# Development of Fourth-grade Students' Understanding of Experimental and Theoretical Probability

Lyn English
*Queensland University of Technology*
<l.english@qut.edu.au>

Jane Watson
*University of Tasmania*
<Jane.Watson@utas.edu.au>

Students explored variation and expectation in a probability activity at the end of the first year of a 3-year longitudinal study across grades 4-6. The activity involved experiments in tossing coins both manually and with simulation using the graphing software, *TinkerPlots*. Initial responses indicated that the students were aware of uncertainty, although an understanding of chance concepts appeared limited. Predicting outcomes of 10 tosses reflected an intuitive notion of equiprobability, with little awareness of variation. Understanding the relationship between experimental and theoretical probability did not emerge until multiple outcomes and representations were generated with the software.

Probability experiences in the primary school usually focus on chance events with well-defined sample spaces, such as those of a die, spinner, or coin, where equally likely outcomes are expected (e.g., Jones, Langrall, & Mooney, 2007; Rubel, 2007). Given the variation that can occur in outcomes from small numbers of trials, the experimental probability (what is observed) does not always align with the expected or theoretical probability. When this occurs, students' intuitive notions of probability, such as equiprobability, can be challenged (Amir & Williams, 1999; Fielding-Wells, 2014; Khazanov, 2008). Although there has been substantial research on the teaching of probability, the literature still appears lacking on students' understanding of the connections between experimental and theoretical probability (Jones et al., 2007).

In addition, the important links between statistics and probability do not appear to have received the necessary attention, despite being a combined topic in many curricula including those in Australia (Australian Curriculum, Assessment and Reporting Authority [ACARA], 2013) and the USA (Common Core State Standards Initiative, 2012). In particular, variation, which lies at the heart of statistical reasoning and is a core component of beginning inference (English & Watson, 2013; Makar & Rubin, 2009; Shaughnessy, 2006), has been neglected in the primary school especially with respect to understanding probability. Yet variation is the foundation of statistics and provides the basis for expectation from which inferences can be drawn regarding decision making. The confidence with which one can make a decision, however, depends on creating a balance between variation and expectation (Watson, 2005). In the present study, fourth-grade students explored these two core understandings in their investigations of chance outcomes in tossing one, then two coins, multiple times; this paper addresses the students' learning involving one coin only. Specifically, we focus on students' expectations of the outcomes, how their expectations changed with repeated trials, and how they related experimental outcomes to theoretical probabilities based on large numbers of trials.

## Variation and Expectation in Probability

Although the research on primary school students' awareness of variation and expectation in statistical experiences is limited, it appears that their intuitive awareness of

2014. In J. Anderson, M. Cavanagh & A. Prescott (Eds.). Curriculum in focus: Research guided practice *(Proceedings of the 37th annual conference of the Mathematics Education Research Group of Australasia)* pp. 215–222. Sydney: MERGA.

variation develops earlier than their appreciation of expectation (Watson, 2005). Recognizing variation in determining expectation may be likened to Konold and Pollatsek's (2002) notion of identifying "signal" within "noise." However, if the signal or expectation with respect to a familiar chance situation is sufficiently strong or highly anticipated, then it may be that the noise is totally ignored or given limited consideration. We argue that a focus on both aspects is needed in developing the foundations of probability.

It is well known that cultural factors can influence students' probabilistic thinking (Amir & Williams, 1999), where out-of-school biases towards certain outcomes can dominate, such as tossing a "six" on the throw of a die being less likely than obtaining a "one" because the former outcome is more valued in many board games. Conversely, Rubel (2010) demonstrated how in-school biases regarding what it means to do mathematics can influence students' common response of "equally likely" when asked, "Which event is more likely, or are they equally likely: 7 tails out of 10 tosses of a fair coin or 700 tails of 1000 tosses?" (p.144). When probed with a real-life example, Rubel found students readily offered the correct response and could justify their reasoning.

Also linked with cultural influences, is the expectation that tossing one coin 10 times will yield *exactly* five heads and five tails. This may be considered an example of the "equiprobability" bias, that is, a tendency to view any random trials of an experiment as sufficient indication of equal outcomes (Amir & Williams, 1999; Khazanov, 2008; Lecoutre, Durand, & Cordier, 1990). Intuitive beliefs about "random devices" can further influence students' responses. For example, beliefs such as luck can contribute to the outcome of tossing a coin, how one tosses it has an impact, or a preferred outcome will be produced, can all impinge on children's understanding of formal probability (Amir & Williams, 1999).

Experiences in which children predict outcomes of trial events and then confirm or disconfirm their predictions enable them to experience the variation that occurs in random events. Past research has highlighted students' use of heuristics, such as representativeness, in making predictions about random events, for example, viewing successive heads on a fair coin more likely to be followed by a tail than another head (e.g., Fischbein & Schnarch, 1997). Other research (Konold, Pollastek, Well, Lohmeier, & Lipson, 1993; Rubel, 2007), however, has indicated that such heuristics do not adequately account for the decisions students make in determining probabilities. We argue that students' lack of awareness of the relationship between variation and expectation is an important factor, playing a major role in decisions regarding chance outcomes and in establishing the important theoretical and experimental understandings.

Given these concerns, the first year of our study served as a foundation period for developing an appreciation of variation and prediction as they arise in situations involving measurement data (English & Watson, 2013), leading to data generated in chance experiments. Specifically, we were interested in the following questions for the probability activity addressed here:

1. What are students' expectations of the outcomes when tossing a single coin?
2. How do students' expectations of outcomes change on repeated trials of tossing one coin?
3. How do students' relate experimental outcomes to theoretical probabilities based on large numbers of trials?

# Methodology

## *Participants*

Four grade 4 classes and one grade 4/5 class from a middle socio-economic Australian school participated during the first year of the three-year longitudinal study (2012-2014). For the present activity, we focus only on the grade 4 students (n=89; mean age = 9.7 years). English was a second language for 43% of these students.

## *Background and Design*

A design-based research approach was adopted, involving engineering an innovative educational environment to support the development of particular forms of learning and studying the learning that takes place in the designed environment (Cobb, Jackson, & Munoz, in press). An important component of this approach was the teachers' participation in professional development sessions in preparation for their implementation of the activity; these were followed by debriefing sessions where we reflected on the students' and teachers' development as well as our own.

Data collection included video- and audio-taping of two focus groups in each class (consisting of two students per group) and all class discussions, which were subsequently transcribed for analysis (non-consenting students were excluded). Our data for this report also included camera shots, student workbook responses, and notes of researcher observations. Other data collected across the year comprised measurement of student progress through assessments and an end-of-year survey on the year's activities. Teacher and student interviews were also conducted.

## *Activity and Implementation*

The activity was created in collaboration with the teachers and formed part of their regular mathematics program in the area of data and probability. The teachers implemented the activity across a whole school day. The authors were in attendance for the entire activity to observe the students' learning in each classroom. Only the first component of the activity is addressed here, namely repeated trials of tossing one coin both manually and with the Sampler in the graphing software *Tinkerplots* (Konold & Miller, 2011).

A detailed lesson outline was prepared for the teachers, as was a workbook for students. The latter included instructions with screen dumps for setting up the Sampler in *TinkerPlots*, which the students were expected to create for themselves having been introduced to the software in a previous activity. The students worked in pairs with a laptop computer to use *TinkerPlots*. Although conducting the activity of coin tossing and running simulations was undertaken in pairs, students were asked to write their own answers and explanations in their workbooks.

The activity commenced with a review of everyday events that are "certain," "uncertain," or "impossible," followed by a brief discussion of the possible outcomes of tossing dice and coins and the certainty of each outcome. Prior to tossing one coin 10 times, students predicted the outcomes. They then recorded the outcomes, combined and plotted outcomes for the class, and discussed how close the results came to their predictions (expectation). Students subsequently used computers to replicate their trials with the *TinkerPlots* Sampler, which can create a much larger number of trials. The students observed variation in all of the trials, with an understanding of expectation emerging as they made predictions and observed the results approach the theoretical values of the probabilities.

The objectives of the activity, which aligned with the requirements of the version of ACARA curriculum released at the time (2012), included: (a) Describing chances/probabilities of coin outcomes; (b) Appreciating that “Heads” and “Tails” cannot occur at the same time; (c) Appreciating that the chance/probability of getting a “Head” on one toss of a coin will not affect the outcome on the next toss; (d) Using fractions to express the chance/probability of various coin outcomes; (e) Recognizing that the probabilities of all possible outcomes of an experiment total one; (f) Recognizing both the variation in outcomes for a small number of trials and a tendency for a large number of trials to approach the expected probabilities; and (g) Relating experimental outcomes to theoretical probabilities.

### *Data Analysis*

The data analysed in addressing the present research questions were sourced from selected items in the students’ workbooks. Only data from consenting students (n=89) were analysed. As not all students answered all questions, the number of responses analysed per workbook item vary, as indicated. The data were clustered according to similarities in the understanding displayed in the workbook responses. The coding of the responses generally reflected their correctness, together with increasing levels of sophistication in understanding. Responses ranged from no response or an idiosyncratic one, through to responses that included explicit reference to core ideas such as randomness, variation, independent events, and clustering of outcome data. The authors collaborated on the coding of the data, with the senior research assistant checking the codes produced.

## Results

### *Research Question 1: Expectations of Outcomes in Tossing a Single Coin*

Students’ predictions of the outcomes of tossing a coin once, their certainty of their predictions, and how these compared with the actual outcomes indicated a basic understanding of chance events. Seventy-five percent of students (n=89) indicated “partially certain” and 24% “uncertain” in regard to their prediction (Items 1 and 2). In explaining how their outcomes compared with their predictions (Item 3), however, only 4% could offer a reason related to chance (e.g., “My hypothosis [sic] was it would land on Heads and it did, however I just guessed any side because there is a 50/50 percent chance landing on either side”). Most responses simply referred to the outcome only. Nevertheless, a greater appreciation of chance events was evident in their response to Item 4, which asked whether the type of coin influences the outcome, with 42% of responses offering reasons pertaining to chance or to all coins having two sides (e.g., “No, it is still a 2 sided object and has the same probability”). Likewise, 88% of responses (n=87) to Item 5 indicated that one person’s toss would not influence another’s, with 55% of students providing a reason related to chance or independence (e.g., “No, because it is an independent toss & it dosen’t [sic] matter what the previous outcome was”).

### *Research Question 2: Expectations of Changes in Outcomes on Repeated Trials*

In investigating how students might change their expectations of outcomes on repeated trials, they were initially asked: (a) how many heads and tails they think would be produced if a single coin were tossed 10 times and why (Item 8), (b) if the coin were tossed another 10 times, whether exactly the same number of heads and tails would be obtained and why (Item 9), and (c) prior to actually tossing a coin 10 times, how many heads they would predict and why (Item 10). On recording the outcomes of actually tossing a coin 10 times, the students were to describe how close their prediction was to the outcome (Item 14).

Nearly half (45%) of the student responses (n=86) to Item 8 stated a definite equal probability of obtaining heads or tails, while 35% expressed uncertainty and/or indicated some form of variation or randomness, such as, "Well I could get any mixed number of all of them so it doesn't exactly matter," and "I would say that half could be 60% and half 40%. It is rarely half, half."

In considering whether the same outcomes would occur on another 10 tosses (Item 9, n=86 responses), the majority of students (77%) did not believe this would be the case. However, only 29% offered an explanation highlighting an awareness of randomness or independence, such as, "No, since it will not affect the other one," and "If I tossed another coin 10 times I could get any number of heads and tails because it is a two sided coin so there is a 1/2 chance."

In predicting how many heads out of 10 they would actually toss (Item 10, n=88 responses), a surprising 42% gave no response or an idiosyncratic reason. This might have reflected their uncertainty in actually making such a prediction. Twenty-seven percent stated a definite 5 heads or equal chance, while the remaining 31% expressed uncertainty due to chance, with 20 students referring to randomness or some form of variation. Responses of the last type included, "6 heads and 4 tails because it's uncertain if you get 5 heads and 5 tails, you always get differents [sic] numbers."

All but eight students (n=87 responses) offered an acceptable answer in describing how close their prediction was to their outcome, with students noting that these aligned or they differed. Sample responses included, "Our prediction was so close we thought head was going to be 5/10 tosses but the outcome was 6/10 tosses of head [sic]," and "I was two heads more than my prediction and two tails less than my prediction."

### *Research Question 3: Relating Experimental Outcomes to Theoretical Probabilities Based on Large Numbers of Trials*

The next component of the activity involved collating the class outcomes of the 10 tosses, with students plotting the total number of heads for each group on a number line in their workbooks (Item 15). The students were to describe the shape of the plot (Item 16), whether this is what they would have expected and why/why not (Item 17), whether every group obtained five heads out of 10 and why not (Item 18), and where the values in the plot were clustered and why (Item 19). Examples of student plots for two different classes are shown in Figure 1. In the first example, the student recorded the frequencies of heads obtained by each of the 9 groups in her class and also highlighted trends in the data. She explained, "It started from going up then down then up then down and then up and then down." The student in the second example created a grid and labelled the group numbers (13 groups) in plotting the frequencies of heads.

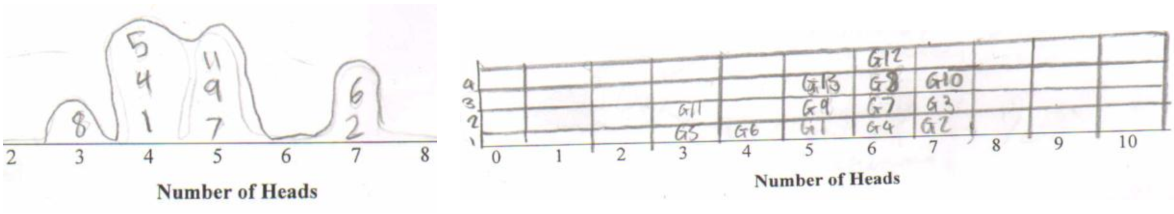


*Figure 1.* Number of heads from 10 tosses for groups in two different classes.

Just over half the students (53%) gave an acceptable description of the class plot (Item 16, n=88 responses), with frequent reference to a mountain shape and some to geometrical properties, such as, “The graph looks like a mountain;” “Looks like a triangle. City. A pyramid;” and “Round. Straight. Horizontal. A symmetrical. Long hill.” Fewer students (33%) mentioned a middle value or a clustering of values on the plot, with comments including, “The middle of the graph has a lot of heads but if you see under five and over five there’s only a small amount (sic) of trials,” “The graph above gets bigger and bigger as it goes to the middle (5) but goes down after the 5, and finishes off to nothing,” and “Its small and then steps up to 6 then down to 3 then 1. It is kind of like a big box in the middle.”

Although all of the students stated that the plot was what they had expected (Item 17, n=88 responses), just under half (43%) provided an explanation that indicated recognition of a middle value or clustering. Examples here included, “I expected the numbers to be in the middle because I didn’t think that it’ll be spread around;” “Yes because it is rare to get very high or very low numbers;” and “It is sort of what I expected. I expected about 50% to be heads but I didn’t expect there to be some odd recordings (8 or 10).”

The responses (n=87) to Item 18 (whether every group obtained five heads out of 10 and why not) were rather disappointing, however. Only 28% offered an explanation based on chance such as, “No, no matter how likely it is, it is not certain,” “No, just because it is an equal possibility doesn’t mean it will be equal,” and “No, the chance of that happening is almost impossible.” In Item 19 (n=88 responses), however, when specifically asked where the values in the plot were clustered and why, just over half (51%) gave an explanation related to a clustering in the middle, such as, “In the middle because that is around 1/2.” Only 30% of responses included a specific reference to chance, such as, “Most of them are in five, six, and seven because they are quite close to the ‘50-50’ chance.”

The final part of the activity involved students in using the Sampler in *TinkerPlots* to make repeated trials of 10 tosses, recording the number and percentage of heads generated each time, and describing the variation they noticed. Generating more samples of trials of 10 tosses with the Sampler clearly enabled the students (n=88) to see the variation in the outcomes, with 91% offering an explanation that corresponded with the data generated. Students’ responses here included, “Most of them cluster together (4-7), they variate [sic] from 30% to 70%,” “Some results are the same and there is a lot of variation,” and “I noticed that the range has a big difference because 10% - 90% is really big.”

The activity concluded with the students increasing the number of tosses to multiples of 100 and 1000, and noting changes in the percentages of heads as the number of repeats increased. The students had some difficulty in interpreting the percentages in their observations (n=88; 41% difficulty), although 59% of responses explained that the results approximate 50% more closely as the number of trials increases. For example, students explained, “As the number of repeats [tosses] increases, the range gets smaller,” and “In the 100's section it varied from 43% - 57%, in the 1000s section it varied from 50% - 51%. The thousands are closer to 50% than the 100s..."

## Discussion and Concluding Points

This paper has addressed fourth-grade students' understanding of variation and expectation in experimenting with tossing a coin, observing variation in expected outcomes, and noting how experimental probability approximates theoretical probability based on large numbers of trials. The results suggest the students were aware they could not predict with complete certainty that a head would be obtained on one toss of a coin, yet very few could offer a reason for this. Students displayed a greater awareness, however, of the irrelevance of the coin type and the independence of the tosses.

An intuitive notion of equiprobability was reflected in students' expectations of the outcomes of tossing their coin 10 times, with only a small number of students aware of the variation that occurs in events of this nature. Few students gave an indication that they understood that such outcomes are "rarely half, half," suggesting their previous experiences had focused on theoretical probability at the expense of the experimental. Although the majority of students were aware that exactly the same outcomes would not be the case on a further 10 tosses, they nevertheless offered little explanation as to why this would be the case. Likewise, the students gave limited indication of their predicted outcomes on tossing their coin 10 times, suggesting weak or no exposure to a question of this nature. Some students did justify their responses with reference to randomness or variation, but on the whole, it appeared that their expressed understanding of variation in chance events was lacking. It was not until the students experimented with *TinkerPlots* to observe and display the outcomes of multiple trials that an understanding of the relationship between experimental and theoretical probabilities developed.

By collating and plotting all of the class outcomes of 10 tosses, the students could see how their group experimental results were aligning with the expected outcomes. The students were generally able to describe the shape of the plot and recognize the clustering of values towards the middle, but not many could offer an explanation for this based on chance concepts. With the introduction of the Sampler in *TinkerPlots*, however, where numbers of tosses of multiples of 100 and 1000 could be performed and changes in the percentages of heads noted, the students were able to describe the reduced variation in the outcomes and offer explanations that corresponded with the data. Although the students had not formally studied percentages, over half were able to explain why the outcomes approximate 50% on large numbers of trials.

Further insight into students' emerging understanding of the relationship between experimental and theoretical probabilities was evident in the transcripts of the class discussions. It was clear that many students had not consolidated the language necessary to verbalize this emerging understanding. The importance of developing students' chance language as they construct their knowledge of, and experiences with, probability across the primary school has long been stressed in the literature (e.g., Jones et al., 2007; Khazanov, 2008; Watson, 2007). More emphasis needs to be placed on the statistical notions of variation and expectation in introducing chance language in the primary school. For example, in problems involving weather forecasting, students need to understand terms such as "It is *likely* to rain means we *expect* rain but the weather may *vary* and be sunny instead." Further, as Khazanov (2008) pointed out, students might have mastered the idea of percent but have difficulty in interpreting it in probabilistic contexts, indicating the need to broaden the meanings associated with this term. We consider it essential to include these foundational terms in students' vocabularies as they express their emerging understanding of probability.

## References


Amir, G. S., & Williams, J. S. (1999). Cultural influences on children's probabilistic thinking. *The Journal of Mathematical Behavior, 18*(1), 85-107.

Australian Curriculum, Assessment and Reporting Authority. (2013). *The Australian Curriculum: Mathematics, Version 5.0, 20 May, 2013.* Sydney, NSW: ACARA.

Cobb, P., Jackson, K., & Munoz, C. (in press). Design research: A critical analysis. In L. D. English & D. Kirshner (Eds.), *Handbook of International Research in Mathematics Education* (3rd edn.). New York, NY: Routledge.

Common Core State Standards Initiative. (2012). *Standards for Mathematical Practice*. http://www.corestandards.org/Math/Practice (Retrieved 21/12/2013).

English, L. D., & Watson. (2013). Beginning inference in fourth grade: Exploring variation in measurement. In V. Steinle, L. Ball, & C. Bardini (Eds.), *Mathematics Education: Yesterday, Today, and Tomorrow*. Proceedings of the 36th Annual conference of the Mathematics Education Research Group of Australasia,( pp. 274-281). Melbourne, Vic.: MERGA.

Fielding-Wells, J. (2014, In press). Where's your evidence? Challenging young students' equiprobability bias through argumentation. Paper to be presented at the *9th International Conference on Teaching Statistics*. Flagstaff, USA.

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

Jones, G. A., Langrall, C. W., & Mooney, E. S. (2007). Research in probability: Responding to classroom realities. In F. K. Lester (Ed.), *Second Handbook of Research on Mathematics Teaching and Learning* (pp. 909-955). Reston, VA: NCTM.

Khazanov, L. (2008). Addressing students' misconceptions about probability during the first years of college. *Mathematics and Computer Education, 42*(3), 180-192.

Konold, C., & Miller, C. D. (2011). *TinkerPlots*: *Dynamic Data Exploration* [computer software, Version 2.0]. Emeryville, CA: Key Curriculum Press.

Konold, C., & Pollatsek, A. (2002). Data analysis as the search for signals in noisy processes. *Journal for Research in Mathematics Education, 33*, 259-289.

Konold, C., Pollatsek, A., Well, A., Lohmeier, J., & Lipson, A. (1993). Inconsistencies in students' reasoning about probability. *Journal for Research in Mathematics Education, 24*, 392-414.

Lecoutre, M.-P., Durand, J.-L., & Cordier, J. (1990). A study of two biases in probabilistic judgements: Representativeness and equiprobability. In J.-P., Caverni, J.-M., Fabre, & M. Gonzalez (Eds.), *Cognitive Biases* (Vol. 68, pp. 563-575). Amsterdam: Elsevier.

Makar, K., & Rubin, A. (2009). A framework for thinking about informal statistical inference. *Statistics Education Research Journal, 8*(1), 82-105. Retrieved from http://iaseweb.org/documents/SERJ/SERJ8(1)_Makar_Rubin.pdf

Rubel, L. H. (2010). Is 7/10 always equivalent to 700/1000? *Mathematics Teacher, 104*(2), 144-147.

Rubel, L. H. (2007). Middle school and high school students' probabilistic reasoning on coin tasks. *Journal for Research in Mathematics Education, 38*(5), 531-556.

Shaughnessy, J. M. (2006). Research on students' understanding of some big concepts in statistics. In G. Burrill & P. Elliott (Eds.), *Thinking and Reasoning with Data and Chance* (pp. 77-98). Reston, VA: National Council of Teachers of Mathematics.

Watson, J. M. (2005). Variation and expectation as foundations for the chance and data curriculum. In P. Clarkson, A. Downton, D. Gronn, M. Horne, A. McDonough, R. Pierce & A. Roche (Eds.), *Building Connections: Theory, Research and Practice*. Proceedings of the 28th annual conference of MERGA, (pp. 35-42). Sydney: MERGA.

Watson, J. M. (2007). Assessing statistical literacy through the use of media surveys. In I. Gal & J. Garfield (Eds.), *The Assessment Challenge in Statistics Education* (pp. 107-121). Amsterdam: IOS Press.