USING DATA COMPARISON TO SUPPORT A FOCUS ON DISTRIBUTION: EXAMINING PRESERVICE TEACHERS' UNDERSTANDINGS OF DISTRIBUTION WHEN ENGAGED IN STATISTICAL INQUIRY

AISLING LEAVY
Mary Immaculate College, University of Limerick
Aisling.Leavy@mic.ul.ie

ABSTRACT

This exploratory study, a one group pretest-posttest design, investigated the development of elementary preservice teachers' understandings of distribution as expressed in the measures and representations used to compare data distributions. During a semester-long mathematics methods course, participants worked in small groups on two statistical inquiry projects requiring the collection, representation, analysis and reporting of data with the ultimate goal of comparing distributions of data. Many participants shifted from reporting descriptive exclusively to the combined use of graphical representations and descriptive statistics which supported a focus on distributional shape and coordinated variability and center. Others gained skills and understandings related to statistical measures and representations yet failed to utilize these when comparing distributions. Gaps and misconceptions in statistical understanding are discussed. Recommendations for supporting the development of conceptual understanding relating to distribution are outlined.

Keywords: *Statistics education research; preservice teacher education; distribution; statistical inquiry, data comparison, teacher knowledge*

1. INTRODUCTION

The teaching of statistics in elementary schools has received increased attention and priority over the past three decades. The release of the *Curriculum and Evaluation Standards for School Mathematics* by the National Council of Teachers of Mathematics (1989), which incorporated a strand focusing on data analysis and probability, and the publication of the *Guidelines for the Teaching of Statistics K-12* (1991) by the American Statistical Association, are two important landmarks. The increased focus on elementary level data analysis and statistics is evident in the proliferation of curricula designed specifically for younger students, such as the *Used Numbers* Project (Technical Education Research Centers and Lesley College, 1989), *Mathematics in Context* (National Center for Research in Mathematical Sciences at the University of Wisconsin/Madison and Freudenthal Institute at the University of Utrecht, 1997-1998), the *Investigations in Number, Data, and Space* (TERC, 1998), and the *Connected Mathematics* Project (Lappan, Fey, Fitzgerald, Friel, & Phillips, 2002). We also see computer software, minitools, and tutorial tools developed for elementary and middle grade students, such as *Tabletop* and *Tabletop Jr.* (Hancock, 1995; Hancock, Kaput & Goldsmith, 1992), *Statistical Minitools* (Cobb, Gravemeijer, Bowers, & McClain, 1997),

Statistics Education Research Journal, 5(2), 89-114, http://www.stat.auckland.ac.nz/serj

Authentic Statistics Stack (Lajoie, 1997), and *Tinkerplots* (Konold, 1998; Konold & Miller, 2001).

The focus of recent curricula and software has been on the notion of distribution and how to support students in understanding distribution. One area which is lacking, however, is an analysis of the ways in which teachers understand and are prepared to teach fundamental notions associated with distribution.

2. SCIENTIFIC BACKGROUND

2.1. DESCRIBING DISTRIBUTION

Distribution refers to the arrangement of values of a variable along a scale of measurement resulting in a representation of the observed or theoretical frequency of an event. Descriptive statistics are indices of distribution: they summarize complex data into measures that can be compared against each other to ascertain the nature of a dataset, and the degree to which two or more datasets are similar. Central elements in the development of a concept of distribution are notions of central tendency, variability, symmetry (skew) and relative frequency (kurtosis), each of which can be modeled using descriptive statistics.

One way to get a handle on distribution is through identifying landmarks and trends (Friel, Mokros, & Russell, 1992) in data. Taken together, identifying data landmarks (outliers, gaps) and generating measures that index certain characteristics of the data (for example, measures of center and variability) provide insights into properties of any given distribution. While descriptive statistics are central components of any treatment of distribution, a focus on them alone can, as cautioned by Makar and Confrey (2005), “aggravate the focus on individual points” (p. 28). Graphical representations serve as useful tools to communicate aspects of a distribution as they facilitate a focus on aspects of the data that may be missed with the use of descriptive statistics alone. Graphics have been described as *revealing* data (Tufte, 1983) and as being superlative to statistical computations in revealing information about data. However, little is known about the ways in which learners use graphical representations to communicate aspects of a distribution.

Research reveals an emphasis on measures of central tendency as a means to describe data distributions resulting in an overemphasis on centers and the corresponding neglect of the variation found within a given distribution (Shaughnessy, 1992, 1997). Variation is a critical component of, and inextricably linked to, the concept of distribution and has been found to play a central role in children’s thinking (Cobb, 1999; Konold & Pollatsek, 2002; Watson & Kelly, 2002). An understanding of distribution requires an awareness of the propensity of a variable to vary and comprehension of how that variability contributes to the notion of the distribution as an aggregate rather than a collection of individual data points.

2.2. USING DATA COMPARISON TO SUPPORT THE DEVELOPMENT OF UNDERSTANDINGS RELATED TO DISTRIBUTION

A critical statistical notion for learners is that of dataset as an entity, in other words developing a ‘statistical perspective’ (Konold, Pollatsek, Well, & Gagnon, 1997). Holding a statistical perspective requires a focus on the dataset as a collective rather than focusing on individual data values. By focusing on comparing distributions students are provided with a conceptual structure that facilitates a focus on aggregate (Cobb, 1999).

More specifically, when comparing datasets the activity leads to consideration of the shape, center and variability of a distribution of data, in turn, providing a context for examination of the distribution as an entity.

Another reason for engaging students in the activity of comparing datasets is due to the focus on variability that is nurtured. Unexplained variation in data creates noise and the primary purpose of many statistical techniques is to unearth the signal within the noise (Konold & Pollatsek, 2002; Wild & Pfannkuch, 1999). Comparing datasets requires a learner to examine the variation both within and between distributions of data. This requires identifying signals or patterns within a dataset worthwhile of attention, and comparing these signals against those emitted by the comparison dataset. Identification and communication of these signals reveals aspects of an individuals' understanding of the notion of distribution.

2.3. PRESERVICE TEACHERS' UNDERSTANDING OF DISTRIBUTION

Given the relatively recent election of data analysis as a focus of mathematics instruction at the K-12 level, it is conceivable that many teachers may be teaching statistical content that they themselves have little experience with as learners. Lajoie and Romberg (1998) comment that statistical concepts may be as new a topic for teachers as for the students they teach and recommend that "teachers must be provided with appropriate preservice and in-service training that will give them the knowledge base they need to feel comfortable teaching about data and chance" (p. xv).

Much of the current research focuses on preservice teachers' understandings of measures used to index distributions of data (Canada, 2004; Gfeller, Niess, & Lederman, 1999; Heaton & Michelson, 2002; Makar & Confrey, 2002). Many of these studies converge on the same finding – preservice teachers' understanding of measures of center tends to be procedural rather than conceptually-based. Leavy & O'Loughlin (2006) report on elementary preservice teacher's fluency in using the mean algorithm but identified gaps in conceptual understanding. Indicators of poor conceptual understanding were lack of understanding of the mean as a ratio, difficulty solving weighted means problems, and poor analog knowledge of the mean (a concept akin to Skemp's (1979) concept of relational knowledge). The prevalence of procedural understandings is further supported by Gfeller, Niess, & Lederman's (1999) finding that computational algorithms were the most prevalent method used by preservice teachers for solving problems related to the mean.

Studies examining variability indicate that the provision of coherent and meaningful statistical activities can lead to gains in understandings of variability. Canada (2004) found that following activities involving chance and computer-generated simulations, preservice teachers' predictions of variability moved from expectations of way too little or too much variation to more realistic expectations of variation. It has also been found that preservice teachers are more likely to use measures of variability to represent a distribution if the data are presented graphically (Makar & Confrey, 2005). This suggests that when choosing methods to represent a distribution of data, merely presenting the data graphically may draw attention to the variability of the data and make variability " ... perhaps more compelling than any measure of center" (pp. 36-37). Other advantages of using graphical representations are identified by Hammerman and Rubin (2004) who reveal that having access to a visual representation of a distribution may influence the value(s) that one chooses to represent that distribution. The authors comment on the low occurrence of comparisons based on means or other measures of central tendency and assert that "seeing a distribution makes it harder to accept a measure of center, especially a mean, as being representative of the entire distribution" (pp. 36-37).

A recent analysis of the literature revealed that only 2% of published research in mathematics education was devoted to probability and statistics (Lubienski & Bowen, 2000). Within this group, there is a visible absence of studies focusing on understandings of *elementary* preservice teachers, resulting in an inadequate picture of the preparedness of our elementary teachers to teach this new and emerging field of study. Unlike their secondary counterparts, elementary teachers are not expected to possess as strong or as broad a foundation in mathematics. However, given their role as primary mathematics educators of young children, a study of preservice elementary teacher's mathematical understanding is warranted. Subject matter knowledge in the preparation of teachers has been identified as a fundamental component of teacher education programs (Ball & McDiarmid, 1990). We now understand that poor mathematics content knowledge may lead to an overemphasis on limited truths and procedural rules (Stein, Baxter, & Leinhardt, 1990), inaccurate explanations (Borko, Eisenhart, Grown, Underhill, Jones, & Agard, 1992), and a lack of understanding of the appropriate representations to utilize when supporting the development of rich mathematical understandings in children (Borko et al., 1992). These relationships between content knowledge and instructional practices make it critical that mathematics teacher educators develop a greater understanding of elementary preservice teachers' statistical understanding.

2.4. PURPOSE OF THIS STUDY

The purpose of this exploratory study was to investigate the development of preservice teachers' understandings of distribution, expressed in the measures and representations used to compare distributions of data. Specific goals of interest were to:

(i) investigate the approaches used to compare distributions of data.
(ii) identify the statistical concepts focused on when reasoning about distributions of data, and examine the ways in which different understandings of these particular statistical concepts support or hinder the description, analysis, and comparison of datasets.
(iii) explore ways to support the development of rich understandings of distribution.

3. METHOD

3.1. PARTICIPANTS

Twenty-three participants were enrolled in a mathematics methods course in a university in the USA, as part of a one-year master's degree program leading to elementary teaching certification. Participants ranged in age from 22 to 55, seven were male. All participants held a bachelor's degree, with majors in Art, Business, Chemistry, Computer Science, Criminal Justice, Early Childhood Education, Economics, English, History, Psychology, Public Relations, Sociology, and Spanish. Three reported taking Advanced Placement Statistics in high school; almost half had no formal coursework in statistics.

3.2. APPROACH

This study employed a one-group pretest-posttest design and involved collection of baseline data or pre-test, an instructional intervention, and a post-test. The study represents a blend of components from two analogous research methodologies: Teaching experiment methodology (TEM, Steffe & Thompson, 2000) and the teacher development experiment (TDE, Simon, 2000). Using a blend of both methodologies supported the documentation of participants' understandings of distribution on entry to the study, the observation of changes in understanding over time, and the focus on the process of student learning and the concomitant teacher actions that promoted advances in statistical understanding. The study departed, though, from a true implementation of either methodology. The dearth of research relating to distribution resulted in the absence of an empirical research base from which to inform the construction of hypothetical learning trajectories, critical components of Teacher Experiment Methodology (TEM).

The whole class teaching experiment was conducted over a 15-week semester in collaboration with two teachers who were members of the four-person research team. Two or more research team members were present in the classroom during teaching sessions and were involved in the daily organization and evaluation of classroom mathematical practices. It was through juxtaposing these multiple perspectives that we gained rich and accurate insights into the development of statistical understanding. Weekly meetings of the research team focused around "taken-as-shared" (Cobb, 1999) interpretations of classroom activity, taken as shared meanings constructed as a result of cycles of construct development that supported the refinement of our own statistical knowledge. Meetings were used primarily as contexts within which to share interpretations of events, devise contexts in which to test these interpretations, and eventually refine and extend interpretations. This researcher-level data analysis supported the development of refined understandings of the process of statistical learning in addition to providing a focus on the development of teachers and their pedagogical understandings, an important component of the Teacher Development Experiment (TDE).

3.3. DESIGN AND TASKS

This study was organized in three parts (see Table 1): Pre-test (P), intervention with instructional components (I), and post-test (PT). The baseline data collection (P) was intended to probe participants' understandings of distribution. Central to this phase were the data collection and representation phases of *the bean experiment* investigation.

During the intervention phase (I), instructional activities supported the development of statistical reasoning, and made use of two tasks described later in more detail, the Beans investigation and Popcorn experiment. Firstly, students worked together to compare distributions of beans grown in different conditions. Secondly, several weeks of instruction focused on the stages of statistical inquiry and what was involved in moving through a statistical investigation (also known as data modeling, see Lehrer & Romberg, 1996). The foci of instruction were derived from analysis of the pre-test activities and related directly to distribution (representativeness, dataset as an entity). In an effort to maintain coherence between the areas of instruction, instructional themes were anchored within the umbrella concept of 'carrying out a statistical investigation'. Instruction was related primarily to the ongoing statistical investigations with the beans; the ability to ask participants to examine and assess their own work facilitated us in highlighting events that occurred in participants' own investigations. In preparation for the instruction, we had digital images of participants work (for example, the graphical representations constructed to compare datasets) or transcriptions of conversations or comments, which we then placed in our power point presentations. These records often became the focus of instruction and presented an opportunity for participants to reflect on and assess their work in light of research being examined as part of course experiences. The primary topics of instruction are presented on table 1. A third component of the instructional phase engaged participants in making journal reflections on statistical understandings related to distribution. The fourth component was involvement in focus group discussions relating to the strategies and approaches used to analyze bean data.

Finally, in the post-instruction phase (PT), we examined changes in participants' understandings about distribution. To that end, during the penultimate class session participants engaged in the popcorn investigation that involved comparison of the effect of refrigeration on the popping of kernels.

Engaging in statistical inquiry was a critical component of the study. From a pedagogical perspective, engaging preservice teachers in statistical investigation provides the potential to highlight statistical (and other) issues they may face when approaching statistical inquiry in their own classrooms. Secondly, engaging in 'real life' statistical investigation engages participants in the activities of statisticians and exposes them to real world 'messy' data. It was our goal that we set up statistical contexts that supported prospective teachers in learning 'statistical concepts in an environment much like the one recommended for students – one that is active … involving authentic data, and offering plenty of opportunities to build their conceptions through experiences with data' (Makar & Confrey, 2005, p. 30). Research focused on activities surrounding two investigations:

The bean investigation A semester-long statistics research project, *investigating optimal conditions for growing beans sprouts* (Appendix A), was one context within which the research was carried out. Participants were presented with an experimental design to determine which growing conditions supported the best growth.

Participants were divided into groups and provided with a bag of 25 lentil beans, a solution (lemon or water), a paper towel, a plastic sealable bag, and a card identifying the light intensity (light/dark) in which the beans would be placed. There were eight groups and four conditions: water/light, water/dark, lemon/light, lemon/dark. Participants were instructed to spray the solution on a paper towel, place the bean on the towel, fold the towel, and place in the plastic bag. The bags were sealed and placed in the labeled light intensity for seven days. The following week the beans were brought to class and the sprouts measured and recorded by the groups. Each group was then responsible, over the course of the semester, for constructing a hypothesis regarding the optimal condition for bean growth and determining what statistical measures or approaches they might utilize to test the hypothesis. Each group was instructed to analyze and compare the data collected by all eight groups and prepare a presentation of the findings. It was by engaging participants in this data comparison scenario that we established the need for individual distributions of data to be represented and indexed in a way that required a focus on distribution and in turn facilitated the comparison of the distributions.

The popcorn experiment The purpose of the post-intervention statistical investigation, *the popcorn investigation* (Appendix B), was to provide a context similar to the bean experiment where participants were once again engaged in a data comparison activity. The requirements of the task were similar to those of the pre-intervention task thus allowing identification and examination of changes in statistical understanding.

The investigation involved two samples of popcorn kernels (n=100 in each sample), one kept at room temperature and the other refrigerated, being popped in an open-top popcorn popper for 4 minutes. The position that each popped corn kernel landed was marked on the plastic 'target sheet' which had been placed beneath the popper. Groups were then presented with the data, asked to examine the distribution, and determine whether refrigerating the kernels influenced the distance that they reached when popped.

Table 1. Outline of statistical experiences over the course of the semester

Phase	Wk	Activity	Specific (teaching) experiences
P	4	Bean investigation: Setting up the experiment–planting	See section 3.3
P	5	Bean investigation: Data collection–measuring beans	See section 3.3
I	6	Instruction relating to asking research questions, collecting data	Instruction focused on overview of data modeling (Lehrer & Romberg, 1996), supporting children in formulating research questions, examples/analysis of questions children construct, types of data generated from questions, overview of categorical and numerical data types, overview of data collection methods (surveys, experiments, observations), supporting children in choosing appropriate data collection methods, identifying and overcoming obstacles and common difficulties faced by children when collecting data.
P	7	Bean investigation: Data representation phase	See section 3.3 and Appendix A
I	7	Bean investigation: Data analysis and comparison	The session was considered instructional in that students were presented the opportunity to learn when working in groups. Activity focused on the analysis and comparison of datasets.
I	9, 10	Reflection on statistical understandings arising from the bean experiment	Revisiting the stages of data modeling in the context of the bean experiment: What have we learned? What difficulties did we face (procedural, content understandings etc.) and how might this apply to classroom teaching of data analysis and statistics?
I	11	Focus group discussion relating to the strategies and approaches used during the bean investigation	The following are examples of questions asked during focus group discussion which focus on statistical misconceptions identified in weeks 1-10: What do descriptive statistics *not* tell us about a distribution? What does standard deviation mean in the context of a distribution of data? What is the relationship between the mean and sample size of the distribution? What role did zero values play when comparing distributions of data? Describe why some groups found the mean of group means? Why did others not? When are graphical representations useful? When/why might a box and whisker plot be used to represent a distribution? What does relative size of quartiles in a box and whisker plot communicate about the data? When/why might you use (or not use) a scatter plot?
I	11	Instruction relating to representing data (using graphical representations), analyzing and comparing data	Instruction on representing data focused on advantages of using graphs, guidelines from research about elements to attend to when constructing graphs, relationship between graphs and the data they represent (categorical/numerical), features of specific graphical representations and inherent advantages and disadvantages of their use (tables,

			pictograms, pie charts, bar graphs, line plots, stem and leaf plots, box and whisker plots), examples of graphs constructed by children, common errors children make when constructing graphs, categorizing representations of data. Instruction on analyzing and comparing data focused on defining distribution, important features of distributions, examining distributional shape (landmarks, bumps, gaps, outliers), locating measures of center on a distribution of data, examining models of the mean (leveling out, balance, fair share), locating indicators of variability on a distribution, thinking about skew when examining distributions, engaging in data comparison (Ben-Zvi, 2003; The “basketball problem,” McGatha, Cobb & McClain, 2002), examining children’s thinking about distribution, common errors children make when using, generating, and describing measures of center and variability.
I	13	Instruction on graphical representations; measures of central tendency and variability; means and weighted means	Instruction on graphical representations focused on revisiting features of graphical representations (Friel, Curcio & Bright, 2001), relationship between box and whisker plots and variability, what is a quartile?, distinguishing between univariate and bivariate data, scatter plots and the data they represent, why we use particular graphs, thinking about our bean data and graphs we used to represent the data (error analysis). Instruction on measures of central tendency and variability distinguishing between measures of central tendency, what research tells us about children’s understanding of the mean and median, features of the mean (e.g., Strauss & Bichler, 1988), revisiting models of the mean, identifying when measures of central tendency are most appropriate, examining use of the mean in the bean data experiment, identifying and indexing variability, representing variability using graphs, looking at use of variability in the bean experiments. Instruction on the weighted mean focused on what is a weighted mean, examining the elevator problem (Pollatsek, Lima, & Well, 1981), analysis of responses on weighted means problems (Leavy & O’Loughlin, 2006), examining the GPA problem (Pollatsek, Lima, & Well, 1981), analysis of responses on the GPA problems (Leavy & O’Loughlin, 2006), thinking about the bean data and weighted means.
PT	14	Popcorn investigation: Representing, analyzing, and comparing data	See section 3.3 and Appendix B
PT	15	Self report on growth in understanding	

3.4. DATA SOURCES

Several types of data were collected: videotape data, observational data, audio taped data of small group interactions, and student artifacts. Weekly sessions were videotaped. During statistical investigations two cameras were used: One focused on the whiteboard at the front of the room (where groups presented the outcomes of their investigation and where the instructor was located); the other moved between groups and focused on particular students during large group discussions. Teacher and researchers observations were recoded during the weekly sessions. There were always a minimum of two researchers in the room during each session, and for four weeks three researchers were present. The third data source was audio taped records of small group conversations during the investigations. Student artifacts were the fourth data sources and took several forms: student journals focusing on the development of statistical understanding, small group reports completed during the bean and popcorn investigations, and large presentation posters used at the culmination of the statistical investigations.

3.5. ANALYSIS OF DATA

In line with TEM and TDE there were two cycles of analysis: ongoing and retrospective analyses. Following each teaching episode, researchers examined the data and met as a group to discuss interpretations of classroom events. Each researcher had responsibility for examining in-depth a subgroup of students over the course of the semester. Discussion focused primarily on the types of understandings about distributions that were evident from videotape and audiotape recordings of participant activity during the class session. Each researcher shared the findings of her analysis and observations and when necessary situated the activity in relation to previous class sessions. The group then identified themes that emerged from the individuals' analysis of the data and constructed assertions relating to the themes. Individual researchers then revisited the data relating to their subgroup of participants with a view to finding supporting or contradictory evidence for the assertions. If the assertion was found to hold for a number of subgroups the cumulative class data were revisited in an effort to seek out evidence to triangulate the data. In cases where there was no supporting evidence we constructed a task or activity to present in the next teaching session (or in a focus group) that would test the assertion. On occasions where misconceptions relating to distributions were identified, statistical tasks were constructed to address misconceptions. It was through these cycles of hypothesis construction, data mining, triangulation, and hypothesis testing that understanding of the participants' understandings of distribution were developed. Figure 1 illustrates the interaction between data collection, analysis teaching episodes.

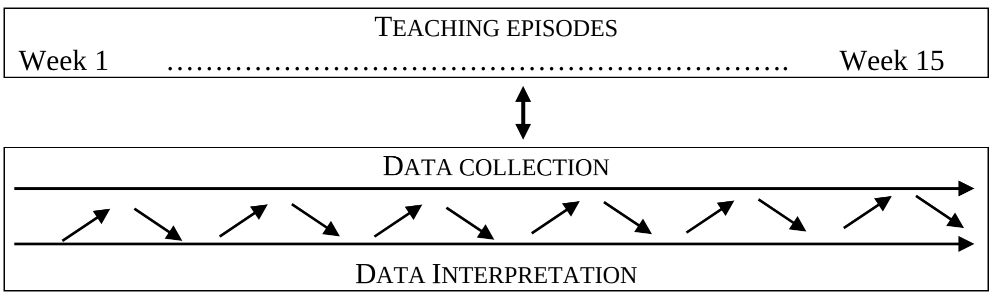

Figure 1. The interaction between data collection and data analysis in the context of a teaching experiment. Adapted from Lesh and Lehrer (2000).

Retrospective analysis at the end of the study involved re-examining participant activity over the entire semester. This allowed for a reanalysis of the findings and identification of supporting and contradictory evidence for the major claims arising form the ongoing weekly analysis of data.

4. RESULTS

4.1. INITIAL UNDERSTANDINGS RELATING TO DISTRIBUTION

Engaging in the bean experiment (Appendix A) provided insights into the types of distributional understandings held by preservice teachers. Participants compared lengths of germinated beans and presented their conclusions for feedback from their peers.

A lack of attention to distributional features of the data was apparent in the dominance of numerical methods for making data comparisons. Three of the groups (A-C) used descriptive statistics, alone, on which to base judgments about data. While use of the mean is appropriate as a comparative measure, absolute reliance on descriptive statistics is limiting as they provide merely one perspective on the data, that of centers, and do not take into account other features of the data (e.g., shape, variability). These groups did not invoke the use of graphical representations as a way to explore the datasets nor did they provide alternative perspectives not immediately apparent through the use of descriptive statistics. Group A's justification of their data comparative method follows:

> Our hypothesis was that the beans in lemon water would grow longer than the ones in water. We compared means because it sort of cancelled out the real high ones and the real low ones [data values representing bean heights] but incorporated every single piece of data. The main limitation is because there are high and low values they skew the data. … our answer was the sprouts in water were 20.2mm and in lemon water were 1.7mm so we were wrong in our hypothesis.

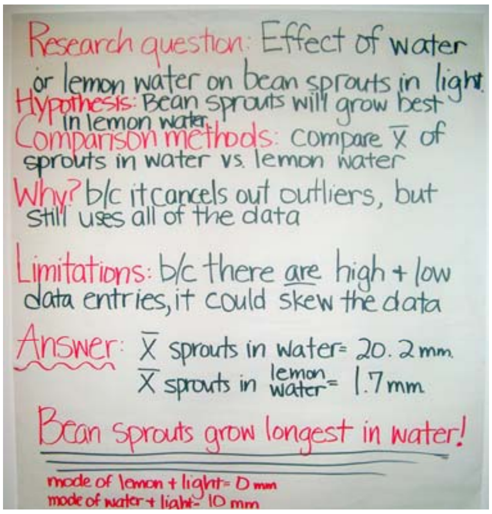

Figure 2. Illustration of Group A's approach to comparing datasets

The remaining five groups used a combination of graphical representations and descriptive statistics as data comparative tools. Interestingly, these students who used graphical representations in the pre-assessment did not have more statistical experiences with data prior to taking the course than their peers who did not construct graphs.

Examination of the graphical representations constructed by these groups, however, reveals that the graphs were used merely to illustrate summary statistics rather than illustrate distributional features. Four of these five groups used bar charts to represent group means resulting in a representation of subgroup means, rather than presenting a picture of the distribution of values along a scale of measurement (see Figure 3). The following is the response of Group D:

> We were interested in how lemon or no lemon affected the growth of beans in light or in the dark. Our hypothesis was that dark and lemon would yield the longest beans so … we were close in the sense that lemon was up there as number 2 and water and dark were number 1. So why did we do a bar graph? Cause it was easiest to show everything. So we got the averages for the four conditions. We didn't count zero cause they didn't germinate. The limitation is that outliers skew the average.

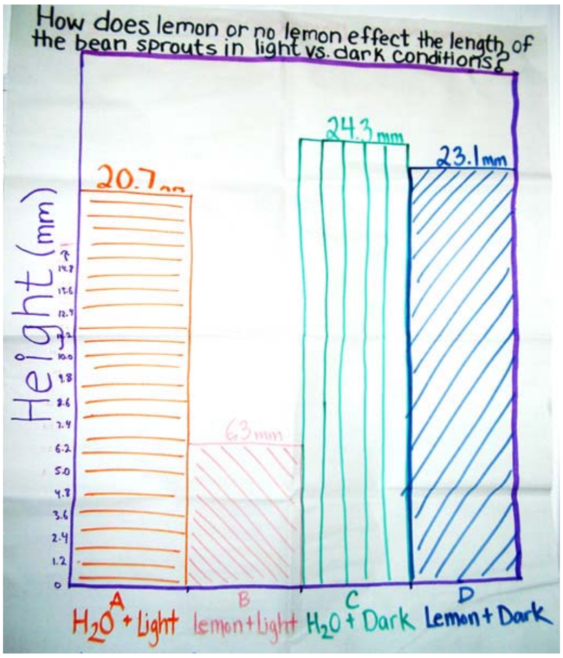

Figure 3. Illustration of group D's use of bar charts to illustrate sub group means

There was evidence of an attempt to represent the variability of the data. This is illustrated in the graphical representation of group G (see Figure 4). The representation resembles a double bar graph consisting of eight groups arranged in pairs; each pair corresponds to the two sets of data collected for each of the four conditions. The height of each bar represents the mean height of beans within each group (rather than the frequency of elements within the group). The line superimposed on each bar represents the range of heights of beans within each sample, thus the 'whiskers' represented the lower and upper limits of bean height. Thus the graphical representation was used as an instrument to report descriptive measures with the bars reporting group means and the whisker lines reporting the range of the data; indicating an attempt to coordinate both center and variability. Examination of group G's justification for inclusion of the range line reveals that they decided to report the variability given the large discrepancies in the sample statistics for beans grown under the same conditions. It seems that the unexplained variation in the measurement of bean heights was creating so much *noise* (Konold &

Pollatsek, 2002; Wild & Pfannkuch, 1999) that it was causing the group to refute the presence of a common signal across samples representing identical conditions. As Stan reported:

> We decided to go with a double bar chart cause of all the inconsistencies in measurement. So we did 8 groups and not 4 we also decided to do a box and whisker plot to show the full range … it is the orange line on the bar graph.

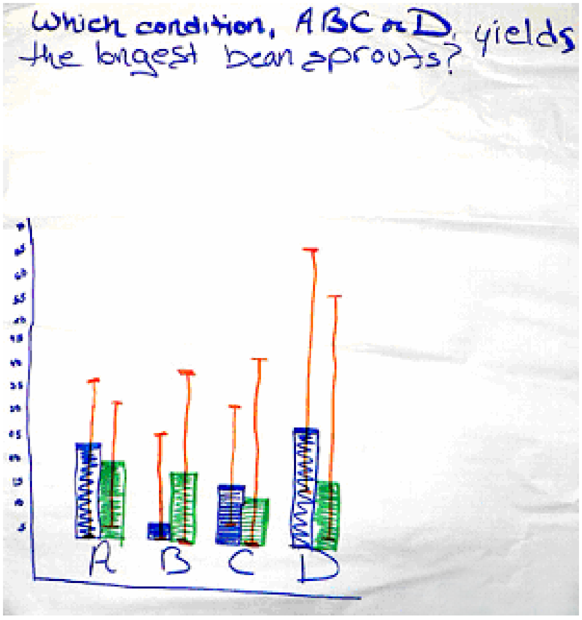

Figure 4. Illustration of Group G's attempt to coordinate variability and center

Analysis of pretest data did not reveal a relationship between participant's experiences with statistics prior to the study and the nature of the analyses carried out in the pretest.

4.2. END-OF-SEMESTER UNDERSTANDINGS RELATING TO DISTRIBUTION

We shift now to present some findings related to how groups coped with the Popcorn experiment, which as explained earlier was used to examine post-intervention performance of the groups. As compared to findings at the beginning of the semester outlined in the previous section, six of the groups shifted to using graphical representations to provide a picture of the distribution of data values. Groups reported that the selection of graphical representations to compare popcorn data was based on the capacity of the representations to highlight distributional features of the data, in contrast to the graphs used at the beginning of the semester which functioned merely to represent group means. This attention to global patterns in distributions of data was evident in the use of representations that highlighted distributional features in the data in particular the use of stem and leaf plots by five of the groups, and to a lesser extent box and whisker plots which were utilized by one group. The increased use of representations did not lead to the neglect of descriptive statistics. Figure 5 shows the data comparison strategy used by one group. Reporting the sample means and the range on the stem and leaf plots maintained the focus on measures of center and variability evident at the beginning of the study. As compared with their previous strategy (see Figure 3) this response represented a coordinated and detailed approach to analyzing data.

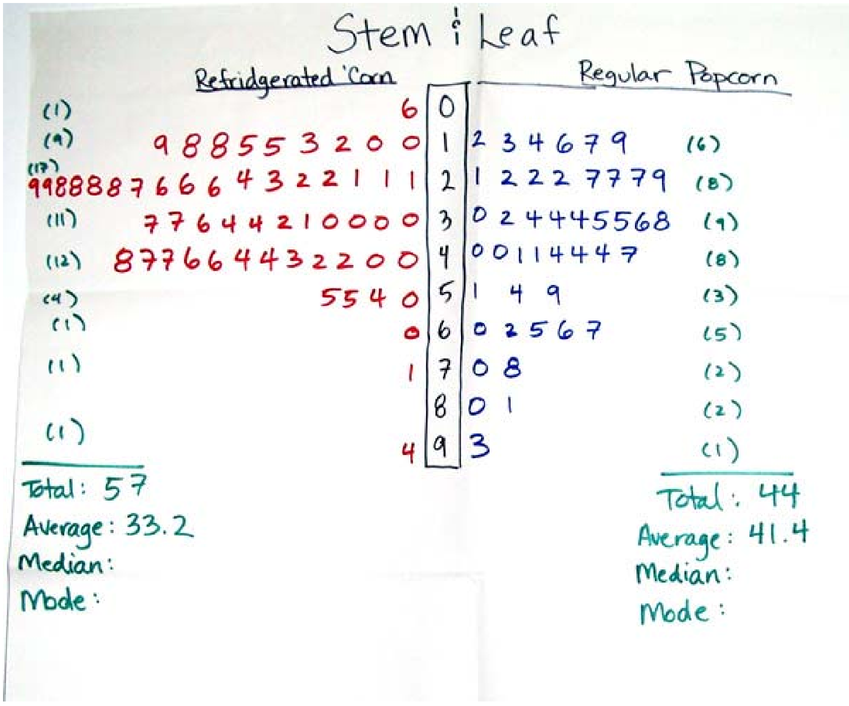

Figure 5. A comparison of distances by graphically illustrating distributional features in conjunction with the calculation of measures of central tendency

Tom: We chose this [stem and leaf plot] cause it gives you a way to see all the points at the same time and it gives you a sense of the distribution while getting all the data points there. And em … we thought that it was pretty representative. We predicted our mean would be higher for the room temp popcorn because of the way the tail skewed and we were right about that (pointing to the means)… and although the refrigerated popcorn did get a better yield it seems like the room temp popcorn did fly a little farther.

Barry: Melissa and I talked a lot about the kind of analysis we could use to think about the distribution curve. ... Well the curve for refrigerated has a bulk (pointing to the values at 10-40mm) where the curve for room temperature doesn’t seem to have a bump it seems to be a much smoother curve. And we noticed that the number of kernels that popped even though it was different it wasn’t all that big of a deal because the sizes were large enough that we could see what the curve would do if there were a ton more data points. You could kinda visually ... kind of .. visualize what the data would look like … how things would fall into that distribution curve.

R: During the discussion, you were trying to decide to do a box and whisker plot and you decided not to. Can you explain why?

Barry: Cause even though I like box and whisker it is difficult for me to verbalize exactly what that [box and whisker plot] represents. And I like how I can look at it and say "okay that's cool" but to try and make descriptions about what that represents is difficult.

Mia: With stem and leaf plots you can see gaps and bumps in the data. While it looks nice on the box and whisker to see where medians lie against the boxes, is difficult to see what numbers are in between the quartiles and see how the distribution looks.

Barry: Even though in box and whisker you can see that the quartile is large you can't see why. When you look at the stem and leaf you can see it is because 60 has the bulge. With the double stem and leaf you can see there is a bump coming out that skews the data one way or another ..

While increased attention to the distributional features of data is a welcome finding, examination of the specific groups who used representations reveals a disconcerting pattern. Five of these groups had used graphs in the bean experiment, while the sixth group had used means. Two of the three groups who relied exclusively on descriptive statistics at the end of the study to make comparisons had not used graphs to represent the bean data in the initial weeks of the study. This finding indicates the development of a broader understanding of the functionality of graphical representations, but only for those who were already inclined to use such representations. Those who set out using descriptive statistics exclusively demonstrated stability in strategy use. Thus we were successful in helping participants understand the differential limitations and advantages of particular graphical representations, as indicated in their reflections at the end of semester; however the stability in strategy use for those who used descriptive statistics indicates that for these participants we were less successful in communicating the functionality of graphical representations as exploratory data analysis tools.

Figure 6 shows the data comparison strategy used by Group A who had used similar descriptive statistics to support their argument in the bean investigation (see Figure 2).

The following transcript is Group A's reasoning for *not* constructing a graph:

Valerie: Cause the question was more about distance and you can't compare individual kernels, we weren't concerned with how they clustered. That's why we didn't do a graph. It was more in terms of the average distance that we looked at it. We didn't do the graph .. cause frankly the double stem and leaf is very messy for me. I understand you can see the bell curve but em it is very … it is too much. I'd prefer more concise data and more gearing towards the average. If it had asked perhaps how would you show the data or …

Robyn: How would you visually represent it?

Anne: Right.

Valerie: That's why we chose average cause the question wasn't so specific.

Closer examination of the dialogue indicates that participants in Group A may have interpreted the Popcorn task as estimating the average distance from the refrigerated to the non-refrigerated kernels. This involves a comparison of averages and does not require the construction of a graphical representation of the data. In this case, participants may have decided against constructing a graph as it did not represent a useful or efficient strategy in the context of this particular task.

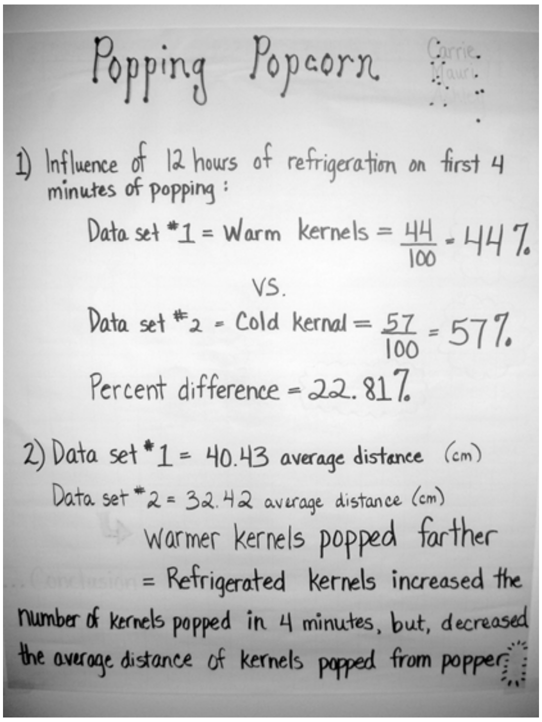

Figure 6. The use of descriptive statistics to describe a distribution of data

Examination of the reports accompanying the popcorn experiment reveals one common theme underlying the choice of descriptive measures: the ability of descriptive statistics to include all data values in the comparison. As Group B stated “averages were used because it is too difficult or rather impossible to compare individual kernels between the two [conditions].” This comment indicates the realization that unequal sample size would make a one-to-one comparison of values impossible, and in any event any one-to-one comparison of values would result in the loss of data (of the larger dataset). The upshot of this is twofold. On the positive side, participants understood the mean as a proportional measure appropriate to use in comparison of unequal size datasets. On the other hand it seems that the group was not prepared, or able, to examine global aspects of the distribution. Their comments indicate that a focus on individual data values would be the alternative had the datasets been equal in size. Thus it may be that this group was still focused on individual values in the datasets rather than possessing what Konold et al. (1997) term a statistical perspective.

When analyzing these groups’ responses across the semester we can see that participants have demonstrated growth in several areas. Firstly, they are now attuned to sample size, something that they did not consider in the bean experiment. Secondly, their choice of descriptive statistics is now grounded in an attempt to include all data values particularly given the realization that datasets are different in size. It seems that this newfound attention to sample size disparity provided greater support for use of the mean and median. Finally, two groups demonstrated some awareness of the limitations of their approaches. Following the group presentations, the act of revisiting their own analyses in light of the presentations (see Step 5d Appendix A, and Step 4b Appendix B) led them to reflect critically on the limitations of not providing a picture of the distribution. The following quotes highlight the discomfort that two groups felt with the data reduction that takes places with the use of descriptive measures.

Group A: After looking at the other groups' presentations we now see that with this method [using means to make group comparisons] we cannot determine if there are clusters of distances

Group C: We liked our method. But the limitation of it [using the mean, median and mode] is that it reduces all data into one number. If we had the opportunity we would re-do our analysis again because seeing other people's graphical representations, it is much easier to see all data represented and to see how much farther it popped in either scenario.

4.3. OTHER ISSUES IMPACTING UNDERSTANDING OF DISTRIBUTION

The mean as a proportional measure Despite focused instruction on the concept of the mean, Amanda posed a question on the last day of class that revealed her difficulty understanding the mean as a data comparative measure.

Jay: Okay then it looks like we'll just find the means of the refrigerated ones and the regular ones. Then we can see if there is any difference between the groups.

Amanda: But if we have different N's doesn't each one [popped corn] in the smaller set get more significance?

This comment indicates that Amanda is grappling with the issue of how to best compare two datasets of unequal sample size. What is evident at the same time is her difficulty understanding the mean as a proportional measure. At this juncture, the instructors recapped on a conversation held several weeks previously in which particular features of the mean were explored and reminded Amanda of the ways in which the mean deals with unequal sample size. Amanda's final comment indicates that she has accepted that the mean is a measure that can be used to compare datasets of unequal sample size however she is of the understanding that there may be some cut-off point representing magnitude in sample size difference when the mean is no longer an appropriate comparison measure.

Amanda*:* I wonder how big does the difference [in sample size] have to be before you can't use the mean to compare them?

The role of zero values when examining and describing distributions Uncertainty regarding whether zero values should be included in calculation of the means for particular conditions surfaced during the bean investigation. During the group presentations Andrew stated his group had "manipulated" the data so as to generate a finding that supported their original hypothesis. In an attempt to further probe this issue, the question of whether it was valid and justifiable not to count zeros was posed in the focus group discussion.

R: Andrew, you said you manipulated the data – how did you do it?

Andrew: We left out all data points that were zero – if we had included them it [the subgroup means] would have averaged out differently –

R: Did you make the decision ahead of time?

Maria: I didn't think it would matter if we counted [the zeros] cause the total number was smaller anyway.

R: The number you divided by didn't count the zeros either?
Maria: No.
R: Is it valid not to count the zeros?
Maria: It would have skewed the data so badly
Stan: I would go back to what the original question was that we wanted to ask. You may care about those that don't germinate. But there may be another scenario where it wouldn't matter.
Andrew: Yeah if you just wanted to find the longest one then yeah those that don't grow don't matter.

We can see in the transcript a number of understandings related to zero. Andrew understands that inclusion of the zeros will change the outcomes of the comparison. Whereas Maria's statement implies the initial belief that not counting zero values would not matter given that the denominator of the mean algorithm adjusts to reflect the number of values being considered. We see, however, that Maria's second comment indicates a change in her thinking. Her comment on skew indicates that the inclusion of zeros will be influential. It was not until Stan commented on the meaning of zero within the context of the research question that the groups' consideration of zero shifted from thinking of zero merely as a number devoid of meaning to consideration of zero as a measure within the context of the investigation.

Examination of the work carried out by groups in both statistical investigations shows that, in general, values of zero were eliminated and not considered as valid data. This has repercussions for reasoning about distributions in that within the context of statistical investigations a measure of zero is a value and its consideration, whether graphical or quantitative, influences the outcome of deliberations. While we never probed the reason behind why zero values were considered inconsequential by a large proportion of the participants, two explanations come to mind. The first hypothesis suggests that a conflation of one type of mathematical understanding of zero and the experimental situation resulted in zero not being considered. From a mathematical perspective participants may have been considering zero as representing the absence of elements, that is, a set of zero objects. In the context of the bean and popcorn experiments, beans that didn't grow and kernels that didn't pop were assigned the value zero, however participants may not have considered the value zero as a quantity but rather as the absence of growth or distance, if a bean didn't grow or a kernel didn't pop then it has no measure and shouldn't be considered. The second hypothesis suggests that participants' actions resulted from over generalizing the property of zero as the identity element in addition. This notion, that zero as an identity element leaves a set unchanged, is true for addition but is not true in the context of the mean. While the numerator of the mean is an addition context, the quantity derived from the addition is then divided by another quantity – in this case zero as the identity element does not hold.

5. DISCUSSION

The first goal of the study was to identify the statistical concepts preservice teachers focus on when analyzing and comparing distributions of data. The findings of this study suggest that elementary preservice teachers' focus is on *summary* rather than *exploration* of datasets resulting in a focus on summary statistics such as measures of central tendency. This overemphasis on centers and corresponding neglect of variation has also been highlighted by Shaughnessy (1992, 1997) in the undergraduate student population. The focus on summary was also evident in that participants did not use graphical representations to support the description and comparison of distributions of data. It seems that when presented with a distribution of data participants did not attempt to represent the distribution in a way that highlighted structural features, thus limiting the accuracy of the conclusions that can be drawn from the data. Shifting attention to exploration *prior* to summarization was not an insurmountable task. Once participants were made aware of the merits of data exploration and their attention drawn to distributional shape, they were eager to utilize a number of alternative measures (for example variation) and representations when comparing and analyzing datasets. Our finding that the increased attention to graphs resulted in revealing aspects of distributions, supporting the findings of Hammerman and Rubin (2004) and Makar and Confrey (2005) who noted a particular emphasis on variability.

The second goal of the study was to examine the ways in which different understandings of these particular statistical concepts support or hinder the description, analysis, and comparison of datasets. As mentioned in the previous paragraph, the overemphasis on measures of central tendency went hand in hand with the neglect of graphs and variability. For many participants, their *lack of exposure* to statistical ideas and statistical inquiry lead to the blanket implementation of measures they were familiar with – the mean invariably. However, once participants' attention was drawn to variation, in concert with an emphasis on how variation is modeled in graphical representations, variation quickly became a central component of participants' understanding of distribution. Similarly, providing experiences which highlighted the functionality of graphical representations, as tools to explore and reveal aspects of distributions, supported a focus on graphs. This resulted in a concentration on the selection of particular representations according to their propensity to reveal features of distributions.

Finally, this research reveals a number of ways to support the development of rich understandings of distribution. Firstly, it is critical that we draw preservice teachers' attention to the notion of distribution; many participants did not hold a distributional view of data. The use of the experimental context supported the construction of distributional perspectives due to the emphasis drawn by the context on the variation of data values along a scale of measurement (i.e., how height of the bean varied within the range of possible heights). Once this notion of distribution was established participants could see the interrelationships between measures of center and variability and the underlying structures of data that they emulate, and recognize the importance of graphs in revealing aspects of data. Secondly, a focus on the dataset as an entity was essential as it provided a meaning and context for the construction of representative values (see also Mokros & Russell, 1995) – values that were initially applied without an underlying rationale. The act of comparing datasets forced the entity view in that the act of comparison required the search for comparison values, each of which needed to be representative of the body of data. Finally, once the notion of distribution is established and the concept of dataset as an entity developed, understandings of distribution can be further nurtured though exposure to the range of measures and representations that support the continued effort of describing, analyzing and comparing distributions. It was surprising to find that many preservice teachers were not adept at constructing, or in some cases even aware of, stem and leaf plots and box and whisker plots. This lack of experience as learners with representations and measures that they may be required to teach in the future is a worrying, yet not surprising, finding as highlighted by Lajoie & Romberg (1998) in their call for teachers to be provided with content experiences in data analysis and statistics. In essence, the study highlights that once provided the *opportunity* to engage in statistical inquiry in conjunction with instruction focusing on data analytic techniques, preservice elementary teachers develop understanding of statistical measures and techniques and utilize them in statistically sound and justifiable ways.

5.1. LIMITATIONS

There are a number of limitations of this study which are related primarily to the research design and the nature of the experiences presented in the class. When considering the research design, the research participants are not representative of the general population of preservice teachers. Given their undergraduate degrees this study may overestimate the mathematical content knowledge of the general population of elementary preservice teachers. Other design limitations, related to the complexities of carrying out research in educational settings, carry with them implications related to the study validity. For example, there not was random assignment of participants to the study and there was not a control group. These factors pose threats in terms of the internal validity of the findings.

5.2. IMPLICATIONS

One challenge prospective teachers face is examining the structure of their own mathematical knowledge in an effort to come to know what it means to understand a particular concept. So when thinking about teaching statistics a preservice teacher may ask what it means for a 6^{th} grade student to understand the mean. This requires examining what it is we know about the mean, how we came to know "it," what "it" contributes to our understanding, and how we might embark on supporting students in developing similar understandings. So, we might agree, if asked, that the mean is a value that is not necessarily represented in the dataset. But if asked to list all we know about the mean we may not list this particular piece of knowledge – so how do we know what we know? How do we start to decompress (Ball & Bass, 2000) our mathematical knowledge? These types of understandings are what Ball and Bass (2000) call knowledge packages and Ma (1999) calls knowledge bundles – they are the fundamental understandings, connections, ideas that teachers need to develop (or may already possess) so that our knowledge becomes more accessible as a resource for teaching. This study provides insights into the ways in which engaging in statistical inquiry, wherein one is accountable for justifying the tools used to explore a distribution of data, challenges learners to question what it is they know and how this knowledge can be used in ways that are mathematically sound and justifiable. In their current form, traditional methods of teaching the pedagogy of data analysis and statistics fail to engage prospective teachers in examining their own knowledge for teaching.

This study also suggests that when considering the mathematical preparation of teachers we cannot assume that preservice teachers have sufficient exposure to statistical measures and techniques, in particular the construction and selection of appropriate graphical representations. Even when preservice teachers demonstrate knowledge of how to generate particular measures; this study shows that they may not recognize situations in which to use these measures, their understandings were what Skemp (1979) would categorize as primarily instrumental with poor relational understandings. It seems that participants did not have an adequate picture of the landscape of data analysis – in other words, what measures are available, when we might use them, and why. These findings suggest strongly that when considering the mathematical preparation of teachers that the focus be placed on what Hiebert and Lefevre (1986) define as knowing *how-to* (procedural understanding) and *why* (conceptual understanding) we use particular statistical techniques and representations.

There are a number of ways in which future research relating to preservice teacher statistical knowledge can build on and extend the findings of this study. It is important to investigate whether similar studies using different types of statistical investigations result in similar conclusions as this study. This would have implications not only for the ways in which we research preservice teacher's statistical understandings but also for the ways in which we structure our instruction in elementary schools. An extension of this work would be to investigate the ways in which providing experiences in data modeling at the preservice level influences the teaching practices of prospective teachers when they enter classrooms.

5.3. RECOMMENDATIONS

The final paragraph of the discussion poses recommendations for ways to support the development of rich understandings relating to distribution. The remainder of this section poses recommendations for ways to support learners build conceptual understanding and skills.

Firstly, while all students seemed to be good consumers of statistical information, in other words they demonstrated skills in graphical interpretation and comprehension (reading the data, reading between the data, and reading beyond the data), the majority exhibited difficulties constructing graphical representations. Efforts should be made to provide preservice teachers opportunities to work with real data and engage in activities related to constructing graphical representations. Secondly, of those students who demonstrated skills in deriving descriptive statistics and constructing graphical representations, relational understanding of these measures was absent. For example, some participants wished to construct scatter plots of the univariate data and persisted in trying to manipulate the data so that it would be amenable to presentation on a scatter plot. This again relates to their primary experiences as consumers of statistics – they have not been in the position of having to select appropriate statistics and representations for particular purposes. This finding calls for a coordinated effort to provide experiences that allow preservice teachers to consider the appropriateness of particular measures and representations for the purposes of data analysis. Thirdly, and not unrelated to the previous points, was the poor conceptual understanding of descriptive statistics and graphical representations, calling for a more conceptual focus in mathematics and methods courses. It also cautions teacher educators against drawing conclusions about conceptual understandings based on demonstrated proficiency in generating and applying measures (such as the mean for data comparison purposes). Finally, our observations lead to the conjecture that gains in understanding demonstrated over the course of the semester were influenced primarily by having access to strategies used by peers when engaging in data description and comparison activities. While classroom teaching experiences supported the development of skills and conceptual understanding, what they seemed not to do was convince participants of the utility of such measures when engaged in data analysis. In other words, our analysis indicated that factors other than classroom teaching were more influential in convincing students to apply new concepts and skills when engaging in statistical inquiry. It became apparent that engagement in small group statistical inquiry acted as a conduit whereby prospective teachers observed and gained access to the complex decision making processes of others when engaged in exploratory data analysis and then compared those decisions against their own. Such experiences provide opportunities for participants to learn in practice, to develop communities of learners who engage in authentic statistical inquiry, and who continuously seek to find more efficiently and statistically justifiable ways of thinking about distributions of data.

ACKNOWLEDGEMENTS

I would like to acknowledge the contribution of Christy Graybeal, Toni Smith and Carolina Napp to the design and implementation of this research.

REFERENCES

Ball, D. L., & Bass, H. (2000). Interweaving content and pedagogy in teaching and learning to teach: Knowing and using mathematics. In J. Boaler (Ed.), *Multiple perspectives on mathematics teaching and learning* (pp. 83-104). Westport, CT: Abex.

Ball, D. L., & McDiarmid, G. W. (1990). The subject matter preparation of teachers. In W. R. Houston, M. Haberman, & J. Sikula (Eds.), *Handbook of research on teacher education* (pp. 437-449). New York: Macmillan.

Ben-Zvi, D. (2003). The emergence of reasoning about variability in comparing distributions: A case study of two seventh grade students. In C. Lee (Ed.), *Proceedings of the Third International Forum on Statistical Reasoning, Thinking, and Literacy (SRTL-3).* [CD ROM with video segments]. Mount Pleasant, Michigan: Central Michigan University.

Borko, H., Eisenhart, M., Brown, C. A., Underhill, R. G., Jones, D., & Agard, P. C. (1992). Learning to teach hard mathematics: Do novice teachers and their instructors give up too easily? *Journal for Research in Mathematics Education*, *23*(3), 194-222.

Cai, J. (2000). Understanding and representing the arithmetic averaging algorithm: An analysis and comparison of US and Chinese students' responses. *International Journal of Mathematical Education in Science and Technology*, *31*(6), 839-855.

Canada, D. (2004). *Preservice teachers' understanding of variation.* Unpublished doctoral dissertation, Portland State University.
[Online: www.stat.auckland.ac.nz/~iase/publications/dissertations/dissertations.php]

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

Cobb, P., Gravemeijer, K. P. E., Bowers, J., & McClain, K. (1997). *Statistical Minitools.* Designed for Vanderbilt University, TN, USA. Programmed and revised (2001) at the Freudenthal Institute, Utrecht University, The Netherlands.

Friel, S. N., Curcio, F. R., & Bright, G. W. (2001). Making sense of graphs: Critical factors influencing comprehension and instructional implications. *Journal for Research in Mathematics Education*, *32*, 124–158.

Friel, S. N., Mokros, J. R., & Russell, S. J. (1992). *Used numbers. Statistics: Middles, means, and in-betweens.* Palo Alto, CA: Dale Seymour Publications.

Gfeller, M. K., Niess, M. L., & Lederman, N. G. (1999). Preservice teachers' use of multiple representations in solving arithmetic mean problems. *School Science and Mathematics, 99*(5), 250-257.

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/SERJ3(2).pdf]

Hancock, C. (1995). *TableTop Software.* Novato, CA: Broderbund Software Direct.

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

Heaton, R. M., & Mickelson, W. T. (2002). The learning and teaching of statistical investigation in teaching and teacher education. *Journal of Mathematics Teacher Education, 5*(1), 35-59.

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

Konold, C. (1998). *Tinkerplots: Tools and curricula for enhancing data analysis in the middle school.* Grant proposal submitted to the National Science Foundation (# ESI-9818946). University of Massachusetts, Amherst.

Konold, C., & Miller, C. (2001). *Tinkerplots* (version 0.23). Data Analysis Software. University of Massachusetts, Amherst.

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

Konold, C., Pollatsek, A., Well, A., & Gagnon, A. (1997). Students analyzing data: Research of critical barriers. In J. Garfield & G. Burrill (Eds.), *Research on the Role of Technology in Teaching and Learning Statistics: 1996 Proceedings of the 1996 IASE Round Table Conference* (pp. 151-167). Voorburg, The Netherlands: International Statistical Institute.
[Online: www.stat.auckland.ac.nz/~iase/publications/8/13.Konold.pdf]

Lajoie, S. P. (1997). The use of technology for modelling performance standards in statistics. In J. Garfield & G. Burrill (Eds.), *Research on the Role of Technology in Teaching and Learning Statistics* (pp. 57-70). Voorburg, The Netherlands: International Statistical Institute.
[Online: www.stat.auckland.ac.nz/~iase/publications/8/5.Lajoie.pdf]

Lajoie, S. P., & Romberg, T. A. (1998). Identifying an agenda for statistics instruction and assessment in K–12. In S. P. Lajoie (Ed.), *Reflections on statistics: Learning, teaching, and assessment in grades K–12* (pp. vii–xxi). Mahwah, NJ: Lawrence Erlbaum Associates.

Lappan, G., Fey, J. T., Fitzgerald, W. F., Friel, S. N., & Phillips, E. D., (2002). *Connected Mathematics Project*. Needham, MA: Prentice Hall.

Leavy, A. M., & O'Loughlin, N. (2006). Moving beyond the arithmetic average: Preservice teachers understanding of the mean. *Journal of Mathematics Teacher Education, 9*(1), 53-90.

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

Lesh, R., & Lehrer, R. (2000). Iterative refinement cycles for videotape analyses of conceptual change. In A. Kelly & R. Lesh (Eds.), *Handbook of research design in mathematics and science education*. (pp. 665-708). Mahwah, NJ: Lawrence Earlbaum Associates.

Lubienski, S. T., & Bowen, A. (2000). Who's counting? A survey of mathematics education research 1982-1998. *Journal for Research in Mathematics Education, 31*(5), 626-633.

Ma, L. (1999). Knowing and teaching elementary mathematics: Teachers' understanding of fundamental mathematics in China and the United States. Mahwah, NJ: Lawrence Erlbaum Associates.

Makar, K., & Confrey, J. (2002, July). Comparing two distributions: Investigating secondary teachers' statistical thinking. Paper presented at the Sixth International Conference on Teaching Statistics, Cape Town, South Africa.
[Online: www.stat.auckland.ac.nz/~iase/publications/1/10_18_ma.pdf]

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/SERJ4(1).pdf]

Mathematical Sciences Advisory Committee, 1984-1985. Academic preparation in mathematics: Teaching for transition from high school to college. Cited in: Beres, R.J. (1988). Statistics for college bound students: Are the secondary schools responding? *School Science and Mathematics, 88*(3), 200-209.

McGatha, M., Cobb, P., & McClain, K. (2002). An analysis of students' initial statistical: Developing a conjectured learning trajectory. *Journal of Mathematical Behavior, 21*, 339-335.

Mevarech, Z. R. (1983). A deep structure model of students' statistical misconceptions. *Educational Studies in Mathematics, 14*, 415-429.

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

National Center for Research in Mathematical Sciences at the University of Wisconsin/Madison and Freudenthal Institute at the University of Utrecht (1997-1998). *Mathematics in context*. Chicago: Encyclopedia Britannica.

National Council of Teachers of Mathematics Commission on Standards for School Mathematics (1989). *Curriculum and evaluation standards for school mathematics.* Reston, VA: Author.

Pollatsek, S. J., Lima, S., & Well, A. D. (1981). Concept or understanding: Students' understanding of the mean. *Educational Studies in Mathematics*, *12*, 191-204.

Shaughnessy, J. M. (1992). Research in probability and statistics: Reflections and directions. In D. A. Grouws (Ed.), *Handbook of research on mathematical teaching and learning* (pp. 465-494). New York: McMillan.

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* (Vol. 1, pp. 6-22). Waikato, New Zealand: Mathematics Education Research Group of Australasia.

Simon, M. A. (2000). Research on the development of mathematics teachers: The teacher development experiment. In A. Kelly & R. Lesh (Eds.), *Handbook of research design in mathematics and science education* (pp. 335-361). Mahwah, NJ: Lawrence Earlbaum Associates.

Skemp, R. R. (1979). *Intelligence, learning and action: A new model for the theory and practice in education*. New York: Wiley.

Steffe, L. P., & Thompson, P. W. (2000). Teaching experiment methodology: Underlying principles and essential elements. In A. Kelly & R. Lesh (Eds.), *Handbook of research design in mathematics and science education* (pp. 267-306). Mahwah, NJ: Lawrence Earlbaum Associates.

Stein, M. K., Baxter, J., & Leinhardt, G. (1990). Subject-matter knowledge and elementary instruction: A case from functions and graphing. *American Educational Research Journal, 27*(4), 639-663.

Strauss, S., & Bichler, E. (1988). The development of children's concept of the arithmetic average. *Journal for Research in Mathematics Education, 19*(1), 64-80.

Technical Education Research Centers and Lesley College (1989). *Used numbers: Real data in the classroom*. Menlo Park, CA: Dale Seymour Publications.

TERC (1998). *Investigations in number, data, and space.* Menlo Park, CA: Dale Seymour Publications.

Tufte, E. R. (1983). *The visual display of quantitative information.* Cheshire: Graphics Press.

Watson, J. M., & Kelly, B. A. (2002, August). *Can grade 3 students learn about variation?* Paper presented at The Sixth International Conference on Teaching Statistics, Cape Town, South Africa.
[Online: www.stat.auckland.ac.nz/~iase/publications/1/2a1_wats.pdf]

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

AISLING LEAVY
R124 Mary Immaculate College, South Circular Road
University of Limerick
Limerick, Ireland

APPENDIX A: SMALL GROUP REPORT GUIDELINES FOR THE BEAN INVESTIGATION

THE BEAN EXPERIMENT: SMALL GROUP ACTIVITY

Group members: ____

Step 1: Re-examine the primary research scenario and experiment design

> Ben likes bean sprouts in his salads and sandwiches. Lately he has been unhappy with the quality of bean sprouts available at the grocery store so he has decided to grow his own. He suspects that lighting conditions and the addition of lemon juice to the water may affect the length of the bean sprouts. Ben wants to grow the longest bean sprouts possible. Briefly describe how he could determine which growing conditions (Light vs. Dark, Plain water vs. Water with lemon juice) support the best growth.

		Solution	
		Water	Lemon
Light Intensity	Light	A	B
	Dark	C	D

Step 2: Construct research questions you may examine in an effort to investigate Ben's suspicion.

Step 3: What comparison method(s) would you utilize to examine the data *and consequently answer your questions.*

Step 4: Construct a hypothesis describing what you will believe will be the outcome of the data analysis.

Now use the data from the bean sprouts to answer your question. Prepare a poster to present your analysis and be prepared to discuss the questions below.

Step 5:

a) Explain why you chose your particular comparison method(s):

b) Explain what (if any) limitations there are of this method(s):

c) Present the answer to your research question. How did you come to this conclusion?

Time to reflect!

d) You have now viewed the presentations and approaches your peers took when approaching the same task. What would you do if you had the opportunity to complete your analysis again – would you change your approach?

If so, why? And what would you do differently?

APPENDIX B: SMALL GROUP REPORT GUIDELINES FOR THE POPCORN INVESTIGATION

USING AN EXPERIMENT TO TEST A CONJECTURE – IT'S ALL ABOUT POPCORN!
SMALL GROUP ACTIVITY

Group members: ____

There are some people who say that refrigerating popcorn kernels prior to popping changes certain characteristics of the kernels. In this experiment we are going to examine one such conjecture.

Step 1: Make a list of factors that you think could be influenced by refrigeration.

Today, we are going to investigate:

> *Does refrigerating corn for 12 hours prior to popping influence either (a) the number of corn that pop over a 4 minute period, or (b) the distance that the corn falls/jumps from the popper?*

Step 2: Record the data for the first experiment: popping kernels for 4 minutes in an uncovered popper in the space below:

Step 3: Record the data for the second experiment: popping refrigerated kernels for 4 minutes in an uncovered popper in the space below:

Step 4:
Now use the data from both experiments to answer the research question. Prepare a poster to present your analysis and be prepared to discuss the questions below.

a) Explain why you chose your particular comparison method(s):

b) Explain what (if any) limitations there are of this method(s):

c) Present the answer to the research question. How did you come to this conclusion?

Time to reflect!

d) You have now viewed the presentations and approaches your peers took when approaching the same task. What would you do if you had the opportunity to complete your analysis again – would you change your approach?

If so, why? And what would you do differently?