REASONING ABOUT VARIATION: RETHINKING THEORETICAL FRAMEWORKS TO INFORM PRACTICE

Chris Reading[1] and Jackie Reid[2]
[1]SiMERR National Centre, University of New England, Australia
[2]School of Science and Technology, University of New England, Australia
creading@une.edu.au

There has been an increasing focus in recent years on theoretical frameworks to describe cognitive development of statistical concepts. There is now a need to encourage the use of these frameworks to inform practice in the teaching and learning of statistics. This paper focuses on frameworks that describe the levels of cognitive development of the concept of variation. Recent research proposing theoretical frameworks on, or referring to, reasoning about variation are synthesised. Discussion follows on the use of theoretical frameworks to inform the teaching and learning cycle for statistics courses, especially the design of curriculum, learning activities and assessment tasks.

INTRODUCTION

Variation is a broad construct with a critical role underpinning statistical thinking and during the past decade statistics education researchers have become more aware of the importance of variation. The use of structured models of cognitive development to develop coding schemes for analysing student responses was an early phenomenon in this focus on reasoning about variation (e.g., Watson, Kelly, Callingham & Shaughnessy, 2003). Detailed information about the cognitive structure of reasoning about variation is needed to better plan for teaching and learning in statistics courses. This paper first presents necessary components of the concept of variation and cognitive frameworks proposed to help articulate how reasoning about variation develops. Then, a variety of teaching and learning situations are described to demonstrate how these frameworks have informed practice. Finally, some implications for future teaching and learning practice are shared.

COMPONENTS OF THE CONCEPT OF VARIATION

The increasing interest amongst statistics education researchers in variation and reasoning about variation in the last decade has provided insights into the concept of variation itself. To better understand the concept it is necessary to review the concept's components. By synthesizing research presented at the Third International Research Forum on Statistical Reasoning, Thinking and Literacy (SRTL3), titled *Reasoning about Variation*, Garfield and Ben-Zvi (2005) were able to propose a set of building blocks necessary for a 'deep' understanding of variability (see components #1 to #7 in Table 1). These components were considered to be increasingly sophisticated ideas (Garfield & Ben-Zvi, 2008) that contribute to a deeper understanding of variation. Two additional important components of variation, #8 recognizing sources of variation (Reading & Shaughnessy, 2004) and #9 resolving expectations with observed variation (Watson, Callingham & Kelly, 2007), have also emerged.

Table 1. Components of Variation

#1 developing intuitive ideas of variability
#2 describing and representing variability
#3 using variability to make comparisons
#4 recognizing variability in special types of distributions
#5 identifying patterns of variability in fitting models
#6 using variability to predict random samples or outcomes
#7 considering variability as part of statistical thinking
#8 recognizing sources of variation
#9 resolving expectations with observed variation

In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8, July, 2010), Ljubljana, Slovenia*. Voorburg, The Netherlands: International Statistical Institute. www.stat.auckland.ac.nz/~iase/publications.php

For a more detailed view, the components can be further expanded. For example, Reid, Reading and Ellem (2008) described components that focus on dealing with explained and unexplained variation, which complement #1, #3, #5 and #8 in Table 1. Also, student conceptions of variation as evidenced in specific tasks can be used to provide more in-depth explanations of the components. For example, Shaughnessy (2006) summarized the work of a number of researchers to present the different conceptions of variability, when dealing with repeated sampling or repeated measures tasks, as: extremes or outliers; change over time; the whole range; the likely range; distance or difference from some fixed point; and sums of residuals.

These deconstructions of the components of the concept contribute to providing an in-depth view of the complexity of variation. However, it is also necessary to track the levels of cognition when dealing with the concept in order to assist in designing learning and assessment tasks that challenge the learner to further develop their reasoning.

COGNITIVE DEVELOPMENT OF STATISTICAL REASONING

Two cognitive frameworks applicable to more general statistical work but useful in informing cognitive levels of variation are now explained. These frameworks are based on the Structure of Observed Learning Outcome (SOLO) Taxonomy, which describes levels of cognitive development and is becoming popular in statistics education research. The SOLO Taxonomy has five modes of functioning, two (Ikonic and Concrete-Symbolic) of which are particularly relevant to developing cognition of specific concepts (Reading, 2004). Within each mode there are three relevant levels: unistructural (U) focusing on one element, multistructural (M) focusing on several unrelated elements, and relational (R) focusing on several elements with inter-relationships. The SOLO Taxonomy is useful because descriptions of levels of cognition are developed from what students are actually doing.

The first framework, for middle school students' statistical thinking (Langrall & Mooney, 2002), was designed to describe the cognitive actions that students engaged in during four different data-handling processes. Various iterations of the framework were proposed and revised, after interviewing students, to provide descriptors at the *idiosyncratic*, *transitional*, *quantitative* and *analytical* levels. These levels are equivalent to the Ikonic mode and then the U, M and R levels in the Concrete-Symbolic mode respectively. The framework was proposed to assess the level of statistical thinking across tasks and variation is included as the spread of data. Although the researchers attempted to make the framework more accessible to teachers and suggested it be used for guiding curriculum and task development, the lack of publications citing Langrall & Mooney's (2002) work suggests that it has not been used as much as hoped.

The second framework for statistical reasoning (Garfield, 2002) was designed to describe the cognitive actions that students engage in when working with statistical concepts. This framework was developed from the same frameworks that Langrall & Mooney (2002) extended, but had five levels: *idiosyncratic*, *verbal*, *transitional*, *procedural* and *integrated*. This provided a more detailed map of cognitive levels by introducing the *verbal* level for those who can articulate but not apply concepts and *procedural* level for those who cannot fully integrate the dimensions of the concept. Garfield (2002) provided specific explanations of how the framework could be used to support student learning and has been cited many times, thus informing both teaching and research.

These studies demonstrate how cognitive levels are being described in statistical processes generally, which informs the description of cognitive levels in reasoning about variation.

COGNITIVE DEVELOPMENT OF REASONING ABOUT VARIATION

Focusing now on articulating the cognitive development of reasoning about variation, various frameworks are considered that deal with acknowledging and describing variation, using variation to support inferences, explaining the source of variation, and linking variation to other concepts.

Acknowledging and describing variation (see #1 and #2 in Table 1) are important first steps for students when undertaking statistical tasks. Reading (2004) used SOLO to develop a hierarchy of the cognitive levels of description of variation. Two U-M-R developmental cycles were observed. The first was based on describing qualitative features of variation, with a relational response linking a number of qualitatively expressed features of variation. The second cycle was based on describing quantitative features of variation, with a relational response linking the features. Another dimension was added to the Reading (2004) levels when Reading and Shaughnessy's (2004) description hierarchy identified cognitive levels based on linking features of the data, in particular the 'extremes' (i.e., the range or spread of the numbers involved) and the 'middle values' (i.e., what is happening to the numbers within the range). This is an important step in moving toward the traditional view of variation as deviations from a central anchor. In the early development of reasoning about variation, students must be carefully guided in the move from the more qualitative (intuitive) description of variation to the more quantitative (numeric) description.

Using variation to support inferences, e.g., comparing and predicting (see #3 and #6 respectively in Table 1) is an important aspect of reasoning about variation. This use will be most effective when the concept has already been understood and linked to other concepts. Reid and Reading (2008) developed a hierarchy of consideration of variation that was based on tertiary students' responses to a variety of tasks. A key discriminator between the levels was the identification of within-group and/or between-group variation: the *weak* level responses recognized either within- or between-group variation, the *developing* recognized both within- and between-group variation and the *strong* linked the two types of variation. Other discriminators included in the descriptions of each level resemble the components acknowledging, describing, recognizing sources and resolving expectations (see #1, #2, #8 and #9 in Table 1). The discriminator *using variation to support inference* was not included as a single component but could be considered as incorporating a number of components, i.e., #2, #3, #5 and #6 in Table 1. More research is needed to give a better description of the *strong* level, as few responses were identified that truly linked the within- and between-group variation.

A broader scope of the first six components of variation was addressed by the four cognitive levels of the Watson et al. (2003) hierarchy. These levels were developed from school student responses to a variety of multiple choice and short answer questions. The first two levels, *prerequisite* and *partial* recognition, articulate to #1 and #2 in Table 1, while the third level, *application*, incorporates #3, #4, #5 and #6. The description of the highest cognitive level, *critical reasoning*, involves complex justifications with a consolidation of concepts. The description of this level provides a dimension beyond the components provided in Table 1.

Explaining the source of variation (see #8 in Table 1) comes naturally to students early on in statistical experiences. Reading and Shaughnessy (2004) developed a hierarchy, parallel to their description of variation hierarchy, to describe cognitive levels in the explanation of causes (sources) of variation. At the lowest cognitive level, extraneous sources of variation were offered. In more cognitively sophisticated responses, relevant explanations for variation were proposed with increasing cognitive sophistication indicated by the discussion of frequencies, proportions and then likelihoods, respectively. The levels in the Reading and Shaughnessy (2004) hierarchies have a strong alignment with SOLO, as the linking of ideas appears at the higher cognitive levels.

Linking variation to other concepts, e.g., expectation (see #9 in Table 1) becomes possible once the concept of variation is consolidated. Watson, Callingham and Kelly (2007) proposed a developmental pathway for understanding both variation and expectation. A SOLO-based model was developed from previous coding descriptions to code responses into one of six levels: *idiosyncratic*, *informal*, *inconsistent*, *consistent*, *distributional*, and *comparative distributional*. At the first three levels variation and expectation develop separately and it is not until the *consistent* level that there are indications of an appreciation of the link between the two concepts. Watson et al. (2007) cited a number of studies where there was "growing appreciation for the interaction between the expectation of the theoretical model and innate appreciation that variation will occur from it" (p. 86).

USING COGNITIVE FRAMEWORKS TO INFORM PRACTICE

Despite an increasing interest in developing cognitive frameworks to explain reasoning about variation, there is little evidence that these frameworks are being used in teaching and learning. Following are some examples of published uses that are evident in the design of curriculum, learning activities and assessment, and in the exploration of students' reasoning about other statistical concepts.

First, the informed design of curriculum is critical to the provision of supported learning. Cognitive development frameworks are useful to inform the sequencing and presentation of content. Changes to the school curriculum to facilitate the further development of statistical literacy were proposed by Watson and Kelly (2008) based on theoretical underpinnings from their longitudinal cross-grades application of the Watson et al. (2003) framework to describe school students' understanding of a number of fundamental statistical concepts including variation. They found that in Year 9, 50% of sampled students could not define variation. Curriculum documents provided for teachers are often not explicit in their definition of concepts such as variation, and so the foundation for student understanding of such concepts is likely to vary from classroom to classroom. The school curriculum needs to be more explicit in addressing these concepts, and teachers must continually reinforce them across the years, ensuring that students develop a conceptually more sophisticated understanding. Modifications to a model for developing a curriculum for an undergraduate statistics course with variation as a core thread were proposed by Reading & Reid (2005) based on their hierarchy of levels of consideration of variation. Fundamental elements of variation were mapped across four key curriculum themes, identifying possible gaps in the curriculum. This helped to identify misconceptions and gaps in students' reasoning about variation.

Second, the design of learning activities should be informed by research to best support student learning. Cognitive frameworks provide a valuable source of information. Garfield and Ben-Zvi (2008) used the first seven components in Table 1 as a basis for the development of a series of lessons that covered topics, such as comparing groups and sampling distributions, that relied on reasoning about variability. The first lesson, based on component #1 in Table 1, helps students to reason informally about variability. The second lesson, based on #2, fosters a deeper understanding of variability by encouraging students to informally estimate deviations from the mean, and to reason about the connections between a measure of centre and spread. Subsequent lessons revisit the concept of variability guided by the other components of variation in Table1.

Hypothetical learning trajectories, predictions of students' thinking and understanding as they progress through activities to reach a learning goal, provide a useful structure on which to base the design of learning activities. The development of such trajectories should be informed by the use of cognitive frameworks. For example, Garfield, delMas and Chance (2007) linked the components in Table 1 to the steps in a proposed hypothetical learning trajectory for developing students' reasoning about variability. This trajectory begins with the basic understanding that data vary (#1) and proceeds through a variety of steps including investigating why measurements vary and processes that lead to variation (#8) to working with particular measures of variability for different types of distributions (#4).

As well as being used to inform the design of suitably sequenced learning activities, cognitive frameworks can be used to inform the modification of learning activities. Reading and Reid (2005) used their hierarchy of levels of consideration of variation to grade minute papers and thus evaluate the effectiveness of the integration of the concept of variation into the learning activities. For one topic, sampling distributions, the generally low level of consideration of variation exhibited in student responses to a minute paper suggested that the learning activities had not adequately conveyed the concept to the learners. Consequently, the learning activities were adjusted to include cognitive conflict situations and interactive applets to better develop the necessary cognitive development of the concept of variation. In fact, Reid and Reading (2008) extended their hierarchy so that it was applicable to a greater range of in-class tasks, not just minute papers. This allows timely identification of gaps in students' reasoning about variation, and thereby aids in the development and modification of appropriate learning activities.

Third, the design of assessment tasks should be informed by the use of cognitive frameworks. Garfield and Ben-Zvi (2005) argued the need for assessment tasks that move away from the traditional focus on definitions, calculations and simple interpretations of spread. They used the first seven components in Table 1 as a framework to provide suggestions for developing assessment tasks to evaluate student understanding of the many aspects of variation. These ranged from tasks of low cognitive difficulty, such as asking students to describe the shape of a distribution when given raw data, i.e., component #1, through to more challenging tasks that allow students to carry out a statistical investigation to reveal how they consider variability, i.e., component #7. They recommended tasks embedded in realistic settings to engage students in complex reasoning situations, while providing a more detailed picture of students' reasoning about the various components of variation.

Cognitive frameworks are also useful for judging whether an assessment task is eliciting a cognitively sound representation of reasoning. Use of the hierarchy of consideration of variation to code students' responses to formative assessment tasks (Reading & Reid, 2005) helped identify those tasks that were better at eliciting useful information about consideration of variation. As a consequence, in later iterations of the course, some assessment tasks were modified to ensure that student responses truly reflected the level of their understanding. For example, one task required data to be presented in a less summarised form than previously, allowing students to better explore various components of variation. In another task the wording was changed to allow a more open style of question, thereby supporting the students in providing a cognitively deeper response.

Finally, the exploration of students' reasoning about other statistical concepts can be informed using cognitive frameworks describing reasoning about variation. Reading & Reid (2006) highlighted the association between reasoning about variation and about distribution. They used the hierarchy that they had developed to identify levels of consideration of variation (Reading & Reid, 2005) as a lens to examine students' reasoning about distribution. By using the hierarchy to firstly code the level of consideration of variation exhibited in students' responses and then examine the level of students' reasoning about distribution, the links between the level of consideration of variation and the level of reasoning about distribution were highlighted. This analysis then informed the development of a hierarchy of reasoning about distribution.

IMPLICATIONS

Although the above uses of cognitive frameworks have informed a variety of teaching and learning situations, there is still untapped potential. These frameworks can inform teaching in four important areas. First, they have the potential to inform cognitively-connected curriculum. Those designing curriculum need to use cognitive development frameworks as a guide so that concepts are developed in appropriate order and at suitable rates. Those being asked to implement a pre-developed curriculum should always ask "what cognitive framework(s) have informed the design of this curriculum?", or even more importantly "does the current order of content allow the necessary cognitive development to occur?". For example, the properties of variation cannot be used to support inferences until the properties themselves are understood. Second, the frameworks have the potential to inform the initial design of sequences of learning activities and also the modification of learning activities to better suit student needs. Those developing learning activities should seek a relevant framework on which to structure a lesson sequence. For example, students should not be expected to engage in lessons requiring the description of variation before they have been provided with opportunities to explore the nature of variation. Third, the frameworks have the potential to inform the choice of relevant assessment tasks for cognition to be assessed. Those developing assessment tasks need to structure tasks with parts that require increasing cognitive sophistication when answered, and to use a wider variety of tasks. For example, initial parts of an assessment task could require students to recognise and/or describe variation, then later parts of the task should require students to work with variability, e.g., using variation to explain comparisons or to resolve anomalies between expectations and observations. Finally, the frameworks have the potential to assist educators in appreciating the linking of variation to other statistical concepts, in helping their students to nurture such linking and in encouraging the development of such linking in students' mental models. For example, students should be provided with opportunities to articulate the effect of changing variation on the shape of a distribution.

Researchers can assist this process by continuing to research relevant frameworks and their application, and by making cognitive frameworks that they develop more accessible to educators, including relevant detail and examples. In the case of reasoning about variation, there already exist a variety of frameworks to deconstruction the concept of variation but there are still areas that need expanding. For example, there is a need for more research into students' reasoning at the high-end of cognitive frameworks to inform post-introductory tertiary statistics courses.

CONCLUSION

What is evident from this examination of applications of frameworks to practice is that a sound understanding of the components of variation (as listed in Table 1) is crucial to the development of students' statistical literacy, is fundamental to the development of reasoning about other key statistical concepts such as distribution, and should underpin curriculum development for statistics courses. This application of cognitive frameworks to practice provides a well-based structure for encouraging increased cognition and should be encouraged in all aspects of teaching and learning.

REFERENCES

Garfield, J. B. (2002). The Challenge of Developing Statistical Reasoning. *Journal of Statistics Education, 10*(3). Online: www.amstat.org/publications/jse/v10n3/garfield.html.

Garfield, J., & Ben-Zvi, D. (2005). A framework for teaching and assessing reasoning about variability. *Statistics Education Research Journal, 4*(1), 92-99.

Garfield, J. B., & Ben-Zvi, D. (2008). *Developing Students' Statistical Reasoning: Connecting Research and Teaching Practice.* The Netherlands: Springer.

Garfield, J., delMas, R. C., & Chance, B. (2007). Using students' informal notions of variability to develop an understanding of formal measures of variability. In M. Lovett & P. Shah (Eds.), *Thinking with Data* (pp. 117-148). Mahwah, NJ: Lawrence Erlbaum Associates.

Langrall, C. W., & Mooney, E. S. (2002). The development of a framework characterizing middle school students' statistical thinking In A. Rossman & B. Chance (Eds.), *Working Cooperatively in Statistics Education. Proceedings of the Seventh International Conference on Teaching Statistics,* [CDROM]. Salvador, Brazil: International Statistical Institute and International Association for Statistical Education.

Reading, C. (2004). Student description of variation while working with weather data. *Statistics Education Research Journal, 3*(2), 84-105.

Reading, C., & Reid, J. (2005). Consideration of variation: A model for curriculum development. In G. Burrill & M. Camden (Eds.), *Curricular Development in Statistics Education: International Association for Statistical Education 2004 Roundtable* (pp. 36-53). Voorburg, The Netherlands: International Statistical Institute.

Reading, C., & Reid, J. (2006). An emerging hierarchy of reasoning about distribution: From a variation perspective. *Statistics Education Research Journal, 5*(2), 46-68.

Reading, C., & Shaughnessy, J. M. (2004). Reasoning about variation. In D. Ben-Zvi & J. Garfield (Eds.), *The Challenge of Developing Statistical Literacy, Reasoning and Thinking* (pp. 201-226). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Reid, J., & Reading, C. (2008). Measuring the development of students' consideration of variation. *Statistics Education Research Journal, 7*(1), 40-59.

Reid, J., Reading, C., & Ellem, B. (2008). Developing assessment items to measure tertiary students' reasoning about explained and unexplained variability. *Proceedings of the 2nd Annual Postgraduate Research Conference: Bridging the Gap between ideas and Doing Research*. Armidale, Australia: University of New England.

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

Watson, J. M., Callingham, R. A., & Kelly, B. A. (2007). Students' appreciation of expectation and variation as a foundation for statistical understanding. *Mathematical Thinking and Learning, 9*(2), 83-130.

Watson, J. M., & Kelly, B. A. (2008). Sample, random and variation: The vocabulary of statistical literacy. *International Journal of Science and Mathematics Education, 6,* 741-767.

Watson, J. M, Kelly, B. A., Callingham, R. A., & Shaughnessy, J. M. (2003). The measurement of school students' understanding of statistical variation. *International Journal of Mathematical Education in Science and Technology, 34*(1), 1-29.