Journal of Mathematical Behavior 32 (2013) 660–672

Contents lists available at ScienceDirect

The Journal of Mathematical Behavior

journal homepage: www.elsevier.com/locate/jmathb

Likelihood and sample size: The understandings of students and their teachers

Jane Watson[a,*], Rosemary Callingham[b]

[a] *Faculty of Education, University of Tasmania, Private Bag 66, Hobart, Tasmania, 7001, Australia*
[b] *Faculty of Education, University of Tasmania, Locked Bag 1307, Launceston, Tasmania, 7250, Australia*

ARTICLE INFO

Article history:
Available online 11 September 2013

Keywords:
Cognition
Hospital problem
Percentage
Probability
Sample size
School students
School teachers

ABSTRACT

The aim of this study was to consider the match of student statistical understanding and teacher pedagogical content knowledge in relation to sample size and likelihood. Students were given two contexts within which to compare the likelihood of events for different sample sizes. Teachers were presented with one of the contexts and asked what their students would do and how they would remediate incorrect responses. The data also provided the opportunity for a detailed hierarchical analysis of students' and teachers' understandings. Analysis of student solutions revealed a wide range of reasoning, some of which was apparently unfamiliar to teachers.

1. Introduction

1.1. Curriculum background

Likelihood, probability, samples and sampling are included in most school mathematics curriculum documents in the 21st century (e.g., Australian Curriculum, Assessment, and Reporting Authority [ACARA], 2013; Franklin et al., 2007; National Council of Teachers of Mathematics [NCTM], 2000). In Australia, for example, conducting large and small numbers of trials appears in Year 6, determining probabilities is in Year 7, and using samples/sampling to collect representative data is in Year 8 (ACARA, 2013). Similarly the NCTM (2000) includes these topics in Years 6–8, with more emphasis on sampling in Years 9–12, as do the *Guidelines for Assessment and Instruction in Statistics Education* (Franklin et al., 2007) at Levels B and C, and the *Common Core State Standards for Mathematics* (Common Core State Standards Initiative, 2010) from Year 6.

The relationship between sampling and probability may not be as explicit in curriculum documents as desired but on one hand it is expected, for example in Australia, that from Year 6 students can "describe probabilities using fractions, decimals, and percentages" (ACARA, 2013, p. 43). On the other hand, the subset relationship of a sample to its population is part of the definition of sample, as is its representative nature. Franklin et al. (2007) address in more detail than other curriculum documents the understanding that the larger the sample the more closely it resembles the population (avoiding purposeful bias). The expectation about using fractions, decimals, and percentages is also relevant when considering characteristics of samples, as parts of a sample may be distinguished and described using one of these forms.

* Corresponding author. Tel.: +61 3 6226 2570; fax: +61 3 6226 2569.
E-mail addresses: Jane.Watson@utas.edu.au (J. Watson), Rosemary.Callingham@utas.edu.au (R. Callingham).

0732-3123/$ – see front matter
http://dx.doi.org/10.1016/j.jmathb.2013.08.003

1.2. The Hospital problem

Although there are many contexts within which understanding of samples and likelihood can be explored, historically the most famous is Kahneman and Tversky's (1972) Hospital problem. They used it to illustrate what they called the *representativeness* heuristic, the intuition based on past experience that any sample is representative of the population. The heuristic focuses on the similarity of the sample statistic to the population parameter rather than on the sample size, which is not seen as explicitly related to the population.

> A certain town is served by two hospitals. In the larger hospital about 45 babies are born each day, and in the smaller hospital about 15 babies are born each day. As you know, about 50% of all babies are boys. The exact percentage of baby boys, however, varies from day to day. Sometimes it may be higher than 50%, sometimes lower.
> For a period of 1 year, each hospital recorded the days on which (more/less) than 60% of the babies born were boys. Which hospital do you think recorded more such days? The larger hospital? The smaller hospital? About the same (i.e., within 5% of each other)? (p. 443)

The problem is set up in such a way that if a person "knows" that larger sample sizes better represent the population, in this case with 50% boys, then it is clear that a sample of 45 babies should have closer to 50% boys than a sample of 15 babies. This knowledge must then be placed in the context of realizing that "60% boys" is a more extreme value than "50% boys" and hence less representative of the overall population of births. Further, samples with more than 60% boys are even more extreme and further away from the expected population percentage.

1.3. Tasks for college students

The Hospital problem and variations have been most commonly used with college or university students. Using surveys with undergraduate psychology students with no background in statistics or probability, Kahneman and Tversky found only 20% of students could make the appropriate choice of the small hospital and over half said the probabilities would be the same. Shaughnessy (1977) replicated these outcomes with other college students, again with over half choosing equality for the two hospitals and the rest evenly split between choosing the two hospitals. Reaburn (2008) used the Hospital problem as part of a pre- and post-intervention survey for a unit in statistics for first year university students. Of the 29 students completing the initial survey, 17 reported having studied statistics in a previous mathematics course; of these 5 were correct and 9 chose equally likely. Of the 8 students who had not studied statistics earlier, 7 gave the correct response. At the end of the unit, 9 students completed the post survey with 4 of 5 who answered equally likely earlier, changing to the small hospital; the other 4 had been correct on the initial survey. Although a small number, Reaburn was encouraged that reasoning about sample size had improved.

In working with mathematically qualified university students, Watson (2000) used the original Hospital problem with 33 pre-service mathematics teachers, most of whom had degrees including at least second year mathematics units. The problem was used in conjunction with a Harvard university case study entitled "Chances Are" (Merseth & Karp, 1997) and students were given the problem to work on overnight and bring their solutions to class the next day. These pre-service teachers' strategies were split into three categories: intuitive, pure mathematics and a mixture of the two. Eighteen students (55%) were initially correct in their solutions with seven using an intuitive approach, eight using mathematics only, and three mixing the two approaches. Of the 15 (45%) incorrect responses, eight were based on intuition, four on mathematics, and three on a mixed strategy. After the classroom discussion there was universal acceptance of the solution and much interchange about the importance of appreciating many different methods of solving the problem.

1.4. Interviews of school students

Interview settings provide the opportunity to probe the reasoning of subjects in more detail than is usually possible in a survey, even when explanations are requested. In a small Australian study Watson, Collis, and Moritz (1995) used a variation on the problem based in the context of taking a random sample of 50 children from a city school and a similar sample of 20 children from a country school (each school with half boys and half girls). The interviewed students were told that one sample was unusual with 80% boys and the students were then asked which of these schools it was likely to have been or whether the schools were equally likely to have been the sampled school. Of the 12 students interviewed, 8 said "same" usually based on "randomness" or percentages not being definite numbers. Of the two choices of the small sample the reasoning was idiosyncratic or based on contextual knowledge: "more boys live in the country"; similarly for the two who chose the large sample: "more people live in the city." In a larger interview study of 41 students (three year 3 and 19 in each of years 6 and 9), Watson and Moritz's (2000) results were similar with 25 saying same chance and only 6 giving an adequate reason to choose the small school: "because there are a lot less children [in the country school], so if you had perhaps a few more [boys] it would bring the percentage up a lot quicker than with [the city school]."

In an intervention study with 42 students (eight in year 6 and 34 in year 9), *ProbSim* software (Konold & Miller, 1993) was used to simulate the collection of samples of size 30 and size 10 from the two schools and create cognitive conflict for students (Watson, 2007). Twenty random samples were collected for each sample size and students kept records of the number of samples where 70% or more were boys. Initially 30 students believed the outcomes would be equally likely, four