided the sample size of the small hospital was too small to have a normal distribution so after a little research (text books) and remembering, I realised that the sample would have a Binomial Distribution.

n: trials = 15
p: success probability = 0.5
where q = 1 – p = 0.5
and x = n
n: trials = 15
p: success probability = 0.5
where q = 1 – p = 0.5
and x = n X 60% = 9.

Also assuming the trials are independent, i.e., the chance of a boy being born is not dependent on the previous births.

$P(X \geq x) = P(X \geq 9) = 0.30$ using the Cumulative Binomial Distribution tables.

The larger hospital with its larger sample size $n > 20$ and p not too close to 0 or 1 can be approximated by the Normal Distribution, where:

$p = 0.5$	$P(X \geq x) = P(X \geq 27) = P(Z \geq (x - \mu)/s)$
$q = 0.5$	$= P(Z \geq (27 - 22.5)/3.35)$
$n = 45$	$= P(Z \geq 1.3416)$
$\mu = np = 22.5$	$= 0.5 - P(0 > Z > 1.3)$
$s = \sqrt{npq} = 3.35$	$= 0.09$, using areas under Normal Probability curve.
$x = n \text{ X } 60\% = 27$	

Hence the Probability that the smaller hospital will have more days in which 60% or more of those births will be boys, is greater than that of the larger hospital.

P(small hospital) = 0.30 ≥ P(large hospital) = 0.09.

This student forgot the continuity correction for the normal approximation to the binomial, which was remembered by the other student who found $P(x \geq 26.5) = P(Z \geq 1.19) = 0.117$. Neither person commented on their surprise or otherwise with the answer, although as seen above the person providing that response indicated an initial belief that the hospitals would have roughly the same number of days with 60% or more boys born.

One other student attempted to calculate binomial probabilities but only calculated the exact probabilities for 9 boys out of 15 and 27 boys out of 45. Although incorrect to answer the question based on "60% or more boys", these values led to the correct conclusion that the smaller hospital was more likely to have exactly "60% boys" born.

The other four students with the correct conclusion used simpler mathematical arguments based on how likely it would be to achieve deviations from the expected number of births to reach the 60% value. After stating that for the total of 60 births per day, 30 on average would be boys, one student argued as follows.

> Given that, for the larger hospital to record 60% or greater boys in a given day would require at least (27/30) 90% of the average number of boys born per day, as compared to only 30% (9/30) or more for the smaller hospital, one would suggest that it is more likely for the smaller hospital to record a greater number of such days.

After displaying a few percent calculations, another student argued as follows.

> Given that 53.3% of the boys are already present (8 for the small hospital and 24 for the large hospital). It only requires 1 more boy for the small hospital to give 60% but requires 3 boys to give 60% for the large hospital. The chances of 1 more boy is [sic] higher than the chances of 3 more boys.

Variations on this theme were expressed with probabilities, the percents represented by each birth, and the days required by the large hospital to redress an imbalance.

Four of the students who used formal mathematics in their solutions made errors that led them to the wrong conclusion. One made an error in the formula when applying the normal approximation to the binomial and one applied a hypothesis test incorrectly to the two hospitals. Both concluded that the hospitals would be equally likely to have 60% or more boys born. A third student did not think there was enough information available to use the binomial distribution, did some calculations to estimate total births with "unknown" probabilities for "number of days greater than 60%", which led to the conclusion that the large hospital would have more "60% or greater" days. The final solution in this group was one mentioned earlier where an incorrect calculation of probabilities (0.21 and 0.23) should have led to the conclusion that the hospitals would have roughly the same number of days with 60% or more boys born.

Only one of the students in this group of 12 was considered to have a weak formal mathematics background. This was the student who used the normal approximation correctly for both hospitals. His university major was in zoology and although he had done no formal mathematics courses at university, he had worked for 10 years in positions involving scientific research. Of this group of 12 using only mathematics for the problem, two were female, one providing a correct solution. Of the 10 males, 7 were correct.

Intuition and Mathematics

Only six students provided solutions that appeared to attempt to justify intuition, including previous experience, with a mathematical argument, of which half were correct. The three BTeach students providing correct conclusions used very different approaches. One provided a graph of a normal distribution curve centred at 50, with the horizontal axis labelled "percentages of boys" and the vertical axis labelled "Num. of days". Under the graph was a paragraph, edited extracts of which are as follows.

> If all things are considered equal both should produce the above distribution. However with the smaller hospital I would imagine there is more chance of (looking at extremes) all babies being boys than in the other hospital with an average of 45. Hence I would envisage that the curve for the smaller hospital would be a bit flatter resulting in ... a lesser value for the 60% mark ... so although it would be flatter it would be longer hence we may find that the larger hospital with its larger sample

> will result in more days closer to 50% resulting in a more pointy distribution while the other is more flat - at this point I would imagine that the smaller hospital may lean towards a greater number.

The second student reported an a priori belief in equality of the hospitals, but then argued that as n approached infinity the ratio of boys to all births would approach 1/2, hence "the smaller hospital will be more likely to have 60% or more boys a day over a year". To confirm this belief the student went to a senior secondary mathematics book, found the formula for the binomial distribution and correctly calculated the probabilities for both hospitals using a calculator. The third student expressed the correct intuition based on sample size and attempted to follow the approach of the previous student. Without a reference for the appropriate formula, the student guessed and achieved the correct relationship of the two probabilities but realised that the actual values were incorrect.

The other three students who mixed intuition and mathematics used basic mathematics. All concluded that the hospitals would have about the same number of days with 60% or more boys born. One argued sequentially and erroneously, that no matter how many children are born, the probability of all of them being boys is 50%. The claim was made individually for up to four babies and then extrapolated to 15, 45, and 60. Another worked with fractions, concluding the number of ways to get 60% or more boys was 6/15 = 2/5 in the smaller hospital and 18/45 = 2/5 in the larger hospital, and hence "the two hospitals record roughly the same number of days". His final comments, however, indicate that he may have become confused by the additional information about the 365 days on which data were recorded. He stated the following.

> Initially I believed that the smaller hospital has a greater likelihood of having ≥ 60% of births being boys because of the greater variation due to the small sample size.
>
> However because the births are independent and at each one the P(B) = 1/2, I believe that over the period of 365 days (which is a fair amount [sic] of trials) the variation will average out so that the two hospitals will record roughly equal number of days where the percentage of boys born was ≥ 60%.

Finally, after translating percents into numbers for the two hospitals and doing comparisons, a student argued, "For the small hospital the fluctuation between 50-60% is only 1-2 children, and 4-5 for the large hospital but they [large] have more children so I think it would be about the same for both."

Only one of the six students in this group was considered to have a weak formal mathematics background. This was the one who made the erroneous claims about all boys occurring 50% of the time independent of sample size. In this group half were male.

Discussion of Results

The first three aims of the study are considered in this section. Several observations can be made about the strategies and success of preservice mathematics teachers in solving the hospital problem. Their success rate (55%) is not surprising compared to the tertiary students surveyed by Kahneman and Tversky (1972), or the prospective teachers surveyed by Fischbein and Schnarch (1997), considering their mathematical backgrounds. The rate did not differ greatly for females (50%) and males (58%). What may be considered surprising is the apparent lack of attempt to engage both intuition and mathematical skills. Only six students (18%) appeared to be working with both.

The students who used intuition only may not have encountered the binomial or normal distributions before but all appeared to understand subsequent discussions of these when presented by other members of the class. They also may not have had immediate access to sources that could supply needed information or felt that the problem was not intended to be that complicated. That those who used intuition only were only successful about half of the time points to the need to make students aware of two pitfalls: one is not looking for mathematics to back up intuitions and the other is knowing which mathematics is appropriate to use. In identifying the general principles that influence intuitions in inappropriate ways during probabilistic reasoning, Fischbein and Schnarch (1997) identified ratio as a schema that caused difficulty for the subjects in their study.

> The general principle identifiable in the problems relating to the effect of sample size is the equivalence of ratios. For example, the concept of ratio is involved in the students' incorrect solution of the problem of the two hospitals. Students are apparently misled by their belief that one must use ratios to solve this problem. Instead, one has to consider another stochastic law, the law of large numbers. As the sample size (or the number of trials) increases, the relative frequencies tend toward the theoretical probabilities. For our subjects, not trained in stochastics, the principle of equivalence of ratios imposes itself as relevant to the problem and thus dictates the answer. It is the evolution of this principle that shapes the evolution of the related misconception and causes it to become stronger as the student ages (p. 103).

As seen in the responses from the preservice teachers in the current study, the influence of ratio, expressed in terms of percent, was significant in about 20% of responses.

The students who presented mathematical arguments with no hint of using intuition or previous experience were more successful (67%) than those who used intuition only (47%), but those who made errors that led to the wrong conclusion illustrate the difficulty when there is no intuition or previous experience to test against one's mathematical calculations. This was particularly true of the two students who had full mathematics majors (one a double major) but made errors with the normal distribution and hypothesis testing. Both were very confident of their initial conclusions.

Of the six who employed both intuition and mathematics, half reached the correct conclusion. These used formal mathematics related to the model appropriate for the hospital problem. The others based arguments on basic mathematics that was misinterpreted in the complex context. The response that indicated the chances of all-boy samples were 50%, was similar to responses about four coin tosses being tails observed by Moritz and Watson (2000), for between 28% and 49% of students in grades 6 to 11. That a preservice teacher has the same difficulty as many students, in an area considered by the author at least as a fundamental part of probability, points to this as an area for research based on instructional intervention.

Of those who used mathematics in their solutions, either with or without intuition, half used basic mathematics and half at least attempted a solution based on the normal and/or binomial distributions. That overall a third of these were incorrect is somewhat worrying for preservice mathematics teachers. Perhaps most disturbing was the treatment of percents as absolute numbers and the misunderstanding of basic probability, but the lack of checking of formal mathematics against other criteria was also a matter of concern.

Of the 10 students considered to have weak formal mathematics backgrounds, eight used intuition, including previous experience, to solve the problem, one used formal mathematics only, and one appeared to mix intuition and basic mathematics (unsuccessfully). Overall half of these students were successful, a result not very different from those considered to have a strong mathematics background (57%). It is hence not possible to make any claims about differing success rates being dependent on the extent of one's previous mathematical background. The only gender difference that might arouse further interest was the observation that 10 of the 12 students (83%) who used mathematics only in their solutions were male. Combined with the 60% of females in the group that used intuition only, one might suggest a hypothesis of gender preference for the use of intuition or formal mathematics. This would be an interesting area for further research that could impact on classroom expectations of future teachers.

Student Reactions

Student reactions to the correct response during class discussion were mixed with some who had used intuition at first seemingly overwhelmed by the theoretical solution. Several mathematics majors lamented their lack of practical experience in statistics that would have assisted in making judgments about sample size. This was summed up by the PhD student who said the following in his evaluation of the case study.

> ...there is no substitute for exposure to practical problems when it comes to understanding statistics. Even though I have completed a few courses in statistics at university level I still found that my intuition let me down when it came to working on this problem.

He further claimed in class discussion that he had never carried out any sampling during his study of statistics. In a sense this problem turned out to be a "leveller" for the groups who worked on it and completed the case study. Since the success rate was not related to previous formal mathematics experience, it was a humbling experience for some and a confidence builder for others.

Several students made comments suggesting they had difficulty because the problem stated the extreme result in terms of boys' births only. One claimed that the problem should be reworded to include girls but then went on to solve it correctly using intuition only. Not appreciating the symmetry of the problem in terms of boys and girls was one of the features included in the Harvard case study (Merseth & Karp, 1997) as a difficulty for high school students. This point was well taken by the preservice teachers and their own difficulties enhanced the relevance of the case study. The fact that a few of the BTeach students experienced the same difficulties with the relationship of percent and actual numbers as the students in the case study and those interviewed by Watson and Moritz (2000), reinforced the need to concentrate on percent in the early years of secondary school.

During the class discussion one of the students who had reached the wrong conclusion would not consider other students' correct intuitions until his original mathematics was proved wrong. Having stated "Bi(n,q) ≈ N(nq, nq(1 − q))", he forgot the square root for the standard deviation. He thus concluded that z = 0.2 for all n and hence both hospitals have roughly the same number of days with 60% or more boys born. Once convinced of his error, however, he quickly produced the following elegant argument for the class.

> As $Bi(n,\Theta) \approx N(n\Theta, \sqrt{(n\Theta(1-\Theta))})$, for a value of $\Theta = 0.5$ and an x-value of .6 or more boys, $z = (0.6n - 0.5n)/\sqrt{(n(0.5)(0.5))} = .2\sqrt{n}$. Hence as n increases the z-value increases. Since we are interested in $P(z \geq 0.2\sqrt{n})$, when n = 15, the probability is greater than when n = 45.

This value reflects the area under the normal curve to the right of 0.2÷n, which decreases with increasing n. With a sketch all students appeared convinced.

Overall in their evaluations, the students were positive about the case study. The next section includes a summary of the simulations used by the BTeach groups as a basis for discussion of how to enhance an intuitive appreciation of the impact of increasing sample size on experimental outcomes.

Suggested Action

A simulation suggested in the Harvard case study (Merseth & Karp, 1997) is to use 15 coins and 45 coins to simulate the births in the two hospitals. The preservice teachers immediately saw the value of this analogy but also realised that it might be difficult to implement in a classroom. It was interesting to observe that without very much previous experience with simulations themselves, it took the groups some time to agree on a workable classroom procedure for using 60 coins. Less noisy ways to provide simulations involved using computer software or a graphics calculator. Windsor (1998) provides the instructions necessary to carry out the simulation on a TI-83 graphics calculator and discusses using it with senior secondary students. BTeach students were given the opportunity to trial this method, found it relatively easy to follow instructions, and felt it was a good application of the technology, if it were available.

The probability simulation software for Macintosh, ProbSim (Konold & Miller, 1993), provides the opportunity to see samples being created, as shown in Figure 1. This software provides an intermediate step between coin simulations and a "magic box" computer program that produces only a final probability or number of days per year when 60% or more births are boys. As seen in Figure 1, the sample size is set at 15 for a random generator (the Mixer) containing a G and B to determine each birth. Using 365 repetitions produces a data record, which can be analysed to count how many days have 9 or more (60%) boys. This result is shown in the Analysis window in Figure 1. A similar outcome for 45 days was also created with the software. Repeated outcomes for different sample sizes were plotted by the BTeach students to document the relative occurrence of extreme results for different sized hospitals to observe the trend with increasing sample size.

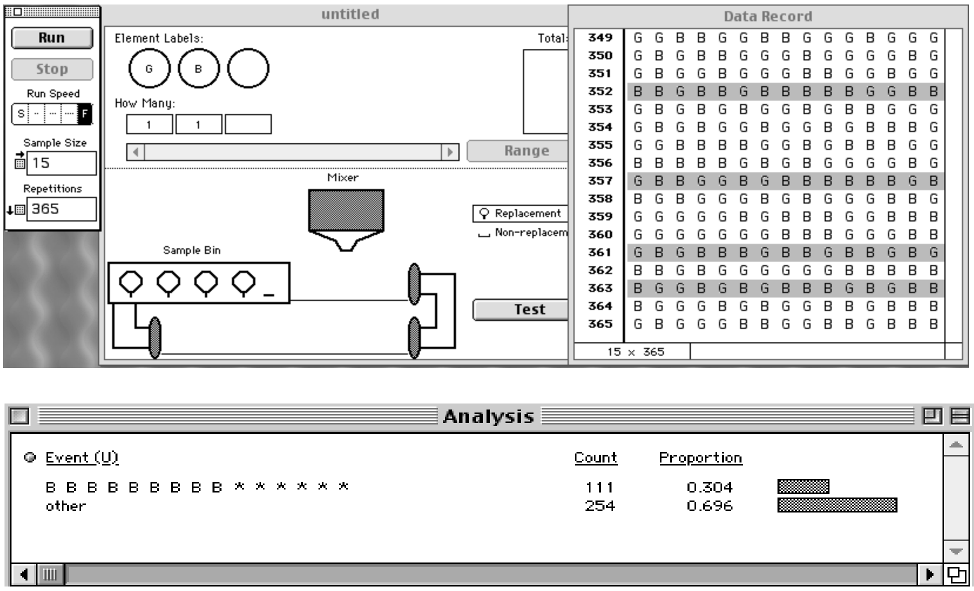

Figure 1. Simulation of 15 births on 365 days, counting the days with 9 or more boys (as an unordered event).

One of the students who had used theoretical mathematics incorrectly produced a spreadsheet simulation in Excel that for one year found extreme results occurring for a 15-baby hospital on 118 days, a 30-baby hospital on 68 days, and a 45-baby hospital on 36 days. For high school students learning to use spreadsheets as part of mathematics or information technology courses, this would be an excellent project. The PhD student noted above wrote a *Mathematica* script to simulate a year's daily births at each hospital. Running this program to simulate 1000 years produced average values of 42.7 days for the large hospital and 110.6 days for the small hospital for the extreme results. These procedures, although impressing the students, also made them appreciate the need to be able to understand and construct such algorithms, not just produce them as finished products. In the end, based on their own experiences and the case study, all students agreed to the importance of using some kind of simulation in conjunction with the theoretical mathematics for the hospital problem. In terms of using the problem at school level (except for the highest level senior course in Tasmania), it was considered imperative to use sampling methods as high level formal mathematics was not available as a tool.

Although, as noted, a few senior secondary students would be able to solve the hospital problem using the normal approximation to the binomial, most high school students would require skills working with fractions, percents, and basic independent probability calculations to confirm the intuitive solution suggested by simulations. The problem hence provides an excellent opportunity to reinforce these skills in an applied setting. The BTeach students supported this suggestion based on the solutions they saw their classmates present and the difficulties some of them experienced.

Conclusion

For people preparing to be mathematics teachers it is disappointing that so few naturally mixed intuition with an attempted mathematical justification in solving the hospital problem. The case study setting of a school class provided an excellent opportunity to discuss this point, as well as the mathematics teachers' role in modelling the relationship between intuition and justification. It is essential that preservice mathematics education programs provide students with problem situations where erroneous conclusions may result, either due to natural intuition without previous mathematical experience or due to faulty calculations without intuition. This is true not only for chance and data but also for other parts of the mathematics curriculum. Not always will the same problem provide both opportunities. The hospital problem is particularly effective in this regard because of the extremes possible in terms of the complexity of solutions that can be created. Problems like this should reinforce the importance of both types of problem solving tools: intuitive and mathematical. It is only when one has both that one can be sure that the solution makes sense.

Acknowledgement

An earlier version of this paper was presented at the Mathematics Education Research Group of Australasia Conference in Fremantle, WA in July, 2000.

References

Fischbein, E. (1987). Intuition in science and mathematics. Dordrecht, The Netherlands: Reidel.

Fischbein, E., & Grossman, A. (1997). Schemata and intuitions in combinatorial reasoning. Educational Studies in Mathematics, 34, 27-47.

Fischbein, E., & Schnarch, D. (1997). The evolution with age of probabilistic, intuitively based misconceptions. Journal for Research in Mathematics Education, 28, 96-105.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgement of representativeness. Cognitive Psychology, 3, 430-454.

Konold, C., & Miller, C. (1993). ProbSim. [Computer software] Amherst, MA: University of Massachusetts.

Merseth, K. K., & Karp, J. B. (1997). Harvard mathematics Case Development Project 1994-1997: Cases of secondary mathematics classrooms. Cambridge, MA: Harvard Project on Schooling and Children.

Miles, M. B., & Huberman, A. M. (1994). Qualitative data analysis: An expanded sourcebook. (2nd ed.). Thousand Oaks, CA: Sage Publications.

Moritz, J., & Watson, J. (2000). Reasoning and expressing probability in students' judgements of coin tossing. In J. Bana & A. Chapman (Eds.), Mathematics education beyond 2000. (Proceedings of the 23rd annual conference of the Mathematics Education Research Group of Australasia, pp. 440-447). Perth: MERGA.

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. Psychological Bulletin, 76(2), 105-110.

Watson, J. M., & Moritz, J. B. (2000). Developing concepts of sampling. Journal for Research in Mathematics Education, 31, 44-70.

Watson, J. M., Collis, K. F., & Moritz, J. B. (1995, November). The development of concepts associated with sampling in grades 3, 5, 7 and 9. Paper presented at the Annual

Conference of the Australian Association for Research in Education, Hobart [on line]. Available: ftp://ftp.swin.edu.au/pub/aare/aare95/conf95/watsj95.475.
Well, A. D., Pollatsek, A., & Boyce, S. J. (1990). Understanding the effects of sample size on the variability of the mean. Organizational Behavior and Human Decision Processes, 47, 289-312.
Windsor, N. J. (1998). Simulating a sampling problem using a TI-83 graphics calculator. Australian Senior Mathematics Journal, 12(2), 32-37.

Author

Jane M Watson, University of Tasmania, Faculty of Education, GPO Box 252-66, Hobart TAS 7001. Email: <Jane.Watson@utas.edu.au>