A FRAMEWORK FOR TEACHING AND ASSESSING REASONING ABOUT VARIABILITY

JOAN GARFIELD
University of Minnesota
jbg@umn.edu

DANI BEN-ZVI
University of Haifa
dbenzvi@univ.haifa.ac.il

SUMMARY

This article is a discussion of and reaction to two collections of papers on research on Reasoning about Variation: Five papers appeared in November 2004 in a Special Issue 3(2) of the Statistics Education Research Journal (by Hammerman and Rubin, Ben-Zvi, Bakker, Reading, and Gould), and three papers appear in a Special Section on the same topic in the present issue (by Makar and Confrey, delMas and Liu, and Pfannkuch). These papers show that understanding of variability is much more complex and difficult to achieve than prior literature has led us to believe. Based on these papers and other pertinent literature, the present paper, written by the Guest Editors, outlines seven components that are part of a comprehensive epistemological model of the ideas that comprise a deep understanding of variability: Developing intuitive ideas of variability, describing and representing variability, using variability to make comparisons, recognizing variability in special types of distributions, identifying patterns of variability in fitting models, using variability to predict random samples or outcomes, and considering variability as part of statistical thinking. With regard to each component, possible instructional goals as well as types of assessment tasks that can be used in research and teaching contexts are illustrated. The conceptual model presented can inform the design and alignment of teaching and assessment, as well as help in planning research and in organizing results from prior and future research on reasoning about variability.

Keywords: Variability; Variation; Learning and teaching statistics; Assessment; Statistics curriculum; Statistics education; Reasoning and thinking; Deep understanding

1. INTRODUCTION

Variability is at the heart of statistics and is the fundamental component of statistical thinking (Pfannkuch, 1997; Pfannkuch & Wild, 2004; Shaughnessy, 1997). Variability is what makes decisions in the face of uncertainty so difficult. Variability is what makes statistics so challenging and interesting and allows us to interpret, model and make predictions from data (Gould, 2004). Therefore, variability should be repeatedly integrated, revisited and highlighted in statistics curriculum and instruction (e.g., Moore, 1990). Moore (1992, p. 426) extends this notion of the centrality of variation by stating that “pupils in the future will bring away from their schooling a structure of thought that whispers ‘variation matters’.” Variability should be centrally emphasized from the earliest grades (in formal and informal activities and discussions) through high school and the introductory college course. Suggestions for how this should be done and evaluated are the focus of this paper.

Statistics Education Research Journal, 4(1), 92-99, http://www.stat.auckland.ac.nz/serj

This article discusses and extends ideas appearing in two collections of papers on research on Reasoning about Variation. Five papers appeared in the November 2004 special issue of the *Statistics Education Research Journal* (Hammerman & Rubin, Ben-Zvi, Bakker, Reading, Gould), and three papers appear in the Special Section on the same topic in the present issue (Makar & Confrey, delMas & Liu, Pfannkuch). These two collections of papers overall provide us with a strong case that understanding of variability is much more complex and difficult than prior literature suggests. For example, even college students or teachers, examined in some of the studies, demonstrate diverse intuitions, misconceptions, and incomplete or shallow understanding (Hammerman & Rubin, 2004; delMas & Liu, 2005; Makar & Confrey, 2005). Shaughnessy, Watson, Moritz, and Reading (1999) have similarly found a lack of clear growth in students' conceptions of variability for a particular task. A full and complete understanding of variability means developing a cognitive model that includes several components and their connections, and using this model to reason about variability in different contexts.

The two collections of papers present general ideas regarding ways for teaching and assessment related to knowledge of variability, but have not examined in much detail bridges to practice, in the form of specific instructional goals and assessment methods. Hence, building from the foundations of the collected set of issues these papers discussed, as well as from our own work, in this paper we suggest an epistemological model and associated actions and behaviors that would be demonstrated when students reason about variability as they encounter data in solving statistical problems. In the following sections, we offer a preliminary but extensive set of ideas both regarding instructional topics as well as regarding assessment tools associated with the model of reasoning about variability, and close with a discussion of implications of the suggested model for practice and future research.

2. DEVELOPING DEEP UNDERSTANDING OF VARIABILITY

In this section, we outline the components of an epistemological model that consists of a set of ideas loosely grouped into seven overlapping categories or areas. These areas can be viewed as building blocks for constructing "deep understanding" of the complex concept of variability. A deep understanding means that instead of being able to recite only fragmented pieces of information, students develop relatively systematic, integrated or holistic understandings. Mastery is demonstrated by their success in producing new knowledge by discovering relationships, solving new problems, constructing explanations, and drawing conclusions. Deep understanding is robust, reflective, i.e., includes awareness of own boundaries, of limits of applicability, and also operable, i.e., can be used to solve real problems. It is characterized by its breadth, by knowing the concept components and their connections, and most importantly, by knowing what is important and what is less important or irrelevant. Deep understanding is similar to "conceptual understanding" and to Skemp's "relational understanding", as opposed to "procedural understanding" (Skemp, 1976). Conceptual understanding similarly implies knowledge of the idea, and how it relates to already acquired ideas. It also requires an understanding of the contexts within which the idea is applicable, as well as its limitations. It enables a person to apply and adapt an idea flexibly to new situations rather than just following learned procedures in familiar situations.

Below we present key ideas in each of the seven areas of knowledge of variability:

1. Developing intuitive ideas of variability

- Recognizing that variability is everywhere (the omnipresence of variability; Moore, 1990, 1997). Individuals vary on many characteristics, and repeated measurements on the same characteristic are variable. Both qualitative and quantitative variables reveal variability of data.
- Some things vary just a little, some vary a lot.
- We can try to understand why things vary: By thinking about and examining the variables we can try to explain the different reasons and sources for variability.
- Variability is a general or global characteristic of a data set. It involves considering data as an entity, rather than as individual points or as a combination of center and extreme values.

2. Describing and representing variability

- Graphs of data show how things vary and may reveal patterns to help us focus on global features of distributions and identify the signal in the noise.
- Different graphs may reveal different aspects of the variability in a data set so it is important to study more than a single graph of a data set.
- We can use one number to represent a global feature (such as variability) of the distribution.
- Different numerical summaries tell us different things about the spread of a data set. For example, the Range tells us the overall spread from highest to lowest value, while the Standard Deviation (SD) tells us the typical spread from the mean. The Interquartile Range (IQR) tells us the spread of the middle half of a distribution.
- While the IQR and SD tell us about variability of data, they are most useful for interpreting variability when we also know the related measure of center (mean for SD, median for IQR) as well as the general shape of the distribution.
- Measures of variability and center (as long as we consider them together) are more or less informative for different types of distribution. For example, the mean and SD tell us useful information about symmetric distributions, in particular, the normal distribution. For skewed distributions, the median and IQR are more useful summaries.

3. Using variability to make comparisons

- When making comparisons of two or more data sets, examining their graphs on the same scale allows us to compare the variability and speculate about why there are differences in the data sets.
- It is helpful to use global summaries of spread and center when comparing groups, rather then comparing individual data points or 'slices' of the graphs.
- It is important to examine both the variability within a group (looking at how the data vary within one or more data sets) and the variability between groups (the variability of measures used to summarize and compare data sets), and distinguish between these two types of variability.

4. Recognizing variability in special types of distributions

- In a normal distribution, the mean and SD provide useful and specific information about variability. For example, if we know the SD in addition to the mean, we can determine the percentage of data within one, two, and three Standard Deviations of the mean. We can also use the mean and SD to estimate the high and low values of the distribution and the Range.
- There is variability in a bivariate data distribution, and we need to consider the variability of both variables as well as the variability for y values given individual values of x.
- The variability of a bivariate data set (covariation) may reveal a relationship between the variables and whether we might predict values of one variable (y) for values of the other (x).

5. Identifying patterns of variability in fitting models

- There is variability involved in fitting models and judging the fit of models (e.g., fitting the normal curve to a distribution of data, or fitting a straight line to a scatterplot of bivariate data).
- The variability of the deviations from the model (residuals) can tell us about the how well the model fits the data.
- Data may sometimes be reorganized and transformed to better reveal patterns or fit a model.

6. Using variability to predict random samples or outcomes

- Samples vary in some predictable ways, based on sample size and the population from which they are drawn and how they are drawn. If we have random samples the variability can be more readily explained and described.
- Larger samples have more variability than smaller samples, when randomly drawn from the same population. However, sample statistics from the larger samples vary less than statistics from smaller samples.
- There is variability in outcomes of chance events. We can predict and describe the variability for random variables.
- In some situations we can link the variability in samples to variability in outcomes, making predictions or statistical inferences.

7. Considering variability as part of statistical thinking

- In statistical investigations, we always need to begin with examining and discussing the variability of data.
- Data production is designed with variation in mind. Aware of sources of uncontrolled variation, we avoid self-selected samples, insist on comparison in experimental studies, and introduce planned variation into data production by use of randomization (Moore, 1990).
- In statistical analysis we try to explain variation by seeking the systematic effects behind the random variability of individuals and measurements (Moore, 1990).
- The ideas listed above are all part of statistical thinking, and come into play when exploring data and solving statistical problems (Wild & Pfannkuch, 1999).

While the ideas described above may seem like a list for a curriculum or syllabus of introductory statistics courses, we point out that in each area, the idea of variability is to be highlighted, discussed and emphasized. This list of increasingly sophisticated ideas offers: (1) the ways in which this body of knowledge can be structured so that it can be most readily grasped by the learner, (2) an effective sequences in which to present material related to variability, (3) a plan for revisiting variability as students progress through the statistics curriculum, and (4) a scaffold on which to build new levels of deep understanding of variability. Supported by theories of constructivist learning (e.g., Bruner, 1973; Cowan 1995), our belief is that progress in students' construction of meanings is not linear but rather complex and is better captured by the image of spiral progression. Therefore, ideas related to variability must be constantly revisited along the statistics curriculum from different points of view, context and levels of abstraction, to create a complex web of interconnections among them. An important part of building this complex knowledge is assessing where students are in developing their deep understanding and reasoning about variability, which is discussed in the next section.

3. ASSESSING STUDENTS' REASONING ABOUT VARIABILITY

Most traditional assessments regarding knowledge of variability focus on knowledge of definitions, calculations, and simple interpretations of standard measures of dispersion (e.g., computing and interpreting the Range, SD, and IQR). On the other hand, alternative assessments focus on intuitions, conceptual understanding, and reasoning about the connections between variability and other concepts such as center and shape. Most innovative assessment tasks in this area have been developed by researchers, as illustrated in the papers in these special issues. However, even such studies lack detailed discussions of how to assess students' knowledge and reasoning about variability. The ways students' knowledge has been revealed or evaluated in the two collections of studies on reasoning about variability is primarily through extended data investigation activities or tasks, and through open-ended survey questions. While these are found beneficial for research purposes, they are not as useful as assessment tasks to provide practical evaluation feedback to teachers and students in real classroom situations, in which time and teacher resources are limited.

It is essential to coordinate the desired learning outcomes in either research or classroom work with ways to assess these aspects of conceptual understanding (Wiggins & McTighe, 1999). Acknowledging that traditional methods of assessing understanding of variability have focused mostly on definitions, calculations, and simple interpretations of measures of spread, we reflected on the tasks used in the research articles reviewed and on their implications, as well as examined prior literature (e.g., Ben-Zvi & Garfield, 2004). This provided us with a basis for suggesting a broader array of assessment items to be used for evaluating students' reasoning about the complex concept of variability. In designing assessments that are aligned to instructional goals, we see a need to develop assessment items that involve a real or realistic context in order to motivate students and encourage them to use their informal knowledge of variability. Below we provide suggestions for developing assessment items, using real or realistic data, to evaluate student understanding of the many aspects of understanding variability. This list is organized in terms of the seven areas listed earlier in which students are expected to develop some knowledge or intuitions related to variability.

1. Assessment - Developing intuitive ideas of variability

- Items that provide descriptions of variables or raw data sets (e.g., the ages of children in a grade school, or the height of these children) and asking students to describe variability or shape of distribution.
- Items that ask students to make predictions about data sets that are not provided (e.g., if the students in this class were given a very easy test, what would you predict for the expected graph and expected variability of the test scores?).
- Given a context, students are asked to think of ways to decrease the variability of a variable (e.g., measurements of one students' jump).
- Items that ask students to compare two or more graphs and reason about which one would have larger or smaller measures of variability (e.g., Range or Standard Deviation).

2. Assessment - Describing and representing variability

- Items that provide a graph and summary measures, and ask students to interpret it and write a description of the variability for each variable.
- Items that ask students to choose appropriate measures of variability for particular distributions (e.g., IQR for skewed distribution) and select measure of center that are appropriate (e.g., median with IQR, mean with SD).
- Items that provide a data set with an outlier that ask students to analyze the effect of different measures of spread if the outlier is removed. Or, given a data set without an outlier, asking students what effect adding an outlier will have on measures of variability.
- Items that ask students to draw graphs of distributions for data sets with given center and spread.

3. Assessment - Using variability to make comparisons

- Items that present two or more graphs and ask students to make a comparison either to see if an intervention has made a difference or to see if intact groups differ. For example, asking students to compare two graphs to determine which one of two medicines is more effective in treating a disease, or whether there is a difference in length of first names for boys and girls in a class.
- Items that ask students which graph shows less (or more) variability, where they have to coordinate shape, center, and different measures of spread.

4. Assessment - Recognizing variability in special types of distributions

- Items that provide the mean and standard deviation for a data set that has a normal distribution and students are asked to use these to draw graphs showing the spread of the data.

- Items that provide a scatterplot for a specific bivariate data set and students have to consider if values are outliers for either the x or y variables or for both.
- Items that provide graphs of bivariate data sets where students are asked to determine if the variability in one variable (y) can be explained by the variability in the other variable (x).

5. Assessment - Identifying patterns of variability in fitting models

- Items that ask students to determine if a set of data appear normal, or if a bivariate plot suggests a linear relationship, based on scatter from a fitted line.

6. Assessment - Using variability to predict random samples or outcomes

- Items that provide students choices of sample statistics (e.g., proportions) from a specified population (e.g., colored candies) for a given sample size and ask which sequence of statistics is most plausible.
- Items that ask students to predict one or more samples of data from a given population.
- Items that ask students which outcome is most likely as a result of a random experiment when all outcomes are equally likely (e.g., different sequences of colors of candies)
- Items that ask students to make conjectures about a sample statistic given the variability of possible sample means.

7. Assessment - Considering variability as part of statistical thinking

- Items that give students a problem to investigate along with a data set, that requires them to graph, describe, and explain the variability in solving the problem.
- Items that allow students to carry out the steps of a statistical investigation, revealing if and how the students consider the variability of the data.

We suggest that teachers and researchers develop items along the lines suggested in the above list, embedding the tasks in contexts that provide real or realistic situations to engage students in reasoning about variability and to reveal their cognitive understanding of this concept. Such tasks will enhance students' ability to negotiate the complex and multifaceted meanings of variability suggested above. The results of these assessments should provide detailed pictures of what students understand or partially understand and how they reason about the many aspects of variability as they learn statistics and carry out exploratory data investigations. In the summary comments below we focus mainly on instructional implications of our deliberations.

4. IMPLICATIONS

We have enjoyed reading and reflecting on the rich collection of papers in the two special issues. As guest editors of these special issues we chose to extend the discussion to how we might build on the implications of the research papers in terms of teaching, curriculum, and assessing reasoning about variability. We hope that teachers of statistics at all levels will build on the results of these studies, finding and using good and rich data sets and contexts that motivate students to describe interpret, and reason about variability.

We proposed in this paper a model of statistical ideas that are part of a deep and multifaceted understanding and reasoning about variability. This epistemological model suggests a sequence of seven increasingly sophisticated areas of variability knowledge. This model can serve to guide and organize results from prior and future research. It can help educators and curriculum developers to plan and design instruction and assessment that aim at achieving deep understanding of variability, in the spirit of Perkins' (1993) "teaching and learning for understanding." We believe it is important to spend ample time on each area, with many examples in rich contexts. Variability should be centrally emphasized throughout the statistics curriculum, from the earliest grades (in formal and informal activities and discussions) through high school and the introductory college course. The areas of the model should be revisited often, making explicit the interconnections among them as well as the connections with other statistical “big ideas” (Garfield & Ben-Zvi, 2004).

Challenging students to reason about variability will require data sets in which variability is revealed in interesting ways, and lead students to make and test conjectures. We encourage teachers to conduct discussions on sources of variability and what leads to the variability in a data set rather than merely collecting data to analyze in graphs and summary measures, or presenting data to be summarized in this way. Activities and discussions that reveal the different aspects and uses of variability can begin in the early grade levels and should be repeatedly included as students progress in their study of statistics. In this way, intuitive, informal notions of variability can gradually be used to construct more formal and complex understanding of variability.

As pointed out in the discussion by Pfannkuch (2005, this issue) and by others (e.g., Garfield & Burrill, 1997), technological tools should play an important role both in developing students’ reasoning about variability as well as in helping to make their reasoning visible to teachers and researchers. We recommend the careful use of appropriate technological tools to facilitate conceptual development and reasoning, and to assist in assessing and evaluating students’ learning (e.g., Ben-Zvi, 2000). It seems clear from the studies in these special issues that reflection is an important component of these activities as well, and that students should be encouraged to communicate their learning and understanding in increasingly appropriate statistical language. We hope that assessments will be created and shared with teachers and researchers that carefully evaluate students’ knowledge and reasoning about variability so that we may measure students’ achievement of this important learning outcome. Finally, we emphasize the critical need for future research focused on the development of a deep understanding of variability and how such an understanding may be advanced over time and through the careful integration of curriculum, technology and assessment.

REFERENCES

Bakker, A. (2004). Reasoning about shape as a pattern in variability. *Statistics Education Research Journal, 3*(2), 64-83. [Online: www.stat.auckland.ac.nz/serj]

Ben-Zvi, D. (2000). Towards understanding the role of technological tools in statistical learning. *Mathematical Thinking and Learning, 2*(1&2), 127-155.

Ben-Zvi, D. (2004). Reasoning about variability in comparing distributions. *Statistics Education Research Journal, 3*(2), 42-63. [Online: www.stat.auckland.ac.nz/serj]

Ben-Zvi, D., & Garfield, J. B. (Eds.) (2004). *The challenge of developing statistical literacy, reasoning, and thinking*. Dordrecht, the Netherlands: Kluwer Academic Publishers.

Bruner, J. (1973). *Going beyond the information given*. New York: Norton.

Cowan, D. A. (1995). Rhythms of learning: Patterns that bridge individuals and organizations. *Journal of Management Inquiry, 4*(3), 222-246.

DelMas, R., & Liu, Y. (2005). Exploring students’ conceptions of the standard deviation. *Statistics Education Research Journal, 4*(1), 55-82. [Online: www.stat.auckland.ac.nz/serj]

Garfield, J., & Ben-Zvi, D. (2004). Research on statistical literacy, reasoning, and thinking: issues, challenges, and implications. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 397-409). Dordrecht, the Netherlands: Kluwer Academic Publishers.

Garfield, J., & Burrill, G. (Eds.) (1997). *Research on the role of technology in teaching and learning statistics*. Voorburg, The Netherlands: International Statistical Institute.

Gould, R. (2004). Variability: One statistician’s view. *Statistics Education Research Journal, 3*(2), 7-16. [Online: www.stat.auckland.ac.nz/serj]

Hammerman, J., & Rubin, A. (2004). Strategies for managing statistical complexity with new software tools. *Statistics Education Research Journal, 3*(2), 17-41. [Online: www.stat.auckland.ac.nz/serj]

Makar, K., & Confrey, J. (2005). "Variation-talk": Articulating meaning in statistics. *Statistics Education Research Journal, 4*(1), 27-54. [Online: www.stat.auckland.ac.nz/serj]

Moore, D. (1990). Uncertainty. In L. Steen (Ed.), *On the shoulders of giants: New approaches to numeracy* (pp. 95–137). Washington, DC: National Academy Press.

Moore, D. (1992). Teaching statistics as a respectable subject. In F. Gordon & S. Gordon (Eds.), *Statistics for the twenty-first century*. MAA Notes, no. 26 (pp. 14-25). Washington, DC: Mathematical Association of America.

Moore, D. (1997). New pedagogy and new content: The case of statistics. *International Statistical Review, 65*, 123-165.

Perkins, D. (1993). An apple for education: Teaching and learning for understanding. *American Educator*, *17*(3), 8, 28-35.

Pfannkuch, M. (1997). Statistical thinking: One statistician's perspective. In F. Biddulph & K. Carr (Eds.), *People in mathematics education* (Proceedings of the 20th annual conference of the Mathematics Education Research Group of Australasia) (pp. 406-413). Rotorua, New Zealand: MERGA.

Pfannkuch, M., & Wild, C. (2004). Towards an understanding of statistical thinking. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking*, (pp. 17-46). Dordrecht, the Netherlands: Kluwer Academic Publishers.

Pfannkuch, M. (2005). Thinking tools and variation. *Statistics Education Research Journal, 4*(1), 83-91. [Online: www.stat.auckland.ac.nz/serj]

Reading, C. (2004). Student description of variation while working with weather data. *Statistics Education Research Journal, 3*(2), 84-105. [Online: www.stat.auckland.ac.nz/serj]

Shaughnessy, J. M. (1997). Missed opportunities in research on the teaching and learning of data and chance. In F. Biddulph & K. Carr (Eds.), *People in mathematics education* (Proceedings of the 20th annual conference of the Mathematics Education Research Group of Australasia) (pp. 6-22). Rotorua, New Zealand: MERGA.

Shaughnessy, J. M., Watson, J., Moritz, J., & Reading, C. (1999, April). School mathematics students' acknowledgement of statistical variation. In C. Maher (Chair), *There's More to Life than Centers*. Presession Research Symposium, 77th Annual National Council of Teachers of Mathematics Conference, San Francisco, CA.

Skemp, R. R. (1976). Relational understanding and instrumental understanding. *Mathematics Teaching, 77*, 20-26.

Wiggins, G., & McTighe, J. (1999). *Understanding by design*. Alexandria, VA: Association for Supervision and Curriculum Development.

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

JOAN GARFIELD
University of Minnesota
315 Burton Hall
178 Pillsbury Drive, S.E.
Minneapolis, MN 55455, USA