CONCEPTIONS OF DISTRIBUTION HELD BY MIDDLE SCHOOL STUDENTS AND PRESERVICE TEACHERS

Daniel L. Canada
Eastern Washington University, United States of America
mathaction@gmail.com

This paper examines how preservice teachers and middle school students reason about distributions as they consider graphs of two data sets having identical means but different spreads. Results show that while both subject groups reasoned about the task using the aspects of average and variation, relatively more preservice teachers than middle school students combined both aspects to constitute an emerging form of distributional reasoning in their responses. Moreover, these emergent distributional reasoners were more likely to see the data sets as fundamentally different despite the identical means used in the task.

INTRODUCTION

The purpose of this paper is to report on research describing the conceptions of distribution held by preservice elementary teachers in response to a task involving a comparison of two data sets. The two data sets were presented in the form of stacked dot plots, and key aspects of statistical reasoning potentially germane to the task were a consideration of the center and variation of the two distributions. Of particular interest was the role that variation, or variability in data, played in the subjects' conceptions of distribution. This interest stems from the primacy that variation holds within the discipline of statistics (Cobb & Moore, 1997; Wild & Pfannkuch, 1999).

Therefore, this study addresses the following primary research question: What are the informal conceptions of distribution held by preservice teachers and middle school students as they compare two data sets? After describing some related research concerning the aspects of distributional reasoning used in an analytic framework, the methodology for the study will be explicated. Then, results related to the research question will be presented, followed by a discussion and implications for future research and teacher training programs.

RELATED LITERATURE

The key elements comprising the conceptual framework for looking at distributional reasoning in the context of the task used in this study were a consideration of aspects of centre (average) and variation, which implies taking an aggregate view of data as opposed to considering individual data elements (Konold & Higgins, 2002). Coordinating these aspects is what enables a richer picture of a distribution to emerge (Shaughnessy, Ciancetta, & Canada, 2004; Makar & Canada, 2005).

Shaughnessy, Ciancetta, Best, and Canada (2004) investigated how middle and secondary school students compared distributions using a task very similar to the task used in the current study reported by this paper. Given two data sets with identical means and medians, subjects showed how they reasoned about averages and variation, with higher levels of distributional reasoning attributed to those responses that conflated both components of centers and spread. The researchers found that subjects' conceptions included notions of "variability as extremes or possible outliers; variability as spread; variability in the heights of the columns in the stacked dot plots; variability in the shape of the dispersion around center; and to a lesser extent, variability as distance or difference from expectation" (p. 29).

Other researchers have focused on the distributional reasoning of secondary teachers. In reporting how subjects compared boxplots, Pfannkuch (2006) proposed a descriptive model that included elements of reasoning with both centers and spreads, among other elements related to making informal inferences. Makar and Confrey (2004) had their subjects' consider the question "How do you decide whether two groups are different?" (p. 353), finding that teachers could use informal language to convey an intuitive recognition of variation within a group as well as between groups. The researchers underscore both the importance and struggle in having their subjects "interpret the difference between these types of variation" (p. 371).

© ICMI/IASE 2008. In C. Batanero, G. Burrill, C. Reading & A. Rossman (Eds.), *Joint ICMI/IASE Study: Teaching Statistics in School Mathematics. Challenges for Teaching and Teacher Education. Proceedings of the ICMI Study 18 and 2008 IASE Round Table Conference.*

Leavy's research (2006) with preservice elementary teachers embraced similar themes to those presented in this paper. Specifically, the researcher "investigated the approaches used to compare distributions of data" (p. 92) via a pre-test, instructional interventions, and a post-test. Many of Leavy's subjects shifted from an exclusive focus on descriptive statistics (especially the mean) to an inclusion of graphical representations as they considered both variability and centers. For those participants who became attenuated to the manifestation of variability in graphical representations, "variation quickly became a central component of participants' understanding of distribution" (p. 106).

Thus getting people to attend to aspects of both average and variation when investigating distributional reasoning helped inform not only the task creation for this research but also the lens for analysis of the subjects' responses.

METHODOLOGY

The task chosen to look at the subjects' thinking about distribution when comparing data sets was called the Train Times task and was motivated by a similar task initially used in previous research (Shaughnessy Ciancetta, Best, & Canada, 2004) The task scenario describes two trains, the EastBound and WestBound, which run between the cities of Hillsboro and Gresham along parallel tracks. For 15 different days (and at different times of day), data are gathered for time the trip takes on each of the train. The task was deliberately constructed so that the EastBound and WestBound train times have the same means, yet different amounts of variation are apparent in the graphs. As a part of the task scenario, subjects were told that the Transportation Department was deliberating whether or not one train was more reliable than the other. Subjects were asked whether they agreed with the hypothetical argument that there was "no real difference between the two trains because the data have the same means" and to explain their reasoning. This methodology follows that of Watson (2002), where subjects are asked to react to a common line of reasoning. A similar technique has been used in other research on statistical thinking (e.g., Shaughnessy Ciancetta, Best, & Canada, 2004; Shaughnessy, Ciancetta, & Canada, 2004). Would subjects be persuaded by the hypothetical argument of "no real difference" in times because of the identical means? Would they argue on the strength of the different modes, which are often a visual attractor for statistical novices reasoning about data presented in stacked dot plots? Or would they attend to the variability in the data, and if so, how would they articulate their arguments?

The subjects who participated in this study consisted of fifty middle school students (24 seventh graders and 26 sixth graders) who came from public schools in northwestern United States and fifty-eight preservice teachers who were taking a ten-week course in mathematics at a university in the same region. The main commonality in the statistical preparation of both subject groups was that their prior learning experiences included the definitions of mean, median, mode, and range. All subjects were given the Train Times task as a written-response item for completion in class, prior to any formal instruction or attention being explicitly given to the concept of variability within a data set. The task was not given as part of a formal evaluation for the course but rather as a way of having the subjects show their initial informal thinking. The task was then discussed in class, and the discussions were videotaped so that further student comments could be recorded and transcribed.

The data, comprised of the written responses and transcriptions of the class discussions, was then coded according to the components of conceptual framework that related to the distributional aspects of centers and spread. Responses could be coded according to whether they included references to centers, to informal notions of variation, or to both. The rubric for coding was adapted from prior work involving other research (e.g., Shaughnessy Ciancetta, Best, & Canada, 2004; Makar & Canada, 2005; Canada & Ciancetta, 2007).

RESULTS

Almost 35% (20/58) of the preservice teachers initially agreed with the hypothetical argument that there was "no real difference between the two trains because the data have the same means". Slightly less–32% (16/50)–of the middle school students agreed. While it might be expected that subjects who were predisposed to think of the mean as the sole or primary summary statistic for a set of data might support the hypothetical argument, a careful analysis of the responses showed different degrees to which subjects relied on centers and variation in their explanations. Thus, in addressing the primary research question ("What are the informal conceptions of distribution held by preservice teachers and middle school students as they compare two data sets?"), results are presented first according to responses that focused primarily on centers, then primarily on variation. Finally, examples of those responses that integrated centers with an informal notion of variation are presented as representing a form of emergent distributional reasoning.

Centers

Out of all subjects, 24.1% (14/58) of the preservice teachers and 42% (21/50) of the middle school students had responses that included what were coded as general references to centers. The exemplars that follow show the initials of the subject as well as a "-T" to designate a preservice teacher and a "-S" to designate a middle school student. The notation (A) or (D) shows whether they initially agreed or disagreed with the hypothetical argument of "no real differences" presented in the task scenario:

KW-S: (A) Because the averages are the same
DW-T: (A) Each train had the same average time.

Although it might be presumed that the subjects equate "average" with the mean, the general responses for center included no specific language. In contrast, the 39.7% (23/58) of the teachers and 34% (17/50) of the students who had specific references to centers were more explicit in what they were attending to:

LT-T: (D) I would probably go with the mode, because it is the most common answer
RB-T: (D) I would go by the median on this one.

Note how subjects LT and RB, in focusing on the mode or median, disagreed with the hypothetical argument of "no real difference." Indeed, although the means for the data sets in Train Times task were identical, the medians and modes differed, and some subjects with specific center responses picked up on these differences:

CM-T: (D) The median and modes are not the same; meaning results varied
LN-T: (D) The median and mode are different. Because the data is very different in its variation.

Here we see CM and LN tying their observations of differences in measures of centers to an informal notion of variation.

Variation

For the purposes of this research, variation need not be defined in formal terms such as a standard deviation (of which these subjects had no working knowledge) but tied to the informal descriptions such as those offered by Makar & Canada (2005). In particular, the essence of variation is that there are differences in the observations of the phenomena of interest. Of all subjects, 32.8% (19/58) of the teachers and only 10% (5/50) of the students gave a more general reference to variation:

TS-T: (D) The trains could all have different times sporadically
LR-T: (D) Because the time patterns are different between the two trains.

Note that in these examples, the theme of differences among data comes out in the natural language of the response. Clearly the sense of variation as differences is a naive and basic idea but one that is fundamental and a potentially useful springboard for a deeper investigation as to how to describe those differences.

A higher percentage of all subjects, 37.9% (22/58) of the teachers and 34% (17/50) of the students, had more specific references to variation. The main motivation for looking at specific characteristics of variation came from the related literature on thinking about variability in data (e.g., Shaughnessy Ciancetta, Best, & Canada, 2004; Makar & Canada, 2005; Canada & Ciancetta, 2007). These characteristics include relative spread, extreme values, and range. For example, consider these exemplars of more specific reasoning about variation:

EK-T: (D) The data is more spread out going EastBound than it is going WestBound
AD-T: (D) Because the times for the EastBound trains are very spread out while the WestBound trains' times are clustered together.

Although the above subjects did not capture formal numerical descriptors of variation about a mean, they did use informal language to convey an intuitive sense of the relative spread of data. Other subjects included variability characteristics in their responses by paying attention to extreme values:

AU-T: (D) No, because the EastBound has more outliers and is more scattered
AN-T: (D) One EastBound train took 59:40 while the longest WestBound train took only 59:15, and that is almost a 30-minute difference.

Note how AU shows sensitivity to the presence of outliers, while AN includes references to the maximal value in each data set. In addition to specific characteristics of variation captured by responses suggesting a focus on relative spread or extreme values, some responses made explicit connection between both maximum and minimum values:

AU-T: (D) EastBound has a higher range, from 58:25 to 59:40, and WestBound's smaller range is from 58:45 to 59:15
DM-T: (A) The range is 1:15 seconds EastBound and 0:30 seconds WestBound.

It was interesting to note that even while acknowledging the different ranges, DM still chose to agree with the hypothetical argument in the task.

Distributional Reasoning

As noted in the previous exemplars, some responses focused more on centers and others on informal notions of variation. However, in line with the previous research (e.g., Shaughnessy Ciancetta, Best, & Canada, 2004), responses coded as distributional needed to reflect an integration of both center and variation. Overall, more than three times as many preservice teachers (43.1% or 25/58) as middle school students (14.0 % or 7/50) had responses that fit the distributional category. Exemplars include the following:

HH-T: (D) Because the mean is an average, and to get an average you will most likely use varying numbers. All the times for the most part on EastBound trains are different. Just because the mean is the same doesn't change that.
AJ-T: (D) The data is different, although the average is the same. We can see, for example, the difference in consistency of the WestBound train, where the times are closer together, and hold nearer to schedule.

Note the richness in the exemplars provided above, as subjects integrate center and spread in their consideration of the two data sets. We see, for example, how HH understands about combining "varying numbers" to get an average. AJ actually lays the groundwork for making an informal inference, in the way that the WestBound may be more reliable train because it holds "nearer to schedule".

DISCUSSION

Of all 58 preservice teachers, 20 (34.5%) initially agreed and 38 (65.5%) disagreed with the hypothetical argument of "no real differences" between the trains. These ratios seem not so different from the 50 middle school students, 16 (32%) of whom initially agreed and 34 (68%) of whom disagreed. Since this research is about how the subjects reasoned distributionally, it was crucial to dig into their explanations. The exemplars show how subject responses could reflect a focus on centers, on variation, or on both.

Since responses could be coded for multiple aspects, including some facets of statistical thinking not reported in this paper, it is interesting to look at the breakdown of who agreed versus who disagreed with the hypothetical argument based on whose responses coded only for center, only for variation, or coded for distributional reasoning (both center and variation). Looking strictly within those groups of responses, the percentage of those subjects who agreed or disagreed is presented in Table 1:

Table 1. Support for "no real differences" by type of reasoning

Type of Reasoning	Preservice Teachers			Middle School Students		
	Subtotal	(A)gree	(D)isagree	Subtotal	(A)gree	(D)isagree
Only Centers	n = 12	75.0 %	25.0 %	n = 18	61.1 %	38.9 %
Only Variation	n = 16	31.3 %	68.8 %	n = 17	17.6 %	82.4 %
Distributional	n = 25	20.0 %	80.0 %	n = 11	18.2 %	81.8 %

Again, the percentages in Table 1 are calculated from the respective subtotals of responses falling within the given types of reasoning (the numbers do not total 58 for the teachers or 50 for the students because some subjects had explanations that went outside the themes of this paper). Comparing those percentages with the total pool of subjects (34.5% of teachers agreeing and 32% of students agreeing), two interesting observations can be made. First, while more than half of all subjects disagreed, the majority of those subjects who only relied on reasoning about centers agreed with the hypothetical argument of "no real differences". Second, and most important, the subjects whose responses reflected distributional reasoning had the highest percentage of disagreement with the hypothetical argument. For both groups of subjects it seems that reasoning only based on centers led to poorer inferences while distributional reasoning led to better inferences.

In connecting these findings to that of previous research, Leavy's (2006) work included both pre-test and post-test analyses, while this study only reports on initial pre-test conceptions. Also, Leavy's subjects were graduate students, while the elementary preservice teachers in this study were undergraduates. However, similar descriptions are found in the initial conceptions of Leavy's subjects, in that an "overemphasis on measures of central tendency went hand in hand with the neglect of graphs and variability" (p. 106). Makar and Confrey's (2004) research uncovered similar findings using a different methodology. By having secondary teachers engaging in a more authentic statistical inquiry, the researchers found that qualitative descriptions of within-group variation "included statements about a distribution being 'tighter' or 'more spread out'" (p. 368). These informal descriptions align well with the articulation of variability given in this study by some middle school students and preservice teachers.

CONCLUSION

This study was guided by the question "What are the informal conceptions of distribution held by preservice teachers and middle school students as they compare two data sets?" Although limited to a single task that was by nature contrived to invite attention to identical means yet differing amounts of spread in two data sets, the research suggests that both preservice teachers and middle school students reflect on aspects of distributional reasoning, and this paper gives a sense of how those aspects manifested themselves in the responses of the subjects. However, just as prior research has shown (e.g., Shaughnessy Ciancetta, Best, & Canada, 2004), we also see that both subject groups make more limited comparisons of distributions when they focus on centers while not attending to the critical component of variability in data. In contrast, the distributional reasoners who attended to both center as well as variation made richer comparisons within the given task.

It is of particular interest how qualitatively similar many of the responses from the preservice teachers were to responses from the middle school students. A reasonable question for curriculum developers at the university level would be how to effectively use this similarity in promoting pedagogical awareness in novice teachers in the area of statistical reasoning. For example, knowing that preservice teachers may come into the university environment with some language and reasoning strategies that mirror those of their prospective students, lessons can be built upon this common ground to help teachers relate to students' emergent conceptions.

While further research is recommended to help discern the most effective ways of moving preservice teachers toward a deeper understanding of distributions, certain tasks such as the one profiled in this paper provide good first steps in helping universities offer opportunities to bolster the conceptions of the preservice teachers they aim to prepare. In turn, these novice teachers can then promote better statistical reasoning with their own students in the schools where they eventually serve.

REFERENCES

Canada, D., & Ciancetta, M. (2007). Elementary preservice teachers' informal conceptions of distribution. In T. Lamberg & L. Wiest (Eds.), *Proceedings of the 29th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 960-967). Lake Tahoe, NV: University of Nevada, Reno.960-967).

Cobb, G., & Moore, D. (1997). Mathematics, statistics, and teaching. *American Mathematics Monthly*, 104 (9), 801-824.

Konold, C., & Higgins, T. (2002). Highlights of related research. In S. J. Russell, D. Schifter & V. Bastable (Eds.), *Developing mathematical ideas: Working with data* (pp. 165-201). Parsippany, NJ: Dale Seymour Publications.

Leavy, Aisling. (2006). Using data comparisons to support a focus on distribution: Examining preservice teachers' understandings of distribution when engaged in statistical inquiry. *Statistics Education Research Journal* 5 (2), 27-45.

Makar, K., & Canada, D. (2005). Pre-service teachers' conceptions of variation. In Chick, H. L. & Vincent, J. L. (Eds.). *Proceedings of the 29th Conference of the International Group for the Psychology of Mathematics Education* (pp. 273-280). Melbourne, Australia: University of Melbourne.

Makar, K. M., & Confrey, J. (2004). Secondary teachers' reasoning about comparing two groups. In D. Ben-Zvi & J. Garfield (Eds.), *The challenges of developing statistical literacy, reasoning, and thinking* (pp. 327-352). Dordrecht, Netherlands: Kluwer.

Pfannkuch, M. (2006). Comparing box plot distributions: A teacher's reasoning. *Statistics Education Research Journal,* 5(2), 27-45.

Shaughnessy, J., Ciancetta, M., Best, K., & Canada, D. (2004). Students' attention to variability when comparing distributions. Paper presented at the *Research Presession of the 82nd Annual Meeting of the National Council of Teachers of Mathematics*, Philadelphia, PA.

Shaughnessy, J., Ciancetta, M., & Canada, D. (2004). Types of student reasoning on sampling tasks. In M. Johnsen Høines & A. Berit Fuglestad (Eds.), *Proceedings of the 28th Conference of the International Group for the Psychology of Mathematics Education* (pp. 177-184). Bergen, Norway: Bergen University College.

Watson, J. (2002). Inferential reasoning and the influence of cognitive conflict. *Educational Studies in Mathematics, 51,* 225-256.

Wild, C., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review, 67,* 233-265.