
Gifted and Talented International

ISSN: 1533-2276 (Print) 2470-9565 (Online) Journal homepage: http://www.tandfonline.com/loi/ugti20

# Gifted Students' Understanding of Statistics: Analysis of Data Arising from a Small Group Teaching Experiment

Aisling M. Leavy

**To cite this article:** Aisling M. Leavy (2003) Gifted Students' Understanding of Statistics: Analysis of Data Arising from a Small Group Teaching Experiment, Gifted and Talented International, 18:1, 17-26, DOI: 10.1080/15332276.2003.11673004

**To link to this article:** https://doi.org/10.1080/15332276.2003.11673004

# Gifted Students' Understanding of Statistics: Analysis of Data Arising From a Small Group Teaching Experiment

AISLING M. LEAVY
MARY IMMACULATE COLLEGE
UNIVERSITY OF LIMERICK
LIMERICK, IRELAND

*ABSTRACT*

*The principal objective of the study was to gain insight into cognitive processes related to three gifted students' understanding of data. A small group teaching experiment was used to uncover students' understanding of the notion of distribution and consisted of a clinical interview phase, teaching phase, and analysis phase. Fifteen teaching episodes followed the initial clinical interview and involved students working together on a variety of mathematical activities. The study examined how students organized data, described and indexed distributions of data, in addition to investigating their conceptualization of the mean. Results suggest that (a) modes and ranges were the most salient features of distributions, (b) typical values were generally located in clusters of data, (c) the mean was not utilized in the constructed measures of typicality, and (d) students had a relatively sophisticated understanding of samples and populations. Implications for curricular treatment of statistics for gifted students are discussed.*

## Introduction

Mathematics instruction for gifted students has often limited students to a battery of standardized tests, advanced placement programs and adherence to selected textbooks. This has restricted the opportunities for gifted students to experience mathematics as a creative and divergent discipline and further reinforces the conventional view of mathematics as a well-structured discipline where answers are either correct or incorrect (Resnick, 1989). A variety of studies have advocated multiple approaches in the education of gifted youths. Arguments have been advanced for an emphasis on individualization (Coutant, 1983; Grant & Piechowski, 1999; Hafenstein & Tucker, 1995; Stein & Poole, 1997), developing problem-solving abilities, stimulating originality, initiative, and self-direction (Rimm, 1990), and investigating real problems and solutions (Dewey, 1938; Olenchak, 1998; Stein & Poole, 1997).

Within mathematics, the domain of statistics offers currently unrealized learning opportunities for gifted students as statistics education has the potential to incorporate many of the features advocated for inclusion in gifted curricula. Examination of the nature and structure of the most recent statistics curricular materials indicates a move towards treating statistics education as an ill-structured discipline (Lajoie, Jacobs, & Lavigne, 1995) requiring the use of critical and creative thinking and an inquiry-based approach to mathematics. Researchers emphasize the importance of students being involved in collecting and analyzing data that will be used for further analyses (Cockcroft, 1982; Curcio, 1987; Thompson, 1999; Tukey, 1977). In gifted education, real problems imply that students take on the role of firsthand inquirers (Reis & Renzulli, 1985), and involvement in analyzing their own data results in students arriving at recommendations that have real consequences. Lehrer & Romberg (1996) refer to this approach as data modeling and describe it as "the construction and use of data".

A major body of research in statistics education focuses on children's and adults' ability to understand particular statistical concepts related to measures of central tendency and/or variability (Leavy & Middleton, 2001; Leavy & O'Loughlin, 2002; Mevarech, 1983; Mokros & Russell, 1995; Pollatsek, Lima & Well, 1981; Strauss & Bichler, 1988). While distributions of data are described using concepts of central tendency, variability, skew and kurtosis, research has focused predominantly on students' understandings of measures of central tendency, in particular the mean. Recent studies in statistics have advanced knowledge relating to undergraduate students' (Leavy & O'Loughlin, 2002; Mevarech, 1983) and children's understanding of the mean (Cai, 2000; Mokros & Russell, 1995; Pollatsek, Lima & Well, 1981), children's notions of the mean as a conceptual or computational act (Gfeller, Niess & Lederman, 1999; Pollatsek, Lima & Well, 1981; Zawojewski, 1988) and the development of children's concepts of the arithmetic average (Strauss & Bichler, 1988; Watson & Moritz, 2000). What is absent from the research base is a coherent model of statistics learning that serves to explain the development of gifted children's understanding of the indices of distribution, and how these measures are related to each other and to the notion of distribution. In addition, little is known of the processes involved in gifted children undertaking data modeling projects. Knowledge of these processes would inform the development of curricular materials in statistics for gifted children in addition to providing insights into the potentialities of statistics education for non-gifted children.

The present research uncovered mechanisms of development that explain how gifted children come to understand the important concepts related to distribution and the decision-making processes used to choose appropriate indices of a given distribution. The establishment of such a model is critical because without a framework for interpreting students' statistical understanding, the resulting knowledge base will remain disjointed, focusing merely on refining our understanding of specific areas of statistics but providing no coherent picture of how individual descriptive statistics contribute to gifted children's notion of distribution.

*Mathematical treatment of distribution*

In recent years, elementary statistics education has moved from a focus on teaching separate descriptive statistics to a more holistic approach that examines distributions of data. Distribution is the "statistical term for the arrangement of the observations along the scale of measurement" (Hardyck & Petrinovich, 1969). If the data from an assortment of observations are plotted, there result a variety of differently shaped curves. A mathematical treatment of distribution entails the use of summary characteristics of the distribution. These are the moments of the distribution described by Hays (1963) as the "expectations of different powers of the random variable". Moments are used to describe the distributions of real random variables. The first moment about the origin of a random variable X is E(X) = mean (a measure of central tendency). The second moment about the origin is $E(X)^2$ and if the mean is subtracted from X prior to raising the variable to the second power then variance $E[X - E(X)]^2$ is considered the second moment. The third moment $E[X - E(X)]^3$ is considered to be a measure of skew, and the fourth moment the kurtosis or peakedness of the distribution. If the entire set of moments (central tendency, variance, skewness and kurtosis) are taken into consideration, the distribution can be described exactly.

At the elementary school level, the focus of attention has predominantly been placed on measures of central tendency. Measures of central tendency, however, do not completely describe a data set. Averages provide a measure of one feature of the data set – the center. To adequately describe a distribution, other indices such as measures of variability are required. Within the context of elementary school, students are not generally introduced to the concept of variability as an additional means to describe distributions of data. There are few studies, if any, carried out on children's attempts to provide accurate descriptions of a set of data. No investigation has been carried out of children's self-constructed measures of variability or spread. No one has investigated if children recognize the need for depictions of variability to supplement measures of central tendency when describing data sets. This study sets out to investigate gifted students' treatment of distributions.

This study examined gifted students' understanding of statistics through administration of tasks designed to probe the following questions:

1. What approaches do gifted students use to organize a set of data?
2. What methods do gifted students use to describe and summarize a set of data?
3. What conceptualizations do gifted students have of "the mean"?

## Method

*Participants*

Three students were involved in the study, each of whom was attending a 3-week mathematics course for academically precocious children at a major South Western university in Ireland. Two of the students were female and had just completed third grade; the other student was male and had completed fifth grade.

*Methodology*

Teaching experiment methodology was used to uncover children's conceptual understanding of the notion of distribution. Steffe & Thompson (2000) state, "the primary purpose for using teaching experiment methodology is for researchers to experience, firsthand, students' mathematical learning and reasoning." Teaching experiment methodology is used in the field of mathematics to assess the learning trajectories of students engaged in mathematics and to construct models of students' mathematics. Involved were a sequence of teaching episodes that served to allow the observation of restructuring in mathematical schemes. Teaching experiment methodology consisted of a clinical interview phase, a teaching phase and an analysis phase.

*Clinical interview*

A clinical interview was administered to each of the students prior to and following the teaching episodes. The purpose of administering the clinical interview prior to starting the teaching episodes was to identify the experiences that students had had in statistics. Due to the interrelatedness of many areas of statistics, for example, the need for children to be able to interpret graphs before they can calculate measures of central tendency on data presented to them in graphical form, the information derived from the clinical interview was essential to the progress of the study. This information provided a baseline to anchor the development of individual children's learning trajectory. In addition, a clinical interview was administered following the teaching episodes. The purpose of this interview was to assess growth and aid in constructing a model of the student's statistical reasoning following the teaching experiment. Both clinical interviews were identical in structure and form i.e. the same tasks were used in both.

Clinical interviews consisted of eleven tasks, lasted approximately 60 minutes and were video taped. Tasks A-C sought to develop a picture of the extent to which students were familiar with graphical methods. Specifically, they assessed the ability of students to construct and interpret bar graphs, line plots, and stem-and-leaf plots. Task D investigated the features of and factors affecting students' self-constructed typical values of a data set. Tasks E and F investigated students' ability to calculate and differentiate between measures of central tendency i.e. the mean, median and mode. Task G investigated the strategies children use to compare sets of data, with a view to examining whether children consider the mean as a measure that can be used in comparison across sets of data. Tasks H-J examined children's understanding related to the mean, in particular they identified whether children understood certain features related to the mean, such as the value of the mean not being higher than any value in the data set it indexes. All of tasks H-J were similar to those developed by Strauss and Bichler (1988). Task H was developed to assess how students understood that the average is located between the extreme values in the data set. Task I investigated the ways in which students came to understand that the average is influenced by values other than the average, and task J evaluates students' understanding that the average value is representative of the values that were averaged. Task K investigated whether students understood the need for additional indices of distribution that supplement measures of central tendency in order to adequately describe a data set.

*Teaching episodes*

Teaching episodes took place after the initial clinical interview and consisted of the three students working on a variety of activities. Fifteen teaching episodes lasting approximately 30-45 minutes each were carried out with the small student group; each episode was video taped. Small group teaching experiment methodology was implemented as small group structure was a common pedagogical practice. These episodes were used to construct models of the students' mathematical thinking and guide children to develop more sophisticated ways of reasoning about data.

The initial teaching episode (Episode 1) was used to further probe the results arising from the initial clinical interview and to test hypotheses arising from the interview. Teaching episodes 2-3 then focused on data collection and organization. Students collected sets of raw data and were required to reorganize the data in some way they could present it to others. These episodes were used to examine the strategies students might use to organize the data, identify strategies that might be useful to students, and facilitate the incorporation of these strategies into their conceptual schema. Two episodes (4-5) were devoted to providing instruction on graph construction; in particular the construction of stem-and-leaf plots and line plots. This was followed by teaching episodes 6-9 devoted to describing and summarizing distributions. These episodes evaluated how students described sets of raw and graphical data and investigated the strategies utilized by children to summarize data. Specifically, the intent was to investigate the features of the data to which students attended when describing the data sets and when constructing typical values. Errors and weaknesses in children's reasoning were identified, and tasks were designed to alert children to alternative and additional strategies. This was facilitated by constructing data sets that incorporated a range of features of interest in the study: outliers, modes, and large variability in data scores. Episode 10 provided experiences in comparing data sets, both raw and graphical. Further episodes (11-13) were designed to help children investigate the features of the mean and develop conceptual understanding of the mean. Children were engaged in balancing distributions around mean values and in constructing distributions around a mean. The final episodes (14-15) pulled together the components taught and focused on helping children integrate their learning experiences. Other unexpected issues that had arisen and been discussed over the course of the teaching experiment (sampling, populations, ethical issues related to data collection etc.) were examined in the final teaching episodes.

*Tasks*

The tasks used in the study were from a variety of sources; the majority were developed by the instructor and others adapted from previously published research. These tasks were designed to elicit models of children's understanding of certain statistical phenomena. Tasks were categorized into groups, each group tapping into a specific area of study: group one tasks investigated children's approaches to organizing data, group two tasks were designed to draw out children's approaches to describing and summarizing data (through the construction of typical values), the third group of tasks evaluated the methods used by children to compare data sets, and the fourth group of tasks explored students' understanding of the mean.

In so far as possible, the data used in the investigation were collected by students. Investigation of children's conceptualizations of typical values involved students' determining typical values for the presented data distributions. The experimenter/teacher then manipulated data in the data set (adding outliers, replacing modal values, etc.) and confronted students with making sense of the newly manipulated data. Problems were presented in several forms (e.g. graphical, tabular, and word problems). Graphical representations generally took the form of line plots and stem-and-leaf plots. These representations were chosen as they facilitated the viewing of individual data values as opposed to other graphical representations in which data are collapsed (pie charts, box-and-whisker plots). Other activities involved presenting students with open-ended problem solving scenarios and engaging students in discussions related to the problem situations.

*Analysis*

The analysis phase involved examining the data gathered from the clinical interviews and the teaching episodes. Ongoing analysis occurred between teaching episodes and a retrospective analysis focused on the cumulative episodes. The main sources of data were the researcher's field notes, video records of the interactions, and samples of the student's written work.

Analysis of the data involved consolidating, reducing and making sense of the utterances and actions of each of the participants during each of the teaching episodes. Key phrases and statements arising from small group teaching episodes were identified, hypotheses relating to the participants' mathematical reasoning were made, and tasks were designed to test these hypotheses (emergent tasks). Emergent tasks were presented in the subsequent teaching episode, examination of the data arising from the emergent tasks involved seeking disconfirming as well as confirming evidence to reject or support the previously constructed hypotheses. When disconfirming evidence was identified the previous hypothesis was rejected and new hypotheses and tasks constructed. If confirming evidence was found, then tasks were designed to either further examine the participant's conceptualizations, or to challenge the students' reasoning in relation to the concept.

Ongoing analysis involved coding and analyzing the audio records of students' responses in the teaching episodes. The purpose was to create descriptions of students' mathematical behaviour and provide explanations of students' progress over time. These descriptions took the form of models that denoted the current understanding of children's mathematical realities. Different trajectories evolved as a result of the constraints of interacting with individual students. Students' contributions to the trajectory were noted and attempts made in task construction to promote the greatest possible progress in learning. Hence the teaching experiment methodology takes the form of a recursive cycle (Steffe & Thompson, 2000) in which hypotheses were formulated, tested, and reconstructed. The language and actions of students resulting from the continuous interaction with researchers served as data to confirm or disconfirm hypotheses.

Retrospective analysis of the record of the teaching experiment occurred after completion of the teaching experiment and was a critical part of teaching experiment methodology. Careful analyses of audio and videotapes activated records of experiences with students. These experiences were re-examined in a historical perspective and provided new insights not available to the researcher at the time the event occurred. Quite often changes in students' mathematical structures become apparent only in retrospect. Hence, the teaching episodes provided the means to generate in-depth knowledge of children's conceptual understanding of data and how it was fostered in a coherent instructional sequence.

## Results

Results of the study present implications for instructional approaches and the design of statistics curricula for gifted students. Results are structured according to the objectives of the primary task groupings.

*Approaches to organizing data*

The practice of organizing data to facilitate interpretation was not a process students entered into naturally. Throughout the study, there was a lack of awareness of the need to organize and categorize data. In cases where students were encouraged to categorize and organize data for display purposes, data were rearranged but only in one case did this rearrangement result in a summarization of the data. *The Spare Time Task*, developed by the McClain group at Vanderbilt, presents data on how a class of seventh graders spends their spare time. This task required students to reorganize the data (consisting of 24 activities) so that the results can be presented on a bulletin board. Examination of the transcript indicates that Jonathan and Tanya merely reorganized rather than summarized the data, their approaches reflecting a temporal rearrangement of the data. Jane was the only student who provided a summarization of the data, using a pie chart.

Researcher: How would you organize the information to make it easier to read?

Jane: I made a pie graph basically I put them into categories and my first one was something calm, basically it is when you are doing something calm. The second was watch TV, third was on the computer, my fourth was moving around, and my fifth was musical .. like something that had to do with music.

Researcher: Very good .. Tanya what did you do?

Tanya: I went by ones having something to do with the other like ..one was read .. 2 was read comic books .. then watch TV which is like comics on the TV .. then MTV and then 5 is watch reruns of the X-files .. then watch old movies and listen to the grateful dead .. which leads to listening to music and you exercise then cause music helps stimulate exercise, play with the dog which is probably my exercise .. play the piano .. because it is sort of relaxing .. 12 ..practice the guitar cause that has to do with music .. talk on the phone

Jane: some people call the phone an instrument

Tanya: then you go to 14 .. which is write letter, which is communication .. then you clean your room so you can find more letters and so on .. and then surf the internet because .. (shrugs shoulders) .. 17 is watch sports on TV .. well you've just surfed on the internet so you feel you want to watch TV then shoot basketball and then you visit with friends on e-mail .. just communicate with them .. and you know if you want to communicate with them .. then you jog to get out the cobwebs from watching TV and stuff.

Researcher: Jonathan, can you explain once again what you did?

Jonathan: So what did I do with all this stuff .. I .. wouldn't do the music category all at once .. and reading a lot so .. I first ..read and then watch some sports .. maybe some exercise .. and play the piano and talk on the phone and .. play with the dog, write letters and visit with e-mail, shoot the basketball .. then I'd jog .. I'd read and I'd watch MTV and listen to Grateful Dead, surf the net and clean up my room, watch reruns of X-files, watch old movies, practice guitar, listen to music, and last read comics.

Researcher: Did you put them in order of what you'd like to do or just ..

Jonathan: Well I kinda .. mainly what I like to do like .. I put the jog and exercise kinda far apart .. and go outside cause you need to go outside every once in a while .. I put them in between like I exercise and then go outside and play basketball and then jog so I go outside .. I put those in between.

What is interesting about the above transcript is that children are aware that the data fall into specific categories. They also demonstrated the ability to construct pie charts and bar charts in the initial clinical interviews. Hence it would seem that they have at their disposal the skills and prerequisites needed to summarize the data using a graphical representation. Nevertheless, two of the children did not consider using graphical representations in any of the tasks in this category of tasks.

*Approaches to describing and summarizing distributions of data*

When describing data sets, students tended to report a) data values that appeared at the highest (modal) and lowest frequencies, and b) the upper and lower values (range) of the distribution. This was illustrated in the *Raisin data* task, a task in which students were required to count the number of raisins in 27 boxes of raisins and graph the result on a line plot. The purpose of using the *Raisin data* task was to evaluate how students describe sets of raw data and to investigate the strategies utilized by children to summarize the data. Examination of a conversation arising from the *Raisin data* task illustrated the students' awareness of the frequency of occurrences of particular values, and Tanya's awareness of clusters and gaps in the data.

Researcher: I want you to look at the line plot on the board of our raisin data. I want you to tell me –what does it look like?
Tanya: Okay .. we have a thing then we have a double space then we have three things then we have a space and then we have three more things and then we have a space. Em, the most common was 32 and the least common was 24.
Researcher: Does anyone else see a pattern?
Jane: Well the least was 24, 27, and 28.
Researcher: Why do you say that Jane?
Jane: 'Cause they are all the same .. all have 2.
Researcher: Very good, Jonathan do you see anything?
Jonathan: Em, … all of them except the numbers 29 and 32 were all in the same range of 2, 3, and 4.

In addition, students referred to clusters of data and to instances of outlying data values. In the following transcript related to the *Paperclip II* activity, each student blew a paperclip on 10 occasions. The 30 resulting data values were entered on a line plot. Examination of the transcript highlights the children's awareness of clusters of data that occur within a distribution of data. Analysis of the data set and the transcript indicated that the children were aware that the majority of data values cluster in the range of 1-10 centimeters. Students also referred to the low frequency of occurrences of scores above 10 centimeters.

Researcher: What about the distribution. When you look at the data values - is there any type of pattern you can see? How would you describe it?
Jonathan: Em, there is a pattern that 1 to 10 they got a lot there compared to all the other places. Yeah the pattern is that there is a lot in the 1 to 10s.
Researcher: Jane, do you have an idea?
Jane: Well like most of us it was like from 1 to 10 and then it started getting above 20 it was only 1's and 2's.

The younger students were reluctant to provide summary measures of data sets. They were inclined toward providing a range of values within which the summary/typical value would fall. Student-constructed ranges tended not to encompass the entire data set. In data sets resembling a normal distribution, students usually chose a range that excluded the lowest and highest data values and contained the modal values of the data set. In situations where the greatest frequency of values lay at the upper or lower end of the distribution (skewed data sets), students provided ranges that enveloped the modal values lying on either end of the data set. This implies a predisposition on the part of younger students to utilize measures of variability rather than central tendency, and also suggests a concentration on modal values. The sixth grade student was, at times, more disposed to providing single summary measures of data sets; however, he predominantly quoted ranges. In the *Raisin data* task the children were asked to provide a single value that represented the number of raisins occurring in a box of raisins. Examination of Jane and Tanya's responses show their predisposition to providing ranges, compared to Jonathan's contribution of a single typical value.

Researcher: So, generally how many raisins do you think are in a box of raisins?
Jane: I think there's a range .. em .. of 31 to 33 raisins in that range .. in a box of raisins.
Researcher: Very good, why do you think that Jane?
Jane: Cause that's what I counted in each of the boxes .. it is kind of close .. the numbers 'cause they are going 31 and 32.
Researcher: Tanya, what do you think?
Tanya: I have an argument to that because in the sun maid raisins they were mostly in the 20s so I'd say 24 to 33 would be the amount.
Researcher: Okay Jonathan .. You are very quiet over there. I know you have a theory .. what do you think?
Jonathan: I em I think that also all the raisins vary from 33 all the way down to 28.
Researcher: Can you provide a typical value for the data set?
Jonathan: Like between 24 and 33.
Researcher: Between 24 and 33, okay very good. So what if I pick a box (of raisins) at random from a bag, Jane how many raisins do you think, based on the data from the board, how many raisins would be in the box that I pick?
Jane: Probably 27 to 33 or something like that.
Researcher: Why do you think that?

Jane: Because .. actually I ..because em .. actually I think 29 between 29 to 32 because 32 and 29 mostly had the most.
Researcher: Tanya, what do you think?
Tanya: Between 29 and 32 actually between 33 and 28 because 29 and 32 are in that category.
Researcher: Why is it important that 29 and 32 are in that category?
Tanya: Because they had the most results shown and they were the most common of all raisins.
Researcher: Jonathan, what about you?
Jonathan: Em, somewhere around 30
Researcher: Why do you pick 30?
Jonathan: Because 29 had a lot … and 32 is right below 30 … no 32 is above 30 and you only got 2 that were 24 and none that were 25 or 26. They are all in the range of 27 and higher. So 30 … yeah 30.

Surprisingly, on no occasion did a student calculate a mean value when requested to provide a typical value. The typical values chosen were predominantly modal values. In this sense typical did not embody a 'fair share' notion nor did it exemplify representativeness. The *Cat & Dog Weights* exercise was used to examine children's self-constructed measures of typicality, and to investigate the effect of extreme measures on these values. The children graphed the data values on a stem-and-leaf plot; the importance of the mode to the children in constructing measures of typicality was evident in the transcript.

It is interesting to note in this task that children demonstrated little reluctance in generating individual typical values when compared to their disposition to provide ranges in previous tasks. One reason that may account for this change in strategy is that this task appeared in a later teaching episode (# 11). Thus children had greater experience with data and may have been more comfortable in constructing individual measures of typicality when compared to tasks presented in earlier teaching episodes. A second possible explanation may lie in the representational form. The data for this task were presented on a stem-and-leaf plot whereas the previous tasks utilized a line plot. Perhaps the representational form (stem-and-leaf plot) invoked the construction of individual typical values whereas previous graphical representations (line plots) invoked measures of variability.

Researcher: Tanya, what do you think is the typical weight of a dog based on the data?
Tanya: About 29 lbs because it's the most common of my data on dogs.
Researcher: Okay, what do you mean by most common?
Tanya: Em, it was the most common number .. in other words the mode.
Researcher: Oh okay because it was the mode. Jane what did you get?
Jane: I got 29 lbs too for the dogs because it was the most common too.
Researcher: Okay good, and Jonathan?
Jonathan: Em .. I got 29lbs because I got .. em 38 and 29 were the mode of lbs and because most dogs em .. are middle size .. I think that 29 fits the dog.
Researcher: Okay, so you pick 29?
Tanya: Yes.
Researcher: Jane?
Jane: Nods in agreement.
Researcher: And Jonathan?
Jonathan: Nods in agreement.

When faced with choosing between several modes in a data set, students attended to the distribution of data values around each mode and selected values based on this criterion. In distributions containing no obvious mode, students tended to choose median values as typical values. Hence situations in which the mode was ambiguous showed potential for developing the concept of median as an important measure of central tendency as children naturally attempted to define the median.

Students were aware of values lying at a distance from the main data corpus. In the majority of cases, students chose to disregard outlying values. As a result, typical values were in no way adjusted to reflect outlying values; rather such values were eliminated from all deliberations. In the *Paper Clip* activity several of Jonathan's values were considerably different from those of Jane and Tanya. The transcript outlines the children's decision-making processes regarding how to deal with the outlying values.

Researcher: What would you say the average distance that one of you could blow a paper clip?
Jonathan: Ah, not including mine because I mostly blew mine above the 20s. So, not including mine the estimated is about 10. Most scored around the area of 1 to 10.
Researcher: Okay. Tell me again why you didn't include your scores?
Jonathan: Mine were all 20 and above. And the rest of them were not as high as that.
Researcher: So you thought it was okay to drop those numbers out of your estimation? Jane, what about you? What do you think is the average distance one of you can blow a paper clip?
Jane: I would think around 10. Because 10 had the most because me and Tanya got most of ours around 10.
Researcher: So you are also going to ignore Jonathan's?
Jane: .. nods affirmatively
Researcher: Why do you think so?
Jane: Because his were mostly higher and they were only like one or two or something like that.

The only instances in which outliers were attended to, and were in fact considered pivotal to the determination of an answer, was when the purpose of the exercise was to contrast two data sets rather than constructing a typical value. Hence, when the task involved comparison of data sets, outliers became seminal in deliberations. The purpose of using the *Skipping Rope* activity, an activity designed by Lappan, Fey, Fitzgerald, Friel & Phillips (1998), was to evaluate the strategies used by children to compare sets of data. As can be seen from the transcript, students did not use the mean to compare the data sets, rather the outlying scores in the task were the primary factors influencing the students' decisions.

| | |
|---|---|
| Researcher: | So can you tell me which class you think jumped rope more? |
| Jane: | I think Mr. Kale's class made most jumps because there's a person who made 300 jumps so mostly Mr. Kale's class had the wide range. As for Miss Rich's class it doesn't have a large range at all because the highest they got was 113 which is about 150 less than 300. Also, Mr K's class doesn't have as many numbers in the 1's (meaning less than 100), Mr. K's class has 24 in the 1's and Miss Rich's .. Miss R's has 27. |
| Researcher: | Very good thank you very much. Jonathan? |
| Jonathan: | Okay, I think Mr. K's class because the kids in Mr. K's class got more higher scores than lower scores like for example his class was the only one that got em 151 and higher .. yeah his class got 151, 160 twice and 300 but Mrs R's class didn't get above em .. ah .. 113 so and em .. Mrs R's class got more in the lower .. more 1's and more 20s, more 10s compared to Mr. K's class. |

*Conceptualization of the mean*

Understanding of the mean was varied. Students displayed no appreciation of the benefit of using mean values to compare data sets, indicating an instrumental rather than functional understanding of the mean. The fourth graders exhibited difficulties reasoning about situations involving the mean, unless information pertaining to the number and value of elements in the data set were included. This indicated a lack of awareness of universal features associated with the mean, features that are independent of the actual elements of the data set. For example, in contexts where both the mean and the largest data element were provided, none of the students were aware that the mean could not exceed the value of the largest element in the data set. The *Barbeque task* (adapted from Strauss and Bichler, 1988) was used to evaluate whether children understood that the mean is located between the extreme values in a distribution of data. Jane and Tanya's responses to the *Barbeque task* indicate a lack of understanding of this particular property of the mean.

| | |
|---|---|
| Researcher: | Do you think that if Amy brought 10 marsh mallows and she brought the most, and the teacher gave them out evenly to people, could people have gotten 12 each? |
| Jane: | Yeah. |
| Researcher: | Tanya, what do you think? |
| Tanya: | I don't think there's enough information. You need the number of people and how many marshmallows that they brought. |
| Researcher: | Okay, so apart from that you could make any type of educated guess or estimate? |
| Tanya: | Em .. not without the number of children that were there. |
| Jane: | Jane nods. |
| Researcher: | Okay. You agree Jane? |
| Jane: | Yeah. At first I thought it could be okay. But now I think I need to know how much every one brought. |

When provided with the same information, the sixth grader Jonathan realized that changing the value of an element of a data set would result in a change in the mean. However, all students understood that the mean value was not necessarily a value represented in the data set. Students initially displayed difficulty constructing data sets around a preordained mean principally due to unawareness that 'mean' referred to the arithmetic mean. Following explanation that arithmetic mean was intended, students succeeded in constructing data sets to reflect a particular mean.

Students did not display any problems determining the relative stability of the mean, median and mode and provided sophisticated arguments for the lack of stability of the mean relative to the other values. For example one student mentioned "if a new value were added to the data set it would have to be shared fairly among the other values when finding the mean .. so .. the mean would change more than the median or mean". In the *Watching Movies* task the students demonstrated sophisticated understanding about the measures of central tendency and their ability to index distributions of data.

| | |
|---|---|
| Researcher: | Which measure, the mean, median or mode will be most affected by adding John's score? |
| Jane: | The mean. |
| Researcher: | Okay Jane, the mean, tell me why? |
| Jane: | Cause the mode is going to stay the same 'cause there is only going to be one 96, the median will not change cause 96 will be at the very end inevitably not close to the middle and the mean will change because it wouldn't be a fair share. |
| Researcher: | Okay, so which would be affected the most then? |
| Jonathan: | The mean. |
| Researcher: | Okay, Tanya? |
| Tanya: | Mean. |
| Researcher: | Why? |
| Tanya: | Because you see what I do with the mean is I add them all up and divided them by the number of people so it'd be most affected if I added 96 to that because there are very few numbers near that. Now that would affect the |

| | |
|---|---|
| | mean a lot. It is a very large number. |
| Researcher: | Okay, what about the mode and the median? |
| Tanya: | The mode .. em wouldn't be affected at all. |
| Researcher: | Why? |
| Tanya: | Because it wouldn't be common. |
| Researcher: | And what about the median? |
| Tanya: | The median still wouldn't be affected .. not very much. |
| Researcher: | Why? |
| Tanya: | It would em .. median would move to 9 but that is all it would do. |
| Researcher: | Okay very good and Jonathan what do you think? |
| Jonathan: | Em .. the mean because em the mean right now with the 96 is 9.88 and if you add the 96 it will go up to 12.25. |
| Researcher: | Okay. |
| Jonathan: | So that is a big leap from 9.88. I still think the rest of them wouldn't be as affected 'cause the median will just move up to the next 8. |
| Researcher: | And what about the mode? |
| Jonathan: | The mode .. it'll stay the same. |

*Comparison of students' responses between the first and second clinical interviews*

Discernable changes in students' statistical reasoning were evident from a comparison of pre- and post- clinical interview data. Certain changes in the frequency and types of data organization strategies were apparent. In the initial clinical interview, numerical data were not reorganized. In the post clinical interview, students organized the data by initially ordering the data and then constructing 'bands' or intervals of data. The use of these approaches emerged as a result of data modeling activities, in which students were involved over the course of the study, wherein students collected and organized data arising from their own investigations. Fewer changes were evident when students were presented with categorical data. While students used more sophisticated and inclusive categories within which to reorganize data, the data remained largely unsummarized. Students remained reluctant to collapse categorical data and 'lose' the visibility of individual data values.

Approaches used to describe and summarize distributions of data also changed between initial and post-clinical interviews. Students maintained their concentration on notions of variability; the researcher cultivated this focus, as variability was an important and desirable feature of students' descriptions. The post-clinical interview data revealed an increase in the use of measures of central tendency, in particular the mean, when describing sets of data. Students used balancing and leveling out analogies of the mean in their descriptions, as these conceptual or kinesthetic notions of the mean had been developed over the course of the study. These attempts at balancing distributions brought about a concomitant direction of attention towards modes and outliers, in particular a factoring-in of the outlier, which had been ignored in the initial clinical interview.

Conceptualizations of the mean also changed. Initial notions of the mean were instrumental (Skemp, 1979) in that children's understanding consisted of having a collection of isolated rules at their disposal rather than an appropriate conceptual schema. This resulted in a demonstrated ability to compute the mean but the absence of a strong conceptual understanding of the nature of the mean. Conceptual difficulties were revealed through students' struggles when solving problems that involved anything more than direct application of the mean algorithm. However, the design and use of specific tasks to emphasize conceptual notions of the mean resulted in a richer understanding of the mean in the post-clinical interview. Improved conceptual understanding was evident in improved performance on tasks examining knowledge of features of the mean, and through the use of the mean to describe distributions of data and to compare data sets.

*Other findings*

Students demonstrated an understanding of the relationship between sample and populations. There was awareness that values occurring with low frequency would not be represented to a great extent in a large data set. This was particularly evident in the discussion on the *Raisin data* activity.

| | |
|---|---|
| Researcher: | If we did the experiment again - what about 24? Would you get any more 24s again? |
| Jane: | No. |
| Tanya: | No, 24 is out of it. |
| Jonathan: | You might get some but 24 is the least. |
| Tanya: | Yeah, it is the least probable. |
| Researcher: | Do you think in the factory that there are a lot of 24s (boxes of raisins) coming out? |
| Jonathan: | Well they probably have a lot coming out but not like as many as the ones up higher, near 27 and above because .. em .. 24 is a lot less than 32 and 29 and the all ones that scored high. |
| Researcher: | Do you think we could get one that is 35? |
| Jonathan: | Probably. |
| Jane: | Yeah. |
| Researcher: | Jane? |
| Jane: | It's kind of like 24, you might not get a lot but you could get some. |
| Tanya: | Em, I agree with her. The possibility is very very slim. If you look at the number of 32's and 29's we found and you add them up to the total number of raisins in the factory. I don't think that 35 would have had a chance. |

Furthermore, when asked about the relationship between samples, the students provided a sophisticated explanation of possible similarities and differences between samples. What is interesting in this excerpt is that the children identify the variability of the data remaining relatively constant between samples and the implicit understanding that measures of location (or central tendency) may fluctuate within the parameters of the range.

| | |
|---|---|
| Researcher: | What if we got another sample of bags of raisins, do you think we would get a pattern like that again? |
| Jonathan: | Probably. |
| Tanya: | It would be possible but if you buy the same two brands it is very very possible. |
| Jane: | They would probably be in the same range. |
| Tanya: | If you point the camera at the results that was in all of the boxes (directing that the video camera scan the chart) … I think we would get the same results .. or close to it. |
| Jane: | They would be in the same range. |
| Researcher: | What do you mean by in the range? |
| Jane: | Like between 24 and 33. |
| Researcher: | Between 24 and 33, okay very good. Jonathan, what do you think – if we were to get another 15 boxes? |
| Jonathan: | I think we would get .. like Jane was saying .. probably in the range of 24 and 33. And there probably wouldn't be much of a difference .. I mean a couple of differences .. but not that much. The same probably. |

## Discussion

Several general observations were made over the course of the study. Firstly, students exhibited a lack of awareness of the relationship between raw data and graphical representations of data. Students rarely considered that a graphical representation initially consisted of a set of data values prior to the graphical construction. This finding suggests that wherever appropriate, raw data sets and their graphical representations should be presented in tandem. Related to this difficulty was the finding that children were unlikely to summarize data sets using graphical representations despite their demonstrated ability to construct graphs when requested, thus suggesting a lack of understanding of the functionality of graphical representations. These findings highlight the importance of engaging children in data modeling projects so as to strengthen the association between question construction, data collection, representation and analysis. Students in the study demonstrated a great deal of interest and motivation when engaging in their data modeling projects. They wrote research questions, hypotheses, designed data collection, collected and analyzed data. Students demonstrated great pride and ownership of the projects indicating both the merit and appeal of self-directed inquiry for gifted students.

Secondly, despite their ability to compute the mean of a data set (when asked), on no occasion did a student calculate a mean value when requested to provide a typical value. Neither did the students consider utilizing the mean as a comparative index when comparing data sets. This inability to spontaneously employ the mean indicates poor functional knowledge of the concept of the mean and supports findings of other research in this area (Leavy & Middleton, 2001; Leavy & O'Loughlin, 2002; Mevarech, 1983; Pollatsek et. al., 1981). This finding necessitates the provision of further studies assessing the function and significance that gifted students attach to the mean. Furthermore, students were very influenced by the occurrence of modal values. The majority of typical values constructed by students were modal in nature. This tendency to choose modes, combined with the propensity to ignore scores that are a distance from the main data corpus and avoid utilizing mean values, resulted in students' self-constructed measures of typicality being unrepresentative of the elements of the distribution. This may reflect a curricular emphasis on developing computational rather than conceptual understanding of the mean, and stresses the need for development of a curriculum emphasizing conceptual understanding of statistical phenomena.

A key finding of the study was that of the sophisticated reasoning children used when referring to samples and populations. Despite the fact that it was the first occasion on which they were engaged in focusing on distributions as individual entities, their ability to draw conclusions about populations based on samples indicated statistical inference abilities previously unrecorded. Further study of this area should be carried out and a component of statistical inference included in elementary curricula for gifted students.

Finally, students were reluctant to provide summary measures of data sets, and were inclined more toward providing a range of values within which the summary/typical value would fall. This tendency to refer to the range suggested that the natural tendency to examine the variability of a data set may have conceptual precedence to examining measures of central tendency. This predisposition to providing measures of variability, rather than measures of central tendency, when requested to provide typical values for a distribution of data is a very interesting and a potentially important finding. The study demonstrated that students reasoned about variability and were very aware of the 'spread-out-ness' or 'bunched-up-ness' of the data, awareness not demonstrated to date by their non-gifted peers. In fact, it was obvious from the study that the student's notion of variability determined his notion of central tendency. The fact that the measures of central tendency were determined by their connection to variability may suggest that students have the natural tendency to examine the variability of a data set prior to examining measures of central tendency. The establishment of this proposition necessitates further study into this area.

## Conclusion

This study suggests that gifted children are spohisticated in reasoning about statistics and employ non-traditional ways of thinking about the importance of data sets. Further research on different prototypes of gifted learners is highly warranted.

### References

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.

Cockcroft, W.H. (1982). *Mathematics Counts*. London: HMSO.

Coutant, M.F. (1983). A constituency for the gifted and talented. In A.H. Roldan (Ed.), *Talented and gifted children, youth and adults: Their social perspectives and culture* (pp. 142-143). Manila: Reading Dynamics (Monroe, NY: Trillium).

Curcio, F. (1987). Comprehension of mathematical relationships expressed in graphs. *Journal for Research in Mathematics Education, 18*(5), 382-393.

Dewey, J. (1938). *Experience and education*. New York: Macmillan.

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.

Grant, B.A., & Piechowski, M.M. (1999). Theories and the good: Toward child-centered gifted education. *Gifted Child Quarterly, 43*, 4-12.

Hafenstein, N.L. & Tucker, B. (1995). *Case studies in diversity: Individual differences in abilities and traits of young gifted children.* Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, April 18-22, 1995).

Hardyck, C.D., & Petrinovich, L.F. (1969). *Introduction to statistics for the behavioral sciences*. Philadelphia: Saunders.

Hays, W.L. (1963). *Statistics for psychologists*. Holt, Rinehart and Winston, Inc.

Lajoie, S.P., Jacobs, V.R., & Lavigne, N.C. (1995). Empowering children in the use of statistics. *Journal of Mathematical Behavior, 14*, 401-425.

Lappan, G., Fey, J.T., Fitzgerald, W.M., Friel, S.N., & Difanis Phillips, E. (1998). *Data about U.S.: Statistics*. (Connected Mathematics Series). Dale Seymour Publications.

Leavy, A.M. & Middleton, J.A. (2001). *Middle grade students' understanding of the statistical concept of distribution.* Paper presented at the American Educational Research Association (AERA) Annual Conference in Seattle, Washington, USA.

Leavy, A.M. & O'Loughlin, N. (2002). *Moving beyond the arithmetic average: Preservice teachers' understanding of the mean.* Paper presented at the American Educational Research Association (AERA) Annual Conference in New Orleans, Louisiana, USA.

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

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.

Olenchak, F.R. (1998). Research Skills. In S. M. Baum, S. M. Reis, & L. R. Maxfield, (Eds.), *Nurturing the gifts and talents of primary grade students* (pp. 253-284). Mansfield Center, CT: Creative Learning Press.

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

Reis, S.M., & Renzulli, J.S. (1985). *The secondary triad model: A practical plan for implementing gifted programs at the junior and senior high school levels.* Mansfield Center, CT: Creative Learning Press.

Resnick, L.B. (1989). Treating mathematics as an ill-structured discipline. In R. I. Charles & E. A. Silver (Eds.), *The teaching and assessment of mathematical problem solving* (pp. 32-60). Reston, VA: National Council of Teachers of Mathematics.

Rimm, S.B. (1990). Parenting and teaching gifted children: A theory of relativity. *Gifted Child Today, 13* (3), 32-36.

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

Stein G., & Poole, P. (1997). Meeting the interests and needs of gifted children: A strategy for teaching and learning. *Early Child Development and Care, 130*, 13-19.

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.

Thompson, W.B. (1999). Making data analysis realistic: Incorporating research into statistics courses. In M.E. Ware & C.L. Brewer (Eds), *Handbook for teaching statistics and research methods.* Mahwah, NJ: Lawrence Erlbaum.

Tukey, J.W. (1977). *Exploratory data analysis*. Addison-Wesley Publishing Company.

Watson, J.M. & Moritz, J.B. (2000). The longitudinal development of understanding of average. *Mathematical Thinking and Learning, 2* (1&2), 11-50.

Zawojewski, J.S. (1988). Teaching statistics: Mean, median, and mode. *Arithmetic Teacher, 35*, 25-27.