THE ROLE OF STATISTICS IN IMPROVING EDUCATION

BURRILL, Gail
Michigan State University,USA
burrill@msu.edu

ABSTRACT

The goal of teaching statistics is to foster an adult population capable of reasoning from and about data and making informed decisions based on quantitative information in the workplace, in their personal lives and as citizens. This paper describes examples of data available to those involved in the design and delivery of education and some of the statistical skills and reasoning necessary to understand, interpret and use that information about schools, teachers, and students to improve their educational systems. Educators in the United States increasingly have access to data such as achievement trends over time, item analyses from high stakes tests, or comparison data for states and comparable systems that can help them develop ways to improve student learning. Unfortunately, few educators have sufficient understanding of statistics to make use of this data to help prevent errors in decision-making.

THE PROBLEM

Citizens are being called upon to make increasingly complex decisions about policies and practices in the socio-political, workplace, and consumer arenas (Franklin & Garfield, 2006; Kader & Perry, 2006). Crucial skills to make informed decisions include the ability to explain, decide, judge, evaluate, and analyze information (Rumsey, 2002). In most US curricula, the study of data analysis begins in the middle grades (ages 12 to 14). The ideas are sometimes revisited in high school, but the level of cognitive complexity is rarely deepened, and the applications are typically procedural. At the tertiary level, education students do not always take a statistics course; if they do, the course is often a mathematical statistics course. Growing evidence suggests that students who learn statistics this way find it difficult to apply statistical concepts in real settings where the concepts are clearly applicable and their use could help prevent errors in making decisions (Garfield & Ben-Zvi, 2007). Even those who have taken a course often do not leave with any lasting understanding of statistical concepts (Clark et al, 2003). They have not grasped that measures of center without corresponding measures of spread do not give an accurate picture of the situation, the need to understand and interpret the variability in data, the importance of considering sample size or the need to have a “standard” in order to measure both attainment and improvement. Many involved in the educational system have never had a course in statistics that would give them the background and tools to make sense of the data about student achievement with which they are confronted. The dilemma is that as data about teaching and learning becomes more readily available and the tools for analyzing the data become more sophisticated, the ability to produce useful information from the analyses is outpacing the capacity of the field to use the knowledge productively.

This paper draws on professional development work with teachers involved in a Mathematics Science Partnership project, Promoting Rigorous Outcomes in Mathematics and Science Education (PROM/SE), funded by the US National Science Foundation. The project was focused on improving mathematics and science learning through assessments of students and teachers, creation of clear standards and expectations across grade levels, and capacity building for teachers and administrators. Directed by William Schmidt at Michigan State University, the project collected baseline data from 60 school districts and 3,000 teachers and their students at five different sites in Ohio and Michigan. For project activities, districts selected teams consisting of an administrator and teachers from each of the district schools and/or grade level clusters, which typically meant about 250 participants involved in the professional development activities. Participants were divided into groups depending on the focus of the sessions. This paper describes issues that emerged in efforts to develop the capacity of the district teams to understand data about student achievement, a preliminary step in the larger PROM/SE research agenda. Sources for the observations in the paper include project staff who worked as small group facilitators, reflections and evaluations collected at professional development sessions, and responses participants gave on worksheets designed to scaffold data interpretation. Given the diversity of schools and teachers in the study, it seems reasonable to infer the issues that emerged might be relevant in the larger educational community. The discussion includes suggestions for possible strategies to address the challenges in enabling educators to use statistics as an effective tool for guiding practice.

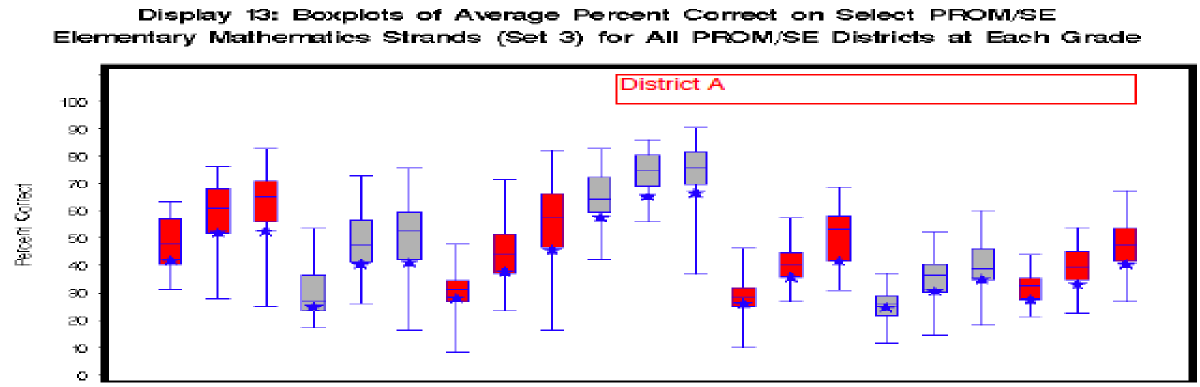
Center And Spread

"The median income is $79,000." "The car averages 36 miles per gallon." School curriculum, the media, and common usage all privilege the center of a distribution of data with little attention to how the data might vary around that center (Shaughnessy, 1997). The research suggests that, while students learn to compute measures of variability such as standard deviation or interquartile range, they often do not understand what those measures represent, either numerically or graphically, how such measures relate to other statistical concepts, nor why they are important (Bakker, Biehler & Konold, 2004; Garfield & Ben-Zvi, 2007; Hammerman & Rubin, 2004). The evidence suggests the system has not been able to equip those involved in education with the knowledge and skills they need to understand and make sense of how variability can help them understand data related to the work of educating students (Ben-Zvi & Garfield, 2004; delMas & Liu, 2005; Reading, 2004).

Boxplots are becoming common as a way to display and summarize educational data, but research has shown some of their features make them particularly difficult to interpret. This is true for educators who encounter many of the same difficulties students have in understanding how to read them: linking length of segments or boxes with frequency of data, misunderstanding the lack of relationship between plot area and frequency, unable to reason from general characteristics of a set of data, problems comparing across groups (Bakker et al, 2004; Makar & Confrey, 2005; Bakker & Gravemeijer, 2004). The next section describes the results of using boxplots to summarize student achievement data for school administrators and teachers.

ANALYZING STATE AND DISTRICT DATA

PROM/SE collected baseline information about student achievement for students in grades 3 to 12 in 60 districts from tests developed by the project. Because the focus of PROM/SE was on curricular coherence both within and across grade levels (responding to the charge that the mathematics curriculum in the United States is a "mile high/inch deep") three tests were administered, for students in grades 3-5, 6-8 and 9-12. Theoretically, this would allow districts to identify grades in which a particular content area seemed to be the focus and to trace progress over the three grades as well as consider student performance on concepts over all the grades. Boxplots (see figure 1 for an example) and other graphical representations were used to display summaries of achievement scores.



Grades 3 4 5 | 3 4 5 | 3 4 5 | 3 4 5 | 3 4 5 | 3 4 5 | 3 4 5

2- and 3-D Shapes | Perpendicular & Parallel Lines | Angles & Measure | Properties Geometric Figures | Angle Sums | Area & Perimeter | Volume & Surface Area

Figure 1: Student Achievement on Elementary Mathematics Strands for all PROM/SE Districts

Project staff designed activities to help educators interpret the different data presented. District teams then met with a project facilitator to work through district data in relation to the overall project data. The facilitators reported a large degree of confusion and recommended revisiting the task.

According to the facilitators' observations and participant responses to "Reading Your District Data" worksheets, teams did not notice in figure 1 for example, that students seemed to perform best on 2 and 3-dimensional shapes and properties of geometric figures, but in the first area by fifth grade, about one fourth of the districts still had average scores below 53%. They did not see that third grade students from three fourths of the districts were able to successfully answer less than about 30% of the problems related to angle measure or sums of angles, which would seem to indicate that these concepts were not yet part of the curriculum. They did not notice the variability by district – such as from 15% to 80% correct on angle measure – nor the apparent lack of progress from fourth to fifth grade for some topics, such as parallel and perpendicular lines. District teams had difficulty making sense of what the data might suggest about the mathematics program in their own districts.

To overcome these problems, the next professional development session focused on enabling participants to understand the difference between percent correct and percentile, read and interpret box plots, describe how different graphs convey different information about the data, and use plots to compare two or more sets of data. Participants considered bar graphs showing individual scores in a class, converted the data to a dotplot on the number line and then considered a boxplot compared to the dotplot (figures 2 and 3).

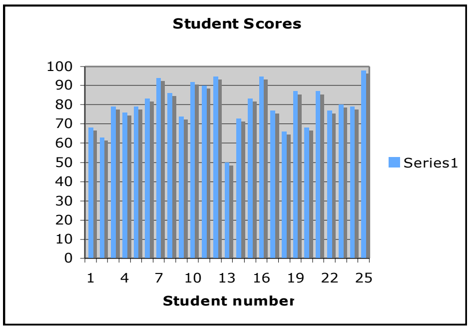


Figure 2: Bar Graphs of Student Scores

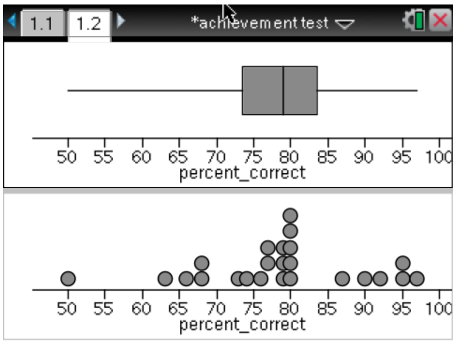


Figure 3: Dot/Boxplot of Student Scores

Participants were asked to answer a series of questions about boxplots, discuss distributions of scores and compare groups (figures 4 and 5).

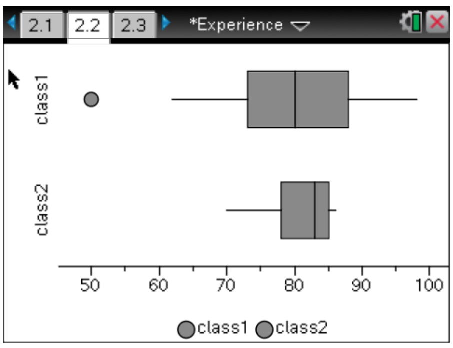


Figure 4: Comparing Classes

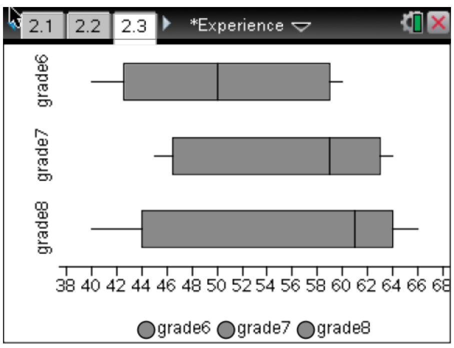


Figure 5: Comparing Grades

The post-session reflections indicated continued confusion. The next professional development session approached the problem from a different perspective, using data collected from the participants. The number of years of teaching experience was used to build understanding of mean and to illustrate the role variability plays in moving from a distribution of a population to a distribution of sample means. The mean years of experience for participants at different tables were collected and displayed in boxplots (figures 6 and 7). The discussion connected these ideas to data from participant schools and districts.

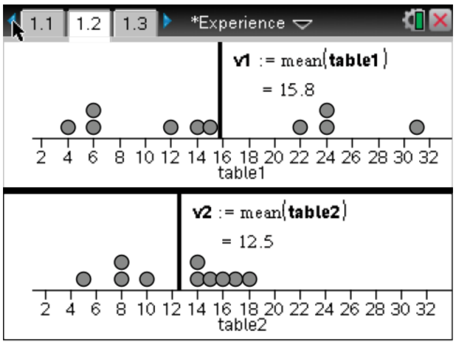


Figure 6 Years teaching experience by table

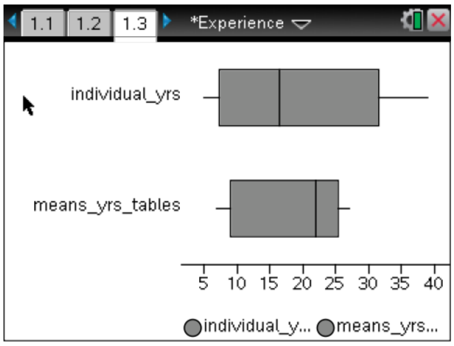


Figure 7 Teaching experience

Although participants were engaged in these and other activities including district central office visits by project staff to explain the data, many were unable to come to grips with the variability, relate medians and quartiles from boxplots to mean scores, or adapt to shifts in distributions such as the scores of all students in a school, mean scores by grade across all districts or the mean scores by grade of all project schools. Teams found it difficult to examine how student performance in a grade varied over a set of content expectations and how the graphs displayed the progression of students in consecutive grades. One large group of districts eventually transformed the boxplots into bar graphs with the median as the height. The project staff ultimately decided to report the data using only a table of mean achievement scores. While the standard deviation was included in the table, very few participants could accurately interpret what this statistic indicated about the data.

The following section describes a second situation in which district teams were asked to think about their achievement data.

INTERPRETING NUMBERS

In another phase of the project, district teams were asked to analyze their scores on the high stakes achievement tests given to all grade 3 to 8 students in each state. Scores are usually reported to districts in a table with student, class and state scores and may include item analyses. While districts (and the media) were keenly interested in the final percentage of students deemed "proficient" by the state based on student test scores, most had not examined their data in relation to other district or state data nor had they done long-term analysis of the trends. When district teams were given the data in table 1, they suggested the three districts should be commended for improvement.

Table 1 Overall Percent Proficient Grades 3-8

District/Year	2005	2006	2007
Brown	32%	29%	48%
Red	36%	28%	43%
Green	41%	33%	51%

No one questioned the drop in 2006 and large improvement in 2007 for all districts; such data should raise questions about the comparability of the tests from year to year or whether the level of proficiency was changed at the state level (as was the case).

In table 2, the large variation in scores for some of the districts between the two years raises the question of sample size. And indeed, the districts represented in the table were primarily small rural districts often with only one class at each grade level. The achievement for a small number of students each year will be greatly influenced by outliers, and the scores might not be indicative of a genuine change in performance. Yet Districts F and J were celebrating their improvement, and the superintendent in District N was accused of leading the district in the wrong direction.

Table 2 Percent Proficient in Grade 8

District	2006	2007	Difference
D	46	46	0
E	38	41	+3
F	27		+10
G	48	39	-9
H	53	53	0
I	47	40	-7
J	38	49	+11
K	47	42	-5
L	24	29	+5
M	56	62	+6
N	61	49	-12

One district established a test preparation course for their high school mathematics students, and at the end of the year, the district recommended the continuation of the course because of its impact on student scores. The data (table 3), however, suggests other factors might be involved as the scores in all but social studies (which fell three percentage points) rose more than in mathematics, suggesting more evidence is needed before a long-term commitment to the program is made.

Table 3 District A Percent at or above proficient on high school state assessment

Year	Science	Social Studies	Writing	English/Language Arts	Reading	Math
2008	64	78	42	52	64	47
2007	50	81	28	43	54	40

Few districts organized their state achievement data to look at the results over time and cohorts of students. When prompted to do so, District B, a large district, noted for example that the scores in grades 3 and 4 were considerably lower in 2008 than in previous years with a relatively large drop for the same cohort of students as they moved from 3rd to 4^{th} (16.6%) and from 4^{th} to 5^{th} grade (17.8%). They had adopted a new textbook series for grades K-5. The data alerted them to the need to monitor scores over the next several years as teachers adjusted to the new series and to attend to the alignment of the content in the new series with the state assessment. (For example, teachers found one item with very low performance asked students to identify the multiplicative inverse of a number; their new text used only the term reciprocal). Overall grade 5 scores were much lower than those of other grades; this might have been due to a variety of factors but in fact turned out to be common across the state and did not seem to be a district problem. No textbook changes were made in middle school, yet the scores in grade 7 for 2008 dropped 12% for the same cohort of students; discussion revealed that one of the three teachers for that grade level was a long-term substitute with little background in mathematics.

Table 4 District B percent proficient over time

	Gr 3	Gr 4	Gr 5	Gr 6	Gr 7	Gr 8	Gr 10
2005	67.4%	62.5%		53.4%	38.9%	47.1%	76.8%
2006	72.4%	74.6%	52.9%	60.6%	55.6%	56.0%	74.5%
2007	80.7%	76.0%	61.9%	65.0%	58.1%	62.7%	70.5%
2008	69.0%	64.1%	58.2%	61.4%	53.0%	64.1%	67.8%

CONCLUSION

The teachers and administrators involved in PROM/SE were typical, with varying degrees of preparation to teach mathematics and a desire to do the best for their students. Yet their training had not prepared them to use tools that would enable them to look systematically at what data could reveal about their focus in classrooms and the achievement of their students. The dilemma is that as data about teaching and learning becomes more readily available, the ability to produce useful information from analyzing the data is outpacing the capacity of the field to use the knowledge productively. This poses serious questions: what degree of statistical literacy should we expect for every one? How can we structure opportunities to enable those in our educational systems to reach this level and to ensure that a significant cadre in a school system is able to use statistics to improve their system?

One suggestion is to engage students in relevant data analysis activities focused on questions where answers could matter. Students may find analyzing student achievement for their school, comparing achievement across states or investigating the typical amount spent on education interesting and perhaps more productive than analyzing hand spans or estimates of a doorknob's height. Exploring activities leading to figures 8 and 9 might help students recognize class averages can have a large variation from year to year for the same teachers.

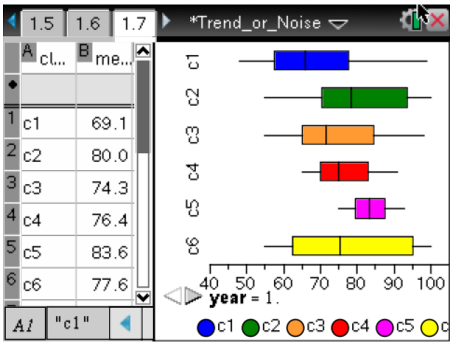


Figure 8 Student scores by teacher 2007

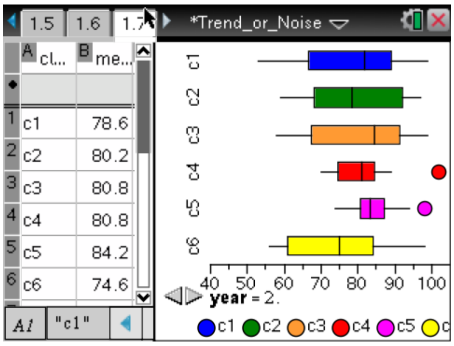


Figure 9 Student scores by teacher 2008

Another suggestion is to consider the use of interactive dynamic software focused on core statistical concepts. For example, figures 10 and 11 illustrate an interactive file for either computers or handheld devices addressing the issue of density vs. length of interval in a boxplot. Figures 12 and 13 illustrate the limitations in drawing conclusions about the shape of a distribution from a boxplot.

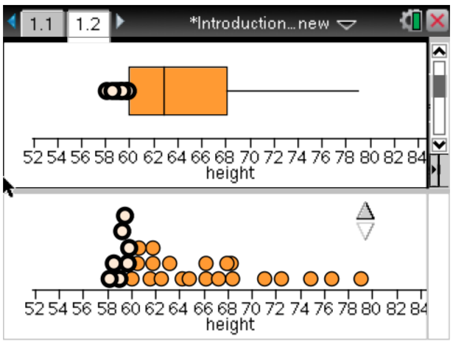


Figure 10 Lower fourth of the data

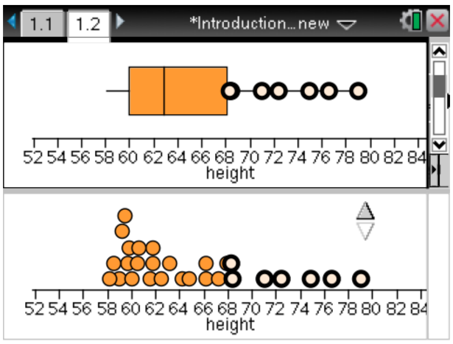


Figure 11 Upper fourth of the data

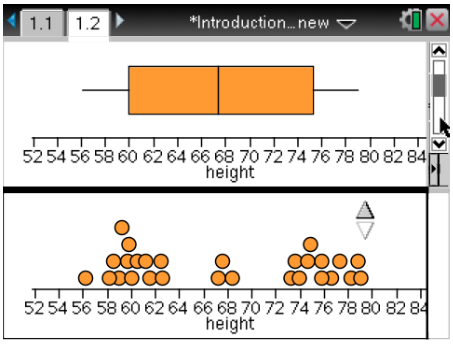


Figure 12 Symmetric distribution A

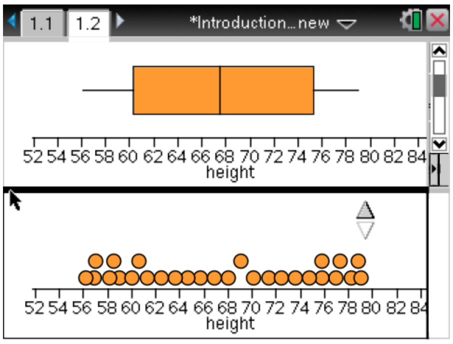


Figure 13 Symmetric distribution B

When such activities are used in combination with questions that engage students in thinking and reasoning about statistical concepts, some evidence is emerging that students gain a better understanding of the concepts (Chance & Rossman, 2006; Garfield & Ben-Zvi, 2008; Pape et al, 2010).

The challenge that faces the statistical education community is how to take advantage of this research and find ways to apply it that will increase the capacity of the field to improve educational practices by reasoning and thinking with data.

REFERENCES

Bakker, A., Biehler, R., & Konald, C. (2004). Should young students learn about boxplots? In G. Burrill and M. Camden (Eds.) *Curricular development in statistics education: IASE 2004 Roundtable on Curricular Issues in Statistics Education.* Voorburg, the Netherlands: International Statistics Institute. www.stat.auckland.ac.nz/~iase/index.php

Bakker, A., & Gravemeijer, K. (2004). Learning to reason about distribution. In D. Ben-Zvi & J. Garfield (Eds.) *The challenge of developing statistical literacy, reasoning, and thinking.* (pp. 147-167). Dordrecht, the Netherlands: Kluwer Academic Publishers.

Ben-Zvi, D., & Garfield, J. (2004). Statistical literacy, reasoning, and thinking: Goals, definitions, and challenges. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 3-15). Dordrecht, The Netherlands: Kluwer Academic Publishers (Springer).

Chance, B., & Rossman, A. (2006). Using simulation to teach and learn statistics. In A. Rossman & B. Chance (Eds.), *Proceedings of the Seventh International Conference on Teaching Statistics*. [CD-ROM]. Voorburg, The Netherlands: International Statistical Institute.

Clark, J., Karuat, G., Mathews, D., & Wimbish, J. (2003). The fundamental theorem of statistics: Classifying student understanding of basic statistical concepts. Unpublished paper.

delMas, R., & Liu, Y. (2005). Exploring students' conceptions of the standard deviation. *Statistics Education Research Journal*, *4*(1), 55–82. www.stat.auckland.ac.nz/~iase/publications.php?show=serj

Franklin, C. & Garfield, J. (2006). The GAISE project: Developing guidelines for K-12 and college courses. In G. Burrill (Ed.). *Thinking and reasoning with data and chance*: National Council of Teachers of Mathematics 2006 Yearbook. Reston VA: National Council of Teachers of Mathematics

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

Garfield, J. & Ben-Zvi, D. (2007). How students learn statistics revisited: A current review of research on teaching and learning statistics. *International Statistical Review*, 75(3), 372-396.

Hammerman, J.K., & Rubin, A. (2004). Strategies for managing statistical complexity with new software tools. *Statistics Education Research Journal, 3*(2), 17-41. www.stat.auckland.ac.nz/~iase/serj/SERJ3(2)_Hammerman_Rubin.pdf

Kader, G. & Perry, M. (2006). A framework for teaching statistics within the K-12 curriculum. In A. Rossman & B. Chance (Eds.). Proceedings of ICOTS 7. Voorborg, The Netherlands: International

Association of Statistical Education
www.stat.auckland.ac.nz/~iase/publications/17/2B3_KADE.pdf

Makar, K. & Confrey, J. (2005). Using distributions as statistical evidence in well-structured and ill-structured problems. In K. Makar (Ed.) *Reasoning about distribution: A collection of current research studies. Proceedings of the Fourth International Research Forum on Statistical Reasoning, Thinking and Literacy* (SRTL-4), University of Auckland, New Zealand, 2-7 July. Brisbane: University of Queensland.

Pape, S. J., Irving, K. E., Owens, D. T., Boscardin, C. K., Sanalan, V. A., Abrahamson, A. L., et al. (2010). *Classroom connectivity in algebra I classrooms: Results of a randomized control trial*. Manuscript submitted for publication.

PROM/SE. Promoting Rigorous Outcomes in Mathematics and Science. Michigan State University Mathematics Science Partnership. W. Schmidt Principle Investigator. Supported by NSF Cooperative Agreement EHR – 0314866

Reading, C. (2004). Student description of variation while working with weather data. *Statistics Education Research Journal*, *3*(2), 84-105. www.stat.auckland.ac.nz/~iase/serj/SERJ3(2)_Reading.pdf

Rumsey, Deborah J. (2002). Statistical literacy as a goal for introductory statistics courses. *Journal of Statistics Education*. 10 (3).

Shaughnessy, 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 meetings of the Mathematics Education Research Group of Australasia) (pp. 6-22). Rotorua, New Zealand: MERGA.