Data and Chance in the senior primary classes

Commissioned research paper

Dr Aisling Leavy

Mary Immaculate College

NCCA

An Chomhairle Náisiúnta Curaclaim agus Measúnachta
National Council for Curriculum and Assessment

June 2020

Importance of data and chance for children's learning

In a rapidly changing world where technological innovation is omniprescent, new competencies and literacies are necessary. The increasing use of data for decision making and prediction requires that educators prioritise the statistical and probabilistic reasoning of children (Franklin et al., 2007). Importantly, these data are not abstract quantities– they are *numbers in context* (Cobb & Moore, 1997) and originate from a range of contexts such as health (e.g. disease), environmental (e.g. pollution), and economic (e.g. sales) contexts. Thus, the driving goal for the study of data and chance in schools is to develop statistically literate citizens who can manage uncertainty and risk, make reasonable evidence-based arguments and critically evaluate data-based claims. As children enter the senior classes they have developed the fundamental statistical and probabilistic understandings, alongside the cognitive resources, that are necessary to evaluate risk and interrogate data.

Investigations develop statistical and probabilistic reasoning

Most children and adults have adequate procedural knowledge to compute statistics and probabilities, however, do not demonstrate understanding of what these measures represent and when you might use them (Hiebert & Lefevre, 1986). This lack of functional literacy arises when school mathematical content is presented in a manner that is isolated from purposeful activity and learning experiences, thus making it less coherent (Bakker & Derry, 2011). What is required are experiences that make powerful ideas accessible to children (Greer et al., 2007) by connecting mathematical ideas within rich contextual situations.

For illustrative purposes, we shall consider statistical investigations. Statistical investigations are rich contextual situations that develop statistical understanding by engaging children in the reasoning and processes of a statistician during data-based enquiry. Moreover, they ensure that children access powerful statistical ideas within authentic and age appropriate learning environments. One particular inquiry cycle, termed the PPDAC (Problem, Plan, Data, Analysis, Conclusion) cycle, has gained traction in statistics education (Wild & Pfannkuch, 1999). Figure 1 illustrates the investigative cycle. The cycle can be used to both guide planning for the teaching of data and to locate and address the primary curriculum learning outcomes in data.

Statistical investigations are motivated by a compelling and meaningful question (ideally posed by the children); are situated within an engaging context; and produce data that is sufficiently complex so as to support reasoning and discussion (Leavy & Hourigan, 2016; Makar, 2018; Shaughnessy, 2007). As children move through different stages of statistical inquiry, they develop understandings of key statistical concepts (see section 3), engage with mathematical processes (see section 4), see how statistical concepts are related to each other and experience their utility in problem solving.

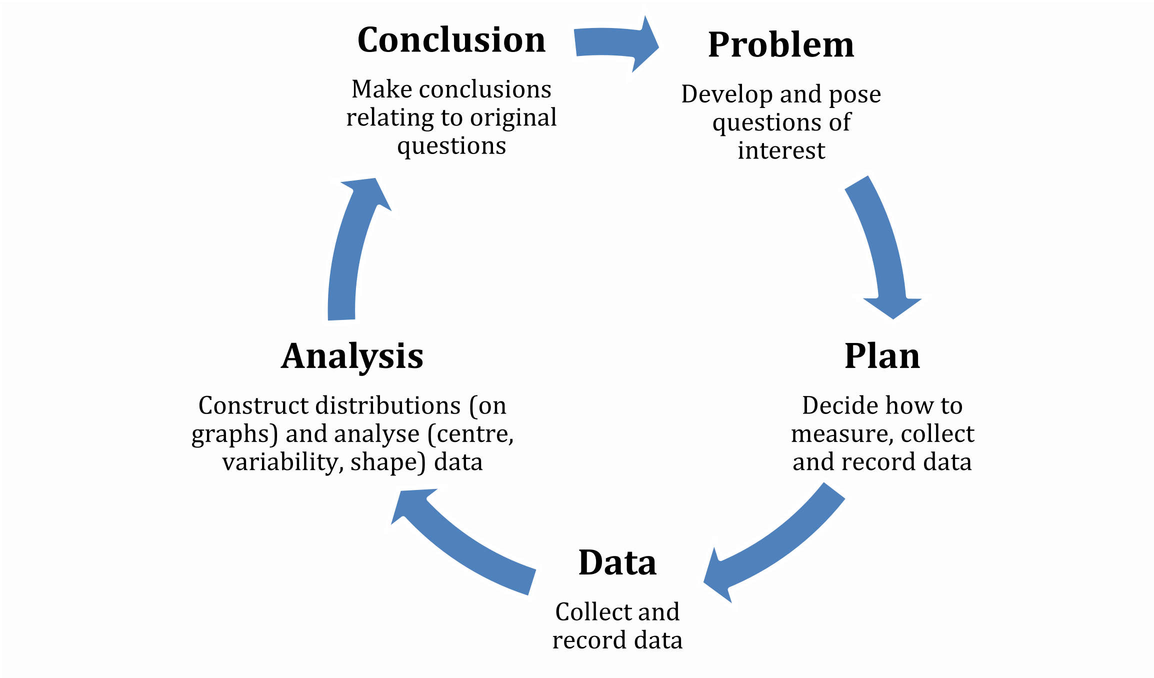

Figure 1: Statistical investigative cycle

Key concepts associated with the domain of data and chance

The key concepts that underpin statistical thinking and reasoning include: distribution, centre, variation, sampling and inference. Opportunities to develop understanding of these concepts occurs in the analysis stage of the PPDAC cycle (figure 1).

Distribution is the arrangement of observations (in other words, data) along a scale of measurement. In figure 2, the data values represent the resting heart rates of children and are arranged using a common scale of measurement (counts of heart beats per minute). The result is a distribution – a picture of the set of data that embodies its structural properties as a whole. A focus on distribution allows the identification of patterns and relationships in collections of data. From a pedagogical perspective, graphing a collection of data, as part of the analysis phase of the PPDAC cycle, allows learners to visualise the distribution of data. When presented with distributions of meaningful data, children demonstrate relatively sophisticated reasoning (Cobb, 1999), recognise landmarks and trends (Friel, Mokros, & Russell, 1992) and focus on variation in distributions (Cobb, 1999; Konold & Pollatsek, 2002; Petrosino, Lehrer, & Schauble, 2003; Watson & Kelly, 2002).

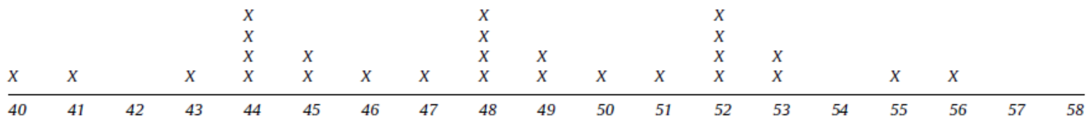

Figure 2: Distribution of 'What is your resting heart rate?' data

The emergence of distribution as a focus in primary level has motivated a re-focus on *centre* and *variability* as big ideas in statistics. These measures describe, summarise and act as representative values for distributions. For example, if asked to describe or summarise the resting heart rates (figure 2), one may refer to the heart rates as ranging (i.e. variability) between 40-56 beats per minute, or a mean of 44.2 and median of 48 beats per minute (i.e. centre). Hence consideration of centre and variability now arise almost naturally from the need to characterise distributions.

Research reveals a number of ways in which children coordinate these centre and variability when describing, representing and summarising data (Leavy & Middleton, 2011). There is also evidence of children's use of informal measures such as modal clumps and intervals and the use of proportional reasoning to determine representative values of distributions (Cobb, 1999; Konold et al., 2002; Lehrer & Schauble, 2002). Research also cautions that procedural fluency in computing centres does not indicate the development of associated conceptual knowledge and recommends postponement of the introduction of algorithms until children have developed conceptual understanding. Beyond computation, one of the overriding features of the mean and median that constitutes difficulty for children concerns the concept of representative value. Hancock et al. (1992) found that students did not recognise instances in which the mean could be used to typify or represent a data set, as indicated by the lack of instances where the mean was used to compare two groups of unequal size. Similarly, students able to calculate medians may not necessarily recognise medians as measures of center or as group descriptors of data (Bakker, 2004; Konold & Higgins, 2003). In fact, many students see the median as a feature associated with a particular data value in the middle of the group rather than as a characterisation of the entire group (Bakker et al., 2005). The ability, then, of students to compute representative values when specifically instructed (Mokros & Russell, 1995) compared to their inability to construct and use representative values in other situations (Hancock et al., 1992) suggests that students may not understand the important role that representative values play in data analysis.

Variability is a key concept in statistics – it is this attempt to account for and model variation in data that defines statistics as a discipline. For example, figure 2 captures the variability in the resting heart rates; if there was no variability then all heart rates would be identical and we would have no interest or need to collect data. Variability plays a central role in children's statistical thinking (Cobb, 1999; Konold & Pollatsek, 2002; Watson & Kelly, 2002; Watson, 2018). Primary children demonstrate strong understandings of variability when engaged in rich contextual investigations that involve comparing data sets (Jones et al., 2001), examining distributions of data (Lehrer and Schauble, 2002) and engaging in data modeling activities (Leavy & Hourigan, 2018; Lehrer & Romberg, 1996).

There is growing recognition of the importance of developing young students' *informal inference*. Informal inference is "the way in which students use their informal statistical knowledge to make arguments to support inferences about unknown populations based on observed samples" (Zieffler, Garfield, delMas, & Reading, 2008, p. 44). For example, children can examine the data in figure 2, or measures of centre and variability calculated from these data, to make inferences or predictions about the resting heart rate of a child who may be absent from school on the day data were collected. Inference fits well within the cycle of statistical inquiry (Pfannkuch, 2006), can be easily incorporated through the use of open questions (Leavy, 2010) and supports the generation of conclusions in the final stage of the PPDAC cycle. In fact, it has been argued that "inference and statistical investigation cannot be separated" (Makar and Rubin, 2007, p. 3). Articulation of uncertainty is an essential component of the statistical thinking process in which decisions and predictions are made on the basis of data in everyday contexts. Probability is the measurement we use to quantify the uncertainty about an outcome. Three main perspectives on the quantification of uncertainty exist. *Classical* probability, referred to as theoretical probability in school curricula and computed based on an analysis of sample space, refers to the ratio of the number of favourable cases in an event to the total number of equally likely cases. The *frequentist* perspective, called experimental probability in school curricula and estimated based on observed results from an experiment or simulation, defines probability of an event as the limiting value of its relative frequency in a large number of trials. *Subjective* probability is considered as a personal degree of belief, is subject to change based on personal judgment and information available about a given outcome, and is believed to be closely related to what people commonly use for everyday reasoning.

Gal (2005) lists five *knowledge* elements that are building blocks of probability literacy as:

- the exploration of *big ideas* (variation, randomness, independence, predictability/uncertainty),
- estimating *probabilities*,
- using *language* to communicate chance,
- understanding the *context* where probabilities are applied,
- considering *critical questions* when dealing with probabilities.

In addition to these knowledge elements, the three *dispositional* elements of having a *critical stance, beliefs and attitudes* and *personal* sentiments regarding uncertainty and risk, are underscored by Gal (2005) as equally important building blocks of probabilistic literacy.

Jones et al. (1997; 1999) propose four levels of probabilistic thinking that take into account these three perspectives (see table 1). Probabilistic thinking they describe as 'children's thinking in response to any probability situation' (1999, p. 488). Probability situations are situations involving uncertainty when engaging in an activity or random experiment where more than one outcome is possible; in turn, the actual outcome cannot be predetermined but only inferred. Aspects of all of these levels of probabilistic thinking described in table 1 can be seen in chance activities at primary level.

Table 1. Jones et al. (1997) Framework for Assessing Probabilistic Thinking

Level	**Description**
1	Students make intuitive and subjective judgments influenced by their imagination and irrelevant aspects – as a result, it is associated with subjective thinking,
2	Students often make inflexible attempts to quantify probabilities – thus it is transitional between subjective and naive quantitative thinking.
3	Students draw on informal quantitative thinking and use more generative strategies in listing the outcomes of two-stage experiments and in coordinating and quantifying thinking about sample space and probabilities.
4	Students demonstrate the use of valid numerical measures to describe the probabilities – consequently, this incorporates numerical thinking.

The relationship between mathematical processes and meta-practices

Research suggests that data and chance activities, of the nature advocated throughout this report, support the development of mathematical proficiency by engaging children in the senior classes with processes such as *reasoning, communicating* and *problem solving*. Abstract and quantitative *reasoning* are supported by the representations constructed in statistical and probabilistic contexts. Quantitative reasoning entails "habits of creating a coherent representation of the problem at hand ... and attending to the meaning of quantities" (CCSI, 2010, p. 6). Furthermore, when explaining the relationship between the real-world context of data and chance problems and the symbolic representation of those contexts, students must articulate their thinking to others and listen to and make sense of others' thinking and explanations – thus providing opportunities to reason abstractly and quantitatively. In addition, the use of technologies such as online graphing tools and simulations can enhance discourse and facilitate high-level mathematics thinking (Moss & Grover, 2007). Using calculators and technologies to perform tedious calculations (such as means and probabilities) or lengthy simulations (spinners or coin tosses for a large number of events) carries the procedural load and thereby frees up valuable classtime to support inquiry and higher order thinking and reasoning (Leavy & Hourigan, 2015).

When classrooms involve children in *doing* mathematics, then classrooms are not silent places (Lampert & Cobb, 2003). When engaged in data and chance activities, groups of students work together in inquiry-based learning environments where participation in (oral and written) *communication* is essential in order to learn. For example, as part of the statistical investigative cycle students must identify a question of interest, plan data collection strategies, communicate the data through a variety of representations (tables, graphs), and when presenting findings must engage in mathematical argumentation, produce evidence and explain their reasoning to others.

Similarly, chance activities require children to choose words to describe likelihoods of events occurring or the fairness of activities; they are then required to take and defend positions against alternative views and use debate to resolve conflicting views and arrive at common understandings around likelihoods. It is through these forms of data and chance communication, that students not only clarify and expand their ideas and understandings of mathematical relationships and mathematical arguments, but they also learn to communicate mathematically (Cobb, Yackel & Wood, 1989).

Statistical investigation is inherently a *problem-solving* process (Marriott, Davies & Gibson, 2009). It starts with a problem; students then collect data, analyse it and draw conclusions. They have to decide whether their conclusions provide insights, and in some cases a sensible solution, to the problem initially posed. Thus, students are active participants in the sense-making process and are responsible for making sense of the problems. Similarly, chance activities engage children in determining whether one outcome is more likely to occur than another. In doing so, they develop ways of effectively planning and being systematic in their organisation of activities. There is an abundance of research advocating the active learning of chance and statistics using real data and a problem solving approach (Franklin & Mewborn, 2006) with the central belief being that using a problem solving approach in the teaching of chance and statistics is of great benefit to both teachers and learners (Groth, 2006).

In summary, this research strongly suggests that when certain meta-practices (in this case, *promoting math talk*, *mathematical modelling* and the use of *cognitively challenging tasks;* cf. Dooley et al., 2014) permeate pedagogical approaches in data and chance, they provide children in the senior classes with the opportunity to engage in reasoning, communicating and problem solving and thus develops *mathematical proficiency*.

Key messages

The study of data and chance has distinctive characteristics that are not encountered in other areas of mathematics and which broaden and enrich the mathematics curriculum. This creates a challenge, however, in that these distinctive characteristics must be addressed explicitly through appropriate pedagogies and activities; otherwise children may not encounter them in their primary education.

More specifically, engaging in *cycles of statistical inquiry* (such as the PPDAC cycle) requires children to communicate and make sense of the very data they encounter and develops their ability to think critically. Through constructing *distributions* of data, and considering *centre* and *variability* within such distributions, children in the senior classes make *informal inferences* and conjectures as they interpret, justify and verify the patterns and relationships they notice with data. Children in the senior classes also require opportunities to work with multiple conceptions of probability: (a) *subjective,* (b) *classical or theoretical,* and (c) *frequentist or experimental*. They benefit from engagement with activities that require them to interpret the *language of chance*, to consider *randomness*, and to understand that events may vary in their degree of *likelihood* and *independence.*

Such tasks support children in developing *dispositions* that enable them to critically evaluate the probabilistic statements they encounter.

Thus, by engaging with these *big ideas* in data and chance, children in the senior classes not only develop the distinctive characteristics that are fundamental to statistical and probabilistic reasoning, they also engage in important mathematical processes such as understanding and connecting, reasoning, communicating, applying and problem solving. These understandings and processes are necessary for investigating, analysing, interpreting, explaining, and making sense of the world in which they live.

Chance favors the prepared mind.

Louis Pasteur (1854)

Glossary

Independence refers to events that are unconnected and one event cannot be predicted from another. Gal argues that understandings of variation, randomness and independence contribute to understanding of the complementary ideas of predictability and uncertainty.

Modal clumps are a distinguishable range or cluster of data at the heart of a distribution of values which suggest the variability and average of the data values.

Predictability and uncertainty relate to our knowledge about the likelihood of a certain event (e.g. rain).

Randomness is the result of a process where an event occurs without some underlying deterministic cause that is fully predictable (Beltrami, 1999).

Variation in the context of probability, refers to the idea that events and processes vary in how certain we are that we can predict how they will unfold (thus it underlies frequentist views of probability).

References

Bakker, A. (2004). *Design research in statistics education: On symbolizing and computer tools.* Utrecht, the Netherlands: CD Beta Press.

Bakker, A., Biehler, R. & Konold, C. (2005). Should young students learn about boxplots? In G. Burrill (Ed.), *Curricular development in statistics education*. International Association for Statistical Education (IASE) Roundtable, Lund, Sweden, 28 June to 3 July, pp. 163–173, http://www.stat.auckland.ac.nz/~iase/publications.

Bakker, A. & Derry, J. (2011). Lessons from inferentialism for statistics education. *Mathematical Thinking and Learning, 13*(1-2), 5-26.

Beltrami, E. (1999). *What is random? Chance and order in mathematics and life.* New York: Copemicusl Springer-Verlag.

Cobb, P. (1999). Individual and collective mathematical development: The case of statistical data analysis. *Mathematical Thinking and Learning, 1*(1), 5–43.

Cobb, G. W. & Moore, D. S. (1997). Mathematics, statistics, and teaching. *The American Mathematical Monthly, 104*(9), 801–823. doi:10.2307/ 2975286

Cobb, P., Yackel, E. & Wood, T. (1989). *Young Children's emotional acts while Doing mathematical Problem solving*. In D.B. McLeod and V.M. Adams, (Eds) *Affect and Mathematical Problem Solving: A new perspective* (pp. 117–48). New York: Springer-Verlag.

Common Core Standards Initiative (CCSI). (2010). *Common Core State Standards for Mathematics*. Retrieved from http://www.corestandards.org

Dooley, T., Dunphy, E. & Shiel, G. (2014). Mathematics in Early Childhood and Primary Education (3-8 years). Teaching and Learning. Research Report No. 18. NCCA, Dublin.

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M. & Scheaffer, R., (2007). *Guidelines and Assessment for Instruction in Statistics Education (GAISE) Report: A Pre-K-12 Curriculum Framework.* Alexandria, VA: American Statistical Association.

Franklin, C.A. & Mewborn, D.S. (2006). "The Statistical Education of Grades Pre-k-12 Teachers. A shared responsibility," *Thinking and Reasoning with Data and Chance*, 68th yearbook of the National Council of Teachers of Mathematics.

Friel, S.N., Mokros, J.R. & Russell, S.J. (1992). *Used Numbers: Middles, Means, and In-Betweens.* Palo Alto, Calif.: Dale Seymour Publications.

Gal, I. (2005). Towards "probability literacy" for all citizens: Building blocks and instructional dilemmas. In G. A. Jones (Ed.), *Exploring probability in school. Challenges for teaching and learning* (pp. 39–63). Dordrecht, The Netherlands: Kluwer.

Greer, B., Verschaffel, L. & Mukhopadhyay, S. (2007). Modelling for life: Mathematics and children's experience. In W. Blum, P. Galbraith, H. W. Henn, and M. A. Niss (Eds.), *Modelling and applications in mathematics education: The 14th ICMI study* (pp. 89-98). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Groth, R. E. (2006). Engaging Students in Authentic Data Analysis. In: *Thinking and Reasoning with Data and Chance* (Eds, G F Burrill and P C Elliott). National Council of Teachers of Mathematics Yearbook.

Hancock, C., Kaput, J. J. & Goldsmith, L. T. (1992). Authentic inquiry with data: Critical barriers to classroom implementation. *Educational Psychologist, 27*(3), 337–364.

Hiebert, J. & Lefevre, P. (1986). Conceptual and procedural knowledge in mathematics: An introductory analysis. In J. Hiebert (Ed.), Conceptual and procedural knowledge: The case of mathematics (pp. 1–27). Hillsdale, NJ: Lawrence Erlbaum Associates.

Jones, G., Langrall, C., Thornton, C. & Mogill, T. (1997). A framework for assessing and nurturing young children's thinking in probability. *Educational studies in Mathematics, 32*, 101–125.

Jones, G. A., Langrall, C. W., Thornton, C. A. & Mogill, A.T. (1999). Students' probabilistic thinking in instruction. *Journal for Research in Mathematics Education, 30* (5), 487-519.

Jones, G. A., Langrall, C. W., Thornton, C. A., Mooney, E. S., Wares, A., Perry, B., et al. (2001). Using students statistical thinking to inform instruction. *Journal of Mathematical Behavior, 20,* 109–144.

Konold, C. & Higgins, T. (2003). *Reasoning about data.* In J. Kilpatrick, G. Martin, & D. Schifter (Eds.), A research companion to NCTM's standards. Reston, VA: NCTM.

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

Lampert, M. & Cobb, P. (2003). Communication and language. In J.Kilpatrick, W.G. Martin, and D. Schifter (Eds.), *A research companion to the principles and standards for school mathematics* (pp. 247-249). Reston, VA: NCTM.

Leavy, A.M. (2010). The Challenge of Preparing Preservice Teachers to Teach Informal Inferential Reasoning. *Statistics Education Research Journal, 9*(1), 46-67.

Leavy, A, & Hourigan, M. (2015). Motivating Inquiry in Statistics and Probability in the Primary Classroom. *Teaching Statistics, 37*(2), 41-47. DOI: 10.1111/test.12062

Leavy, A. & Hourigan, M. (2016). Crime Scenes and Mystery Players! Using interesting contexts and driving questions to support the development of statistical literacy. *Teaching Statistics, 38*(1), 29-35.

Leavy, A. & Hourigan, M. (2018). The role of perceptual similarity, data context and task context when selecting attributes: Examination of considerations made by 5-6 year olds in data modelling environments. *Educational Studies in Mathematics, 97*(2), 163-183. DOI: 10.1007/s10649-017-9791-2

Leavy, A.M. & Middleton, J.A. (2011). Middle grades students' understanding of typicality. *Journal of Mathematical Behaviour, 30*(3), 235-254.

Lehrer, R. & Romberg, T. (1996). Exploring children's data modeling. *Cognition and Instruction, 14*(1), 69–108.

Lehrer, R. & Schauble, L. (2002). Inventing data structures for representational purposes: Elementary grade students' classification models. *Mathematical Thinking and Learning, 2,* 51–74.

Makar, K. (2018). Theorising links between context and structure to introduce powerful statistical ideas in the early years. In Leavy, A.M. Meletiou-Mavrotheris, M & Paparistodemou, E. (Editors). *Statistics in Early Childhood and Primary Education: Supporting early statistical and probabilistic thinking*. Springer.

Makar, K. & Rubin, A. (2007, July). *Beyond the bar graph: Primary teachers' uses of informal inference to teach statistical inquiry.* Paper presented at Fifth International Research Forum on Statistical Reasoning, Thinking, and Literacy (SRTL-5), University of Warwick, UK.

Makar, K. & Rubin, A. (2009). A framework for thinking about informal statistical inference. *Statistics Education Research Journal, 8*(1), 82-105. [Online: http://www.stat.auckland.ac.nz/~iase/serj/SERJ8(1)_Makar_Rubin.pdf]

Marriott, J., Davies, N. & Gibson, L. (2009) Teaching, Learning and Assessing Statistical Problem Solving. *Journal of Statistics Education, 17(*1) DOI: 10.1080/10691898.2009.11889503

Mokros, J. & Russell, S. J. (1995). Children's concept of average and representativeness. *Journal for Research in Mathematics Education, 26*(1), 20–39.

Moss, L.J. & Grover, B.W. (2007). Not just for computation: Basic calculators can advance the process standards. *Mathematics Teaching in the Middle School, 12*(5), 266-271.

Petrosino, A. J., Lehrer, R. & Schauble, L. (2003). Structuring error and experimental variation as distribution in the fourth grade. *Mathematical Thinking and Learning*, *5*(2&3), 131–156.

Pfannkuch, M. (2006). Informal inferential reasoning. In A. Rossman & B. Chance (Eds.), *Working cooperatively in statistics education: Proceedings of the Seventh International Conference on Teaching Statistics*, Salvador, Brazil. [CDROM]. Voorburg, The Netherlands [Online: http://www.stat.auckland.ac.nz/~iase/publications/17/6A2_PFAN.pdf]

Shaughnessy, J. M. (2007). Research on statistics learning and reasoning. In F. K. Lester Jr (Ed.), *Handbook of research on mathematics teaching and learning* (2nd ed., pp. 957–1009). Greenwich, CT: Information Age.

Watson, J. M. (2018). Variation and Expectation for Six-year-olds. In Leavy, A.M. Meletiou-Mavrotheris, M & Paparistodemou, E. (Editors). *Statistics in Early Childhood and Primary*

Education: Supporting early statistical and probabilistic thinking. Springer.

Watson, J. M. & Kelly, B. A. (2002). *Can grade 3 students learn about variation?* Paper presented at the sixth international conference on teaching statistics, Cape Town, South Africa.

Wild, C. & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review, 67*(3), 223–265.

Zieffler, A., Garfield, J., delMas, R. & Reading, C. (2008). A framework to support research on informal inferential reasoning. *Statistics Education Research Journal, 7*(2), 40-58. [Online: http://www.stat.auckland.ac.nz/~iase/serj/SERJ7(2)_Zieffler.pdf]

NCCA