Original Article

What do the stats tell us? Engaging elementary children in probabilistic reasoning based on data analysis

Mairéad Hourigan and Aisling Leavy

Department of Language, Literacy and Mathematics Education, Mary Immaculate College, University of Limerick, Limerick, Ireland
e-mail: Mairead.Hourigan@mic.ul.ie

Summary As part of Japanese Lesson study research focusing on 'comparing and describing likelihoods', fifth grade elementary students used real-world data in decision-making. Sporting statistics facilitated opportunities for informal inference, where data were used to make and justify predictions.

Keywords: Teaching; elementary statistics education; probability; informal inferential reasoning; real-world context.

INTRODUCTION

Probability is all around us – weather forecasts, lotteries, sports statistics, gambling, insurance and population trends. In order to make informed decisions based on these data, children need to have informal understandings of probability, i.e. the study of how likely an event is (Van De Walle 2013). The understandings of probability that children develop in the elementary years lay the foundations to support them in becoming critical consumers of data and in making informal inferences.

The challenge of teaching statistics in the elementary years

Much research exists on the complexities involved for adults in understanding probability (Leavy et al. 2013; Tversky and Kahneman 1974). These difficulties are exacerbated for younger learners, where everyday intuitions pose obstacles to developing probabilistic understandings (Fischbein 1975; Metz 1998). The challenge for teachers is not merely to provide more experiences in probability for younger learners but also to provide meaningful experiences, which promote correct and robust understanding of concepts. Hence, when designing pedagogical experiences relating to probability, educators need to take into account the importance of prerequisite mathematical understandings, the presence of naïve understandings and biases relating to probability and the ways in which everyday meanings of probabilistic terms may pose challenges for the development of understanding (Greer and Mukhopadhyay 2005; Kaplan et al. 2010; Leavy and Hourigan 2015; Lecoutre 1992).

RESEARCH CONTEXT

Participants

This study describes research carried out in a College of Education[1] in Ireland. Twenty final year pre-service elementary teachers participated in the study during the concluding semester of their 3-year teacher education programme. Participants had completed their mathematics education courses (three semesters) and all teaching practice requirements (at junior, middle and senior grades) and self-selected into mathematics education as a cognate area of study.

Research design

All pre-service teachers, and two mathematics educators, engaged in *Japanese Lesson Study*, which is an approach for studying teaching that utilizes detailed analyses of classroom lessons (Fernandez and Yoshida 2004; Lewis and Tsuchida 1998). The research was conducted over a 12-week semester. Pre-service teachers worked

[1]A third-level institution that provides initial teacher education for pre-service elementary teachers.

collaboratively in four small groups to design and implement a 'study lesson'. The researchers' selection of the topics was informed by international curricula and the recommendation of professional bodies. Each group was given a particular probability concept that would be presented to students in the following sequence: describing likelihoods, comparing and explaining likelihoods and ordering likelihoods of events and the Law of Large Numbers (Leavy and Hourigan 2015). This article focuses on the second of these: comparing and explaining likelihoods.

The process of designing and implementing the respective study lessons involved three phases. While the first phase involved *research and preparation* of a study lesson, the second phase, the *implementation* stage, involved teaching the lesson in a fifth grade (10–11 year olds) classroom. Subsequently, following detailed discussion, the original lesson design was modified prior to the second implementation or reteaching the lesson with a second different fifth grade class. The second implementation was videotaped. The activities in this second lesson are described in this article. The final *reflection* phase involved each lesson study group making a presentation at the end of the semester regarding the process and outcomes of their work.

This process of Lesson Study, involving the planning and the implementing of lessons in classrooms, facilitated the design of sequences of instruction to support the development of statistical and probabilistic reasoning with elementary children. The study lesson we describe has been designed, taught and revised through the lesson study process with 10–11 year olds in two different elementary schools.

STUDY LESSON: COMPARING AND EXPLAINING LIKELIHOODS

Designing probability experiences for elementary school children: a focus on sports contexts

The goal of the classroom-based inquiry, and the ensuing sequence of instruction, was to help children reason about and describe particular unpredictable events using the vocabulary of probability. While the children had previously been introduced to the formal language of probability, they needed additional opportunities to use the language of probability in a meaningful way to describe and compare the likelihood of various outcomes. In designing these opportunities, we were mindful of children's everyday use of statistical and probabilistic language that may conflict with the mathematical meanings and hence wanted to reinforce the mathematical nuances of the probability language and terms.

As it is not possible to teach probability devoid of a context, we selected a sports context to motivate children to use the language of probability. In particular, we challenged children to make predictions about future results and rankings of the national teams competing in the Rugby 6 Nations Competition. For many situations, it is possible to determine the theoretical probability of a particular event, e.g. the probability of obtaining an even number on a die can be accurately represented mathematically using fractions (i.e. 3/6). However, there are many everyday events such as sporting events where it is not possible to accurately determine the outcome, which makes decision-making in the face of uncertainty even more complex. In these situations, statistics are tools that prove helpful in analysis of past trends and data and facilitate making inferences in situations of uncertainty (Jones et al. 2007). Data can be examined for patterns in order to draw conclusions and in turn make predictions about future outcomes, i.e. develop and evaluate inferences (Tarr et al. 2006). In the sporting context presented, in order to make informed predictions about future performance of the teams, children were challenged to investigate and make inferences arising from analysis of sports data. The data were presented in the form of league tables and other performance data arising from games in previous years/seasons. Given the sports context, the influence of home advantage and prediction bias was also examined. The Rugby contexts we used can be easily exchanged for sports more familiar to children of any nationality, such as American Football or Australian Rules, as these games have similar statistics compiled and readily accessed through the media and Internet.

SEQUENCE OF INSTRUCTION

Lesson warm up

Research reveals than students may hold multiple, and contradictory meanings, for chance language such as possible, impossible and certain (Fischbein et al. 1991; Watson and Moritz 2003). Remaining cognisant of this, as an introduction to the topic, the teacher reviewed the language of probability that students had been exposed to in the mathematics curriculum. Each

child was given a set of colour-coded probability language cards (Figure 1). It was decided to colour code them in order to facilitate the teacher to gauge the range and nature of response to various scenarios and in turn focus class discussion.

Words were selected from the range of probability language, and children were asked to explain the meaning of the selected word. Their response to this activity highlighted very few instances of the misuse of chance language. Moreover, the majority of children had a good facility with the chance language and demonstrated a readiness to use this language to respond to scenarios:

Teacher: The word 'impossible'. What does that mean?

Jim: It can't be done.
Teacher: Ok. Let's talk about another word: 'likely'.

Susie will you explain that word?

Susie: It means it probably will happen, there is a chance that it mightn't, but it probably will.

Teacher: And another word, the word: 'even'.

TJ can you explain what 'even' means?

TJ: It means a 50:50 chance. It might happen but it might not.

While in this activity, children demonstrated an understanding of the selected probability language, during subsequent activities, some misconceptions became apparent. For example, the word 'even' was sometimes used incorrectly in informal conversation. In one case, when predicting the outcome of a card activity, a child used the word 'even' to describe the chances of picking an identical card. When asked what does 'even' mean, the child replied *the same*. This suggests that the child had generalized the language 'even' considering the language 'even' to mean the same thing as 'equal to' (as opposed to the probability of an event happening is exactly the same as it not happening (50:50)).

Fig. 1. Probability language cards

We then developed and explored the sports context. We presented sports scenarios with a view to having children gain more practice using probabilistic language to describe the likelihood of these events occurring. The teacher made a number of statements and asked students to hold up a probability language card (Figure 1) that best represented the likelihood of the event occurring.

Teacher: What is the chance that Ronan O' Gara[2] is going to walk into the classroom?
Most childen hold up card that says 'unlikely' or 'impossible'

Teacher: Ria, you are holding up 'impossible'. Tell me why you think it is impossible that Ronan O'Gara is going to walk into the classroom now?

Ria: It just is. He is hardly just going to walk in.

Teacher: Tadgh, your card says 'unlikely'. Can you explain your thinking?

Tadgh: He might walk in, you never know. There is a chance.

The class was informed that while it was not possible to have Ronan visit the classroom, they would watch a video clip of him playing rugby instead. Children were instructed to watch the clip carefully because when it was paused they would have to make a prediction. A clip of Ronan O' Gara taking a penalty kick was played. Once he had kicked the ball, the video was paused. The class was asked to select a card to respond to the statement: *What is the probability of Ronan O'Gara scoring the penalty*? Children who were less familiar with rugby gave generic responses when justifying their answers. For example, Jessica said 'I picked "even chance" because he might score and he

[2]Ronan O'Gara was a well-known member of the Irish rugby team.

mightn't. You just never know'. Analysis of our data revealed that children who possessed more experiences and knowledge of the sport were more specific in their justifications:

Kate: It is likely he will score because he has a very good record.

Tom: I picked 'certain' as he is at a good distance and angle.

Tess: I think it is unlikely that he will score. He has been having a bad season.

Harry: It is 'Certain' he will score because the ball is heading towards the goal.

It is interesting that some of the children at this stage immediately focused on the players' past performance when considering the factors which affected his chances of success. The end of the clip was played; this facilitated children in comparing their predictions with the observed outcome, i.e. he did score the penalty kick. The teacher explored the scenario further, making all children aware that prediction is complex given that so many factors affect the kicker's success, e.g. *What if the player had a sore toe*, *would that increase or decrease his chances*?

Developing familiarity with the context

After children had made predictions and provided justification for their predictions, they were given the opportunity to make predictions about the ranking of the countries who partook in the *2013 Royal Bank of Scotland 6 Nations Rugby Championship*. At the time of the research, four of the five rounds for the 2013 season had been played. The Championship takes the form of a round-robin tournament where each team plays every other team once. Given that some of the children had little or no experience of the championship or of reading and interpreting the resulting league tables, it was necessary to familiarize the class with the structure of the Championship. The teacher provided an overview of the Championship and presented information pertaining to the league tables arising from the round-robin tournament. Children were supported in developing understanding of the context through carefully directed teacher questioning. The teacher asked questions such as *Who are the six countries who compete in this competition*? *How many matches do each nation play*? *How many matches have they played so far this year*? *How does the points system work*? *How many points do you get for a win*? (Answer: 2 points) *How many points do you get for a draw*? (Answer: 1 point) *How many points do you get for a loss*? (Answer: 0 points).

Using statistics to make predictions: using league table data

When all children had developed the appropriate prerequisite knowledge about the structure of the championship, they were presented with the results to date by means of a league table (Figure 2). The table below represents a simplification of such tables commonly used in the media. In particular, the researchers decided to omit data regarding the total number of game points scored by a team within all the matches played to date as well as the total number of game points scored against a team to date and the overall game points difference. Instead the league table focused only on match outcomes. The resulting table presents data relating to number of games played (Played), the number of games each team had won (Won), drew (Drew) or lost (Lost) in addition to table points achieved from these results (Pts). Initial questions posed by the teacher focused on the structure of the table. For example, she asked: *Who is at the top of the table*? *How many points do they have*? *Looking at the points only*, *how many matches have Ireland won*? These initial questions did not involve probabilistic reasoning and involved merely locating relevant data on the table and are commonly referred to 'Reading the Data' (Bright and Friel 1998).

Once children were aware of the workings of the competition and the mini-league table, they were challenged to make predictions using

Position	Team	Played	Won	Drew	Lost	Pts
1	England	4	4	0	0	8
2	Wales	4	3	0	1	6
3	Scotland	4	2	0	2	4
4	Ireland	4	1	1	2	3
5	Italy	4	1	0	3	2
6	France	4	0	1	3	1

Fig. 2. 2013 Royal Bank of Scotland 6 Nations League Table (after four matches)

knowledge derived from the mini-league table data. They were asked to make and justify their predictions in response to questions such as

What is the probability of Italy winning the six nations?

What is the probability of Wales finishing above England in the competition?

What is the probability that Scotland will finish last on the table?

These questions required them to consider both previous results (as indicated on the league table) alongside possible scores in the upcoming game. In the beginning, the majority of children tended to provide responses without reference to patterns in the data. It seemed that a number of the children who were selecting 'even' in response to various statements believed that each scenario was fair, i.e. that each possibility had an equal chance of happening. In this situation, as can be seen in the transcript below, the teacher drew children's attention to the data presented in the table. A careful analysis of these data supported them in making data-based inferences.

Teacher: What is the probability that Scotland will finish last on the table?

[*The majority of children hold up cards with* '*Even*' *or* '*Likely*'].

Tia can you tell me why you think it is likely?

Tia: [silence] I am not sure…

Teacher: Have a think about it for a minute. Kim what do you think?

Kim: An even chance.

Teacher: Even chance, why would you say that?
Kim: It could go any way.

Teacher: Ok. When deciding the chances, we need to examine the data and use this to make the best decision. Let's think. Hands up, how many matches have Scotland left to play?

Jim: 1.

Teacher: Each team has to play 5 times and at the moment each team has played 4 games. Everyone has 1 more match left to play. If Scotland lost that match is there a chance that they will go to the bottom of the league table?

Louise: Yes

[*The majority of the class are responding* '*No*']

Teacher: Louise, why do you say yes?

Louise: Because all the rest [of the teams] might go up.

Teacher: Ok. How many points did we say you get if you win a match?

Louise: 2 points.

Teacher: They [pointing to France in the table] have 1 point now. If they got 2 points that would give them how many points?Alisha: 3.

Teacher: Would that bring them up above Scotland?

Harry: No.
Teacher: So what are the chances of Scotland finishing at the bottom of the league table after 5 matches? Have a look at the league table, make a prediction and use your cards to show me what you think.

[*Children hold up the cards that best represent the chance matching the possibility of Scotland coming bottom of the league table*]

Ok. I see lots of cards that say 'impossible'. Lee, why did you say 'impossible'?

Lee: Because if France wins they'll get 2 more points and they'll move a little bit up the table cause then they would have 3 points. But Scotland already have 4 points and the most France could get is 3.

Our research indicates that there is the tendency for some children not to use the data to inform their predictions. We found it necessary to direct the attention of children towards using the data on past performance to justify their predictions about the ranking of the teams at the end of the league.

Using data regarding previous results

Attention then moved to the upcoming fixtures. The first game that was discussed was the game between Ireland and Italy. After enquiring about the

date of this fixture, attention was turned to the venue. Many children were aware that this game would be played in Rome. Children were then presented with a table containing the score statistics for the last years of Ireland versus Italy matches. Within the table, reference was made to which team had the 'home advantage' (Figure 3). The impact of playing a match at an 'away venue' was discussed fully. Children were aware of the possible disadvantages for the Irish team of playing 'away', making reference to the potential negative impact on Ireland's chances of winning. They made particular reference to possible effects of jet lag, a different climate, a smaller amount of supportive fans at the game venue and an unfamiliar pitch.

Teacher: Why do you think it is a disadvantage for Ireland to play in Rome?

Adam: The Italian support might help Italy to win.

Mark: Because they are not used to their [the Italian] pitch and they [the Italian team] are used to their own [pitch].

Children were then asked to use the statistics for the last 3 years of Ireland versus Italy games to make a data-based prediction on the probability of Ireland winning the match. Children were asked to record their prediction as well as their reasoning (Figure 4).

In both classes, children responded that Ireland were 'likely' or 'certain' to win. While some students were able to justify their decision from the data, e.g. *They have won every year so maybe they're a stronger team*, the teacher used this response to introduce the idea of 'bias' in making objective predictions:

Italy V Ireland

Year	Italy	Ireland
2010	11 points	29 points (home)
2011	11 points(home)	13 points
2012	10 points	42 points (home)
2013	?	?

Fig. 3. Score statistics for the 2010–2012 Ireland versus Italy matches

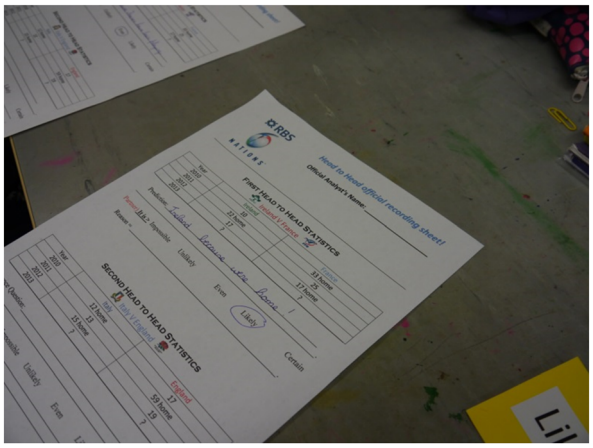


Fig. 4. Recording of students' prediction and justification

Who here wants Ireland to win? Hands up [*Every child in the class puts their hands up*]. Maybe then we might be biased. Biased means that we are one-sided, we might think that the team we support, Ireland, is definitely going to win without considering the facts. So maybe bias has affected your prediction!

The teacher told the class that they were now going to make predictions about another match that would take place the same weekend: Wales versus England. They were asked to consider if they might be biased when they predict who is going to win that match. The nature of responses suggests that they had gained an appropriate understanding, e.g. *No because it's not one of our countries*. *We are not from England or Wales*.

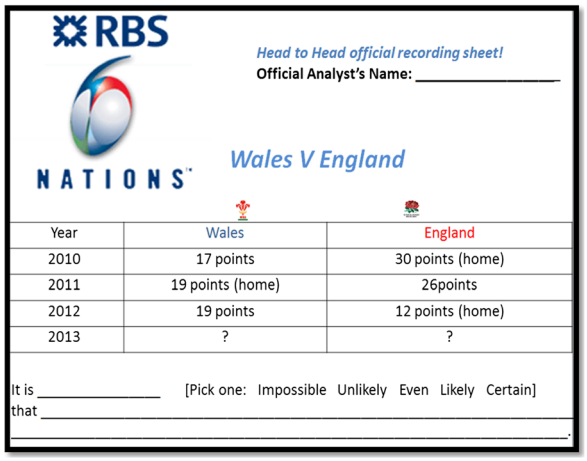
Children were then asked to pretend that they were TV analysts and to use the statistics regarding previous Wales versus England matches (Figure 5) to make and justify their prediction as to who would win the upcoming game.

While children's predictions varied, they demonstrated an increased ability to refer to data from previous matches in their justifications. They paid particular attention to a number of factors such as the impact of the location of the game (i.e. home versus away games), the ratio of wins to losses and most recent performance.

Tom: It is likely that Wales will win because they are at their own pitch

Tess: It is an even chance that either could win on Welsh ground because they have both won 'away' before

David: It is likely that England will win because


RBS NATIONS

Head to Head official recording sheet!

Official Analyst's Name: _____

Wales V England

Year	Wales	England
2010	17 points	30 points (home)
2011	19 points (home)	26points
2012	19 points	12 points (home)
2013	?	?

It is _____ [Pick one: Impossible Unlikely Even Likely Certain]
that _____
_____.

Fig. 5. Score statistics for the 2010–2012 Wales versus England matches

they have won 2 of the last 3 years
Jake: It is likely that Wales will win as they seem to be getting stronger each year, they won last year and they are used to their pitch

REFLECTIONS

This sequence of instruction draws on the connections between statistics and probability. It facilitates the children to use sports statistics to make predictions about future events, i.e. teams that will finish at the bottom/top of the Championship league table as well as outcomes to upcoming games. The activities support children in justifying their predictions by making reference to the data. The opportunity to analyse data and make inferences as a result are worthwhile activities. This article demonstrates that if the context is motivating and adequate support is provided, children in the senior classes of elementary school can make informed predictions and have a taste of the potential role of statistics and probability in their lives.

On reflection, the authors would consider a number of key prerequisites for this lesson. Firstly, children must have ample opportunity to use the language of probability before this sequence of instruction. Children must also be comfortable with the workings of a league table in order to engage fully with the activities. While these factors are important in facilitating optimum access to students, the biggest challenge for teachers in implementing this lesson is keeping the 'mathematics focus'. As is evident from this article, good questioning is essential in order to focus on the 'Stats', which are central to students' predictions and justifications. Otherwise, children may focus on periphery issues and create statements such as 'It is likely that a player will get injured in the game'. This sequence of instruction has the potential to develop relevant and transferable understandings.

References

Bright, G.W. and Friel, S.N. (1998). Graphical representations: helping students interpret data. In: Lajoie, S.P. (ed.) Reflections on Statistics: Agenda for Learning, Teaching, and Assessment in K-12, pp. 63–88. Hillsdale, NJ: Lawrence Erlbaum Associates.

Fernandez, C. and Yoshida, M. (2004). Lesson Study: A Japanese Approach to Improving Mathematics Teaching and Learning. Mahwah, NJ: Lawrence Erlbaum Associates, Publisher.

Fischbein, E. (1975). The Intuitive Sources of Probabilistic Thinking in Children. Dordrecht, The Netherlands: Reidel.

Fischbein, E., Nello, M.S. and Marino, M.S. (1991). Factors affecting probabilistic judgments in children and adolescents. *Educational Studies in Mathematics*, **22**, 523–549.

Greer, B. and Mukhopadhyay, S. (2005). Teaching and learning the mathematization of uncertainty: historical, cultural, social, and political contexts. In: G.A., Jones (ed.) Exploring Probability in School: Challenges for Teaching and Learning, pp. 297–324. Dordrecht, The Netherlands: Kluwer.

Jones, G.A., Langrall, C.W. and Mooney, E.S. (2007). Research in probability responding to classroom realities. In: Lester, F.K. (ed.) Second International Handbook of Research in Mathematics Teaching and Learning, pp. 909–955. Charlotte, NC: Information Age.

Kaplan, J., Fisher, D.G. and Rogness, N.T. (2010). Lexical ambiguity in statistics: how students use and define the words: association, average, confidence, random and spread. *Journal of Statistics Education*, **18**(2), 1–22. http://www.amstat.org/publications/jse/v18n2/kaplan.pdf

Leavy, A.M., Hannigan, A. and Fitzmaurice, O. (2013). If you're doubting yourself then, what's the fun in that? An exploration of why prospective teachers perceive statistics as difficult. *Journal of Statistics Education*, **21**(3), 1–26.

Leavy, A. and Hourigan, M. (2015). Motivating inquiry in statistics and probability in the primary classroom. *Teaching Statistics*, **37**(2), 41–47. DOI:10.1111/test.12062.

Lecoutre, M.P. (1992). Cognitive models and problem spaces in "purely random" situations.

Educational Studies in Mathematics, **23**, 557–568.

Lewis, C., and Tsuchida, I. (1998). A lesson is like a swiftly flowing river: how research lessons improve Japanese education. *American Educator*, **22**(4), 12-17, 50-52.

Metz, K.E. (1998). Emergent ideas of chance and probability in primary-grade children. In: Lajoie, S.P. (ed.) Reflections on Statistics Learning, Teaching and Assessment in Grades K-12, pp. 149–173. Mahwah, NJ: Erlbaum.

Tarr, J., Stohl Lee, H. and Rider, R.L. (2006). When data and chance collide: drawing inference from empirical data. In: Thinking and Reasoning with Data and Chance (68th Yearbook), pp. 139–149. Reston: NCTM.

Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science*, **185**, 1124–1131.

Van De Walle, J. (2013). Elementary and Middle School Mathematics: Teaching Developmentally (8th edn). Pearson International Edition. Boston: Pearson /Allyn and Bacon.

Watson, J.M. and Moritz, J.B. (2003). The development of comprehension of chance language: evaluation and interpretation. *School Science and Mathematics*, **103**, 65–80.