Journal of Mathematical Behavior 30 (2011) 235–254

Contents lists available at ScienceDirect

The Journal of Mathematical Behavior

journal homepage: www.elsevier.com/locate/jmathb

Elementary and middle grade students' constructions of typicality

Aisling M. Leavy[a,*,1], James A. Middleton[b]

[a] Mary Immaculate College – University of Limerick, South Circular Road, Limerick, Ireland
[b] Center for Research on Education in Science, Mathematics, Engineering, and Technology, Arizona State University, Tempe, AZ, USA

ARTICLE INFO

Available online 5 May 2011

Keywords:
Statistical reasoning
Mathematical thinking
Elementary students
Middle grade students
Typicality
Data and statistics

ABSTRACT

This study addresses the measures chosen by students when selecting or constructing indices to properties of distributions of data. A series of individual teaching experiments were conducted to provide insight into the development of five 4th to 8th grade students' conceptualizations of distribution over the course of 8 weeks of instruction. During the course of the teaching experiment (emergent) statistical tasks and analogous teacher activities were created and refined in an effort to support the development of understanding. In the process of development, attempts were made by students to coordinate center and variability when constructing measures to index properties of distributions. The results indicate that consideration of representativeness was a major factor that motivated modification of approaches to constructing indices of distributions, and subsequent coordination of indices of variation and center. In particular, the defining features of student's self-constructed "typical" values and notions of spread were examined, resulting in a model of development constituting eight "categories" ranging from the construction of values that did not reflect properties of the data (Category 1) to measures employing conceptual use of the mean in combination with other indices of center and spread (Category 8).

1. Introduction

1.1. Concept of distribution

This study is concerned primarily with describing the development of descriptors that represent distributions of data. Distribution is a statistical term for the arrangement of the observations along a scale of measurement (Hardyck & Petrinovich, 1969). In practical terms, if individual data points (e.g. *frequency*, *magnitude*) from an assortment of observations are arranged (i.e. plotted, tabulated) using a common scale of measurement, the result is a picture of the set of data that embodies its structural properties as a whole. Repeated sets of observations will result in sets of differently shaped plots—each of which varies in its location (e.g. center) and its density (e.g. variation). Here is where our primary interest lies: How do students conceptualize these sets of data? What are the features of students' depictions of distribution specifically related to variation and center that hold promise for coherent development of statistical reasoning for complex data sets?

Briefly, in formal statistics, distributions are indexed by *parameters*. Parameters are numerical, descriptive summaries of a population. Theoretical distributions are often described on the basis of the center and the variability of the distribution. In an applied situation (such as a group of students examining the distribution of a particular variable), once the variable and scale

* Corresponding author. Tel.: +353 87 4177551/61 204978; fax: +353 61 345598.
E-mail addresses: Aisling.Leavy@mic.ul.ie (A.M. Leavy), James.Middleton@asu.edu (J.A. Middleton).
1 Tel.: +1 480 965 9644.

0732-3123/$ – see front matter
doi:10.1016/j.jmathb.2011.03.001

of measurement have been selected, a group of scores may be obtained and the distribution of scores can be represented using tabular, graphical, or algebraic means. A mathematical treatment of distribution entails the use of such summary characteristics of the distribution that identify algebraically center and variation. The mean (μ) is the most commonly used measure of *center*. It is the weighted average of the possible values of a variable and is also referred to as the expected value of the variable, $E(X)$. Interpreted in terms of the density of a distribution, the mean can be considered as its center of mass. *Variation* represents the spread of the distribution around a measure of location and is typically denoted by some difference function between individual values in the distribution and its collective center. The variance is the most commonly used index of variation. These measures have been described as emerging "as ways of describing how specific data sets are distributed within the space of values" (Cobb, 1999, p. 10). More commonly, these characteristics describe different geometric properties of a distribution including its center, its variation, its symmetry, and its peakedness. Each of these properties contributes to the statistician's understanding of the peculiarities of a set of data.

Both the Principles and Standards for School Mathematics (NCTM, 2000) and the Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report (Franklin et al., 2007) emphasize the importance of statistics and the centrality of the notion of typicality (i.e. representativeness). Throughout the grades there is a visible emphasis on distributions of data with specific attention to descriptions of shape (grades 3–5), measures of center and spread (grades 6–8), and the selection and calculation of summary statistics (grades 9–12). Researchers and practitioners are only beginning to identify and understand the ways in which elementary and middle school students come to index distributions and the extent to which students are able to provide rich and comprehensive descriptions of distributions of data in the ways envisioned by the principles and standards.

2. Theoretical perspective

2.1. Understandings arising from research on the teaching and learning of statistics

2.1.1. Central tendency

Studies of descriptive statistics have focused on children's conceptions of central tendency and representativeness (Cai, 1998, 2000; Hancock, Kaput & Goldsmith, 1992; Mevarech, 1983; Mokros & Russell, 1995; Pollatsek, Lima, & Well, 1981). Findings suggest that the concept of the mean, in particular, is difficult to understand. One plausible explanation is that many students find it difficult to understand that the mean can be a number not actually represented in the data set. Another factor may be the use of the 'fair share' analogy to describe the mean, a comparison which makes little sense when the calculated mean represents a value that cannot possibly be represented in reality (e.g. a mean of 3.2 persons per household). Moreover, the simplicity of the typical 'sum/n' formula may lead to the misconception that the mean is elementary and straightforward (Mokros & Russell, 1995; Strauss & Bichler, 1988).

Beyond computation one of the overriding features of the mean that constitutes difficulty for children concerns the concept of representative value. Hancock et al. (1992) in a study of fifth through eighth graders found that students had difficulties reasoning about aggregate (e.g. mean). Students did not recognize instances in which the mean could be used to typify a data set, as indicated by the lack of instances where the mean was used to compare two groups of unequal size. This difficulty understanding the representative nature of some descriptive statistics extends to conceptualizations of the median. Research has indicated that procedural fluency in computing the median does not indicate the development of associated conceptual knowledge; students able to calculate medians may not necessarily recognize medians as measures of center or as group descriptors of data (Bakker, 2004; Konold & Higgins, 2003). In fact, many students see the median as a feature associated with a particular data value in the middle of the group rather than as a characterization of the entire group (Bakker, Biehler, & Konold, 2005). The ability, then, of students to compute representative values when specifically instructed (Mokros & Russell, 1995) compared to their inability to construct representative values in other situations (Hancock et al., 1992) suggests that students may not understand the role that representative values play in data analysis.

While we are aware of the inherent difficulties associated with the mean, and to a lesser extent the median, as representative values, we have less insight into other measures that students may use to represent data sets, and the ways in which these measures relate to and index landmarks and trends in the data. Regardless of what measure is used, these summary statistics can have little meaning for students unless the data set can be thought of as more than a series of numbers (Mokros & Russell, 1995) – an understanding of the data set as a unit must exist. When this understanding is not present, and the data set is only seen as a set of disjoint numbers, representative values such as the mean and median have little significance. The importance of having a distributional view of data is illustrated in Bakker's (2004) study of 7th graders. Bakker found that students did not take the distribution into account when considering mean and median; this was particularly evident in symmetric distributions where students did not realize that the mean and median would be similar in value. Perhaps then, one reason why efforts to support students construct understandings of the mean as a representative value have not been as successful as anticipated may be because a distributional perspective on data has not been used.

2.1.2. Variation

Variation has received scant attention both instructionally and in research terms, compared to central tendency (Shaughnessy, 1992, 1997). This is somewhat surprising as variation is a key concept in statistics and fundamental to all aspects of data analysis – it is the attempt to account for and model variation in data that defines statistics as a discipline.

It has been established that variability plays an important, if not central, role in children's statistical thinking (Cobb, 1999; Konold & Pollatsek, 2002; Lehrer & Schauble, 2002; Watson & Kelly, 2002). In a study of 2nd grade students Jones et al. (2001) concluded that when comparing data sets students displayed two conceptions of spread: firstly, a "close-together or far-apart" schema, and secondly, a concept of range. Lehrer and Schauble (2002) used the construction and examination of distributions of data as a context to focus 4th graders on the concept of variability, as expressed by error of measurement and its role in shaping distributions. They found that engaging children in data modeling activities (Lehrer & Romberg, 1996) supported children in developing understandings of variability, in generating plausible explanations to account for sources of variation, and in using variability as an index when comparing distributions of data. Watson and Kelly (2002) found that 3rd grade children's understanding of variation improved following a 10-week instructional unit focusing on data and chance. Highest gains were found in understanding of variation related to chance; smaller yet significant improvements were demonstrated in understanding variation in data and graphs and variation in sampling. All these studies demonstrate that there exists potential, currently untapped in our elementary classrooms, for developing young children's understanding of sources of variation and ways to model it. Research on older students reveals that an overemphasis on central tendency and on the mean in particular, leads students to use measures of dispersion only reluctantly when describing data sets (Reading & Pegg, 1996). These reported studies presenting distributional perspectives have been central in identifying the ability of students to understand and make use of variation in their descriptions of data sets. What is not known is the role of variation in influencing the construction of representative values and at a more fundamental level we know little of the extent to which variation is a feature observed by children in their consideration of distributional structure.

2.1.3. Distribution as an entity

The emergence of distribution as a focus of instruction in primary level statistics has motivated a re-focus on central tendency and variability as measures which describe and summarize distributions. Hence consideration of typicality and spread now arise almost naturally from the need to characterize distributions. In many senses, a focus on distribution is fast becoming a prerequisite for work with ideas of typicality and spread; this is contrary to the traditional order of teaching primary level data wherein central tendency and variability were introduced prior to distribution.

A number of studies have examined how children relate diverse indices of data structure to each other and especially to the notion of distribution. These studies highlight how children coordinate descriptive measures when describing and summarizing data. For example, analyses of classroom teaching experiments have found that elementary and middle grade students use a variety of measures to describe distributions, including measures of center and spread such as modal clumps, ranges, and intervals using proportional reasoning to determine representative values (Cobb, 1999; Konold et al., 2002; Lehrer & Schauble, 2002). Conceptions of average or typicality vary across students and include conceptions of average as modal (values occurring with the greatest frequency) and middle (akin to the median). These conceptions are not static and students frequently move back and forth between chosen measures of average within a given data set as well as between sets of data with different characteristics. One critical skill underpinning the ability to reason about distribution is the ability to view data as an aggregate. Holding an aggregate view of data, as compared to seeing the data as a set of individual values, supports a distributional perspective on data and support the view of data as "a group with emergent properties that often are not evident in any individual member" (Konold & Higgins, 2003, p. 202). Research has shown that with careful attention to instructional design (Friel, Curcio & Bright, 2001), [technological] tools (Bakker & Gravemeijer, 2004; Cobb, 1999; McClain & Cobb, 2001) and contexts, school children can move from focusing on individual data values to holding an aggregate view of data (Ben-Zvi & Arcavi, 2001; Ben-Zvi & Sharett-Amir, 2005). Research exploring learning trajectories and cognitive models of statistical thinking has resulted in significant gains in understanding of how learner's reason about statistics. Nevertheless, there remains a great deal of research to be carried out to investigate how to maximize the full potential of learners. The present study represents one step in the journey to "build connections between levels of children's statistical thinking" (p. 139, Jones et al., 2001) through the development of learning trajectories to focus on the constructs outlined by previous researchers.

3. Purpose of the study

The underlying goal of the study was to move participants toward a way of constructing representative indices that reflected properties of distributions of data. To meet this goal we needed to focus on two aspects of statistical understanding that contribute to the development of representative statistical indices. First was the need to examine the types of statistical measures students select to depict properties of a distribution of data. In this line of inquiry, we attempted to ascertain what (formal or informal) indices students utilized in describing distributions, and how these indices developed and contributed to an understanding of distribution shape and structure. Second was the need to identify the defining features of student's self-constructed "typical" values. Investigation of this question culminates in the construction of an integrated model of the students' learning trajectories when grappling with increasingly sophisticated notions of typicality together with tasks and activities that motivated consideration of distributional features of individual data sets when constructing representative indices.

4. Method

The study design involved four stages: The implementation of a pilot study to develop initial tasks and starting points of the main study; engagement in individual teaching experiments consisting of initial and final clinical interviews and a series of teaching episodes; a retrospective analysis of the clinical interviews and teaching episodes; and the integration of cases.

4.1. Pilot study

The purpose was to develop and test the tasks to be used in the teaching episodes in the main study. The second purpose was the development of an initial hypothetical learning trajectory of children's statistical thinking related to indexing and typifying distributions. The pilot study was implemented over a 3-week period with three students who attended a "Thinking Statistically" course at a Southwestern University in the United States. Two students were female and due to enter 4th grade the following month, the other student was male and entering 6th grade. The researcher was the classroom teacher. Course sessions were carried out for two and a half hours daily over the 3-week period. Small group teaching experiment methods were implemented to efficiently pilot tasks and generate a range of plausible routes for students' thinking. Small group teaching experiments are a smaller instantiation of the traditional teaching experiment; however they occur with a small group of children rather than a classroom of children. Insofar as possible the data used in the investigations were collected by the students. Over the duration of the course nineteen teaching episodes lasting approximately 20 min each were carried out with the group; each episode was videotaped. The analysis phase involved examining the videotape data from the teaching episodes, field notes and samples of the students' work. Results were used to develop a coherent set of tasks that were subsequently utilized in the main study.

Four main points of guidance were derived from the pilot study: (1) students exhibited a lack of awareness of the relationship between raw data and graphical representations of data. This finding guided the main study in that wherever appropriate, raw data sets and their graphical representations were presented in tandem; (2) students located typical values in clusters of data indicating an awareness of variability and an attempt to coordinate concepts of center and variation, reflecting findings of research examining statistical reasoning in the context of distributions of data (Cobb, 1999; Konold et al., 2002; Lehrer & Schauble, 2002). The consistency with which this response occurred drove the creation of tasks and the analytic frame for the main study; (3) students were reluctant to provide summary measures of data sets and were more inclined toward providing a range of values within which the summary/typical value fell. This tendency to refer to the range suggested their natural tendency might have been to examine the variability of a data set prior to examining measures of central tendency; and (4) 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/representative value. Neither did the students consider utilizing the mean as a comparative index when comparing data sets.

In summary, since students' thinking originated in the notions of typicality over center, ranges of values over points, and because their conception of the mean was not central to their conceptions of representativeness, we located the main study squarely in examining how children think about distribution and how we can fruitfully stimulate them to develop indices that require attention to center and spread.

4.2. Participants

Participants in the main study fulfilled a number of criteria. We wanted participants to represent a range of grade levels and ages, and we wanted participants to represent learners who were relatively proficient at mathematics; consequently we sought students who were performing at or above average in mathematics. It was also important that participants had not received any specialized instruction or experiences with data prior to commencement of this study. Five students (3 female, 2 male) from a variety of urban elementary and middle schools in the Southwest participated in the study. Participating students were: Savannah, a 4th grade student (9 years and 1 month), Yan, a 5th grade student (10 years and 4 months), Hope, a 7th grade student (12 years and 5 months), Rachael, an 8th grade student (13 years and 11 months) and Jason, an 8th grade student (14 years and 2 months). Both Savannah and Yan were in mainstreamed mathematics classrooms, and both were identified as somewhere between the 50th and 75th percentiles in terms of mathematical performance. Hope had tested into advanced mathematics and described herself as "struggling" to maintain a B grade. Both Rachael and Jason were also in advanced mathematics; Jason described himself as "about average in the class but it takes a lot of work," while Rachael stated that "about half of the class do better than me in tests." The 4th grade student had had few data handling experiences, limited mainly to pictogram and pie chart interpretation. All other students had received curricular instruction on the construction and interpretation of graphical representations in addition to experiences calculating measures of central tendency and variability (the range).

4.3. Teaching experiment

Teaching experiments, carried out on an individual one-to-one basis, were used to create descriptions of children's mathematical behavior and provide explanations of their progress. The teaching experiment consisted of a series of individual

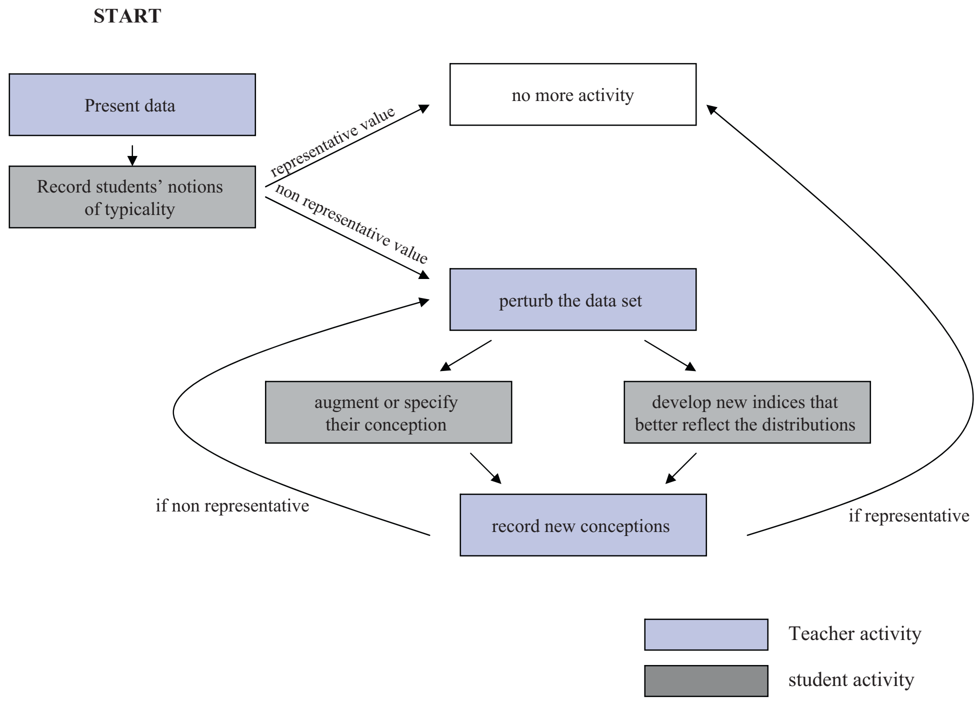

Fig. 1. Teaching experiment cycle of activity.

clinical interviews and individual teaching episodes that occurred over an 8-week period. Data arising from these episodes were analyzed both as the study progressed and retrospectively. Descriptions of mathematical activity took the form of a model of approaches used to construct measures of typicality.

In this study we examined the adequacy of the students' indices of typicality and how they changed as we presented and perturbed distributions of data. The general heuristic used is presented in Fig. 1 and involved us presenting data and recording students' notions of typicality. The data were perturbed in ways that motivated students to augment or specify their conception, or develop new indices that better reflect the distributions (e.g. skew, range, modality). New conceptions were recorded and the cycle repeated if necessary. Therefore by perturbing distributions we were altering the variability of data, not just presenting new tasks. There were occasions that merited the cessation of the cycle of activity. Once children presented a new approach that appeared to work, they were presented with a new data set, the purpose of which was to check the viability of the new approach. If the new approach was found to be viable then the cycle of activity stopped. Thus when children developed conceptions of typicality that were representative for a specific instance, then activity within a particular cycle ceased.

4.3.1. Clinical interview phase

A clinical interview was carried out with each of the students prior to and following the teaching episodes. The initial clinical interviews consisted of 11 tasks administered to each of the students in the same order (Appendix A). The purpose of administering the clinical interview prior to starting the teaching episodes was to identify the experiences that students had had in statistics and to determine their initial approaches to describing and typifying distributions. Initial clinical interviews were generative (Clement, 2000). Generative interviews are designed to support open interpretation of the episodes and focus on discovering new constructs and processes based on analysis of the observations arising from the interviews. As observational categories are not fixed ahead of time, it was intended that analysis of data arising from these type of clinical interviews would generate observational categories and elements that were components of a model of students' statistical understanding. This information derived from the clinical interview provided a baseline to anchor the development of each individual child's learning trajectory. In addition, a clinical interview was administered following the collection of teaching episodes. Analysis of data arising from this final clinical interview was comparative in that it examined the nature of representative values constructed and the degree to which the values indexed properties of the distributions as compared to responses on the initial clinical interview. Inclusion of the 11 initial clinical interview tasks supported assessment of growth in students' conceptualizations of data. The final clinical interview also contained elements of convergent interviews (Clement, 2000) due to the inclusion of six new tasks designed to confirm or disconfirm a priori assertions of the researcher generated from analysis of teaching episode data. The convergent interviews involved detailed coding of small segments of responses on these tasks. As is the case with convergent interviews, specific criteria for recognizing phenomena were constructed prior to interviews. Data arising from the new tasks were analyzed according to these predefined criteria in an effort to examine the reliability of assertions developed from analysis of generative interviews and teaching episodes. Analysis of data arising from these tasks aided in constructing models of statistical reasoning. Clinical interviews lasted 60–90 min, due to the inclusion of some data collection and graphing activities, and were audio taped.

4.3.2. Teaching phase

Teaching episodes took place following the initial clinical interview and were carried out on a one-to-one basis with each student. Six teaching episodes each lasting approximately 50 min were carried out with four students; each episode was audio taped and several sessions were videotaped. Eight teaching episodes were carried out with Savannah, the 4th grade student. These additional teaching episodes were required to address difficulties conceptualizing the meaning of graphically presented data.

Arising from the clinical interview, we constructed hypotheses regarding individual students' statistical understandings. Hypotheses were specific to participants and their responses and consisted of conjectures regarding the statistical reasoning underpinning responses on tasks. Tasks were designed to probe these understandings and test the hypotheses. Once a task was presented, we guided "the child through the problem, while being guided by the child's own approach" (Piaget, 1963). This guiding took many forms. At times it took the form of questions that probed students' understandings, on other occasions it reminded the student of approaches they had used in previous teaching episodes and asked about the suitability of such approaches, and to a lesser extent it presented examples of other students' approaches used in similar situations. It was not the students' solution that was important, rather their untutored individual approach, i.e. the students' mathematics. We did not try to lead the student toward a statistician's view of the problem or to teach the correct approach to the student – throughout the teaching experiment our focus was on learning more about the students' reasoning.

Within any one teaching episode, participants were asked to respond to a specific task; the response generally required the construction of a typical value. In much of the literature the notion of typicality assumes that we are speaking only about measures of center; however, in the context of this study typicality refers to any measure or index (such as measure of center and/or measures of variability) used to represent the data. In other words, typical values refer to values that *index* or *represent* the collective data values of the distribution (a distributional index). It should be noted that the term "typical value" was rarely used, as the problem context and ensuing task were designed so as to establish the need for the construction of such values. Additionally, our desire was to avoid the use of language that might suggest attention to any particular measure. A selection of tasks used throughout the study is presented in Appendix B. Of these tasks, Savannah was administered Tasks 1–5 and Task 9. Rachael, Hope and Jason were administered all tasks presented in Appendix B.

The density of observations that occurred combined with the co-construction of tasks provides an in-depth picture of the factors that promoted the changes in mathematical understanding and promoted development. The recording of minute changes in participants' constructions of typicality over an 8-week period of time reflects microgenetic studies that examine mathematical development. Microgenetic studies of cognitive development (Siegler, 1996; Siegler & Crowley, 1991) utilize a general conceptual framework for thinking about cognitive development in our analysis; we found it useful to adopt elements of this microgenetic framework in our attempts to make explicit the connections between teacher activity and students' learning.

4.3.3. Analysis phase

The primary sources of data in the analysis phase were video recordings of individual teaching experiment sessions, field notes, and copies of the students' work. Videotape analysis focused on explanations and descriptions provided by participants in addition to actions and gestures made in relation to a given representation and/or data set (e.g. pointing to individual data values or clusters of data). Analysis of the strategies and responses to tasks presented in the initial clinical interview resulted in the construction of individual hypothetical learning trajectories (Simon, 1995) for each child. The trajectory provided a sense of direction and set out the possibilities for instruction in terms of learning goals. The responses and strategies utilized in each of the teaching episodes defined the direction of subsequent episodes; as a result the trajectory of each participant's path was often in flux. This resulted in extending and modifying the trajectory and goals throughout the teaching experiment to fit with and extend the students' mathematical activity. In addition to examining the main ideas identified at the beginning of the study, unanticipated questions (Steffe & Thompson, 2000) emerged over the course of the study. Based on a participant's unanticipated responses and attempts to index a specific distribution, hypotheses were generated about the participant's approach. Tasks were then designed to test the hypotheses, the participant's responses to the tasks were coded and a new hypothesis generated based on the analysis. These cycles of hypothesis formulation, testing, and reconstruction continued throughout the course of the teaching experiment.

Analysis of the data involved consolidating, reducing and making sense of participants' activity in the teaching episodes. Key phrases and statements arising from an individual teaching episode were located, hypotheses relating to the participant's mathematical reasoning were made, and tasks were designed to test these hypotheses (emergent tasks). Emergent tasks had not been constructed a priori to the initiation of the study; rather they were developed to test the reliability of new observational categories arising from analysis of the generative interviews. 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 were found then tasks were designed to either (a) further examine the participant's conceptualizations, or (b) challenge the participant's reasoning in relation to the concept. For example, The Winter Olympics Problem (Task 2, Appendix B) was a task initially designed to explore the nature of representative values constructed when a participant is presented with a skewed data set. In response to this task, one participant presented the value of '1' (the mode in this case) as a representative value. It was hypothesized that the student considered modes as representative values, so an emergent task, Performance in Statistics (Task 6, Appendix B) was designed to test this hypothesis, this time however presenting a negatively skewed data set. The mode was chosen again as the representative value, and key statements made indicated the students' belief in the representativeness of the mode for all data sets. A second emergent task, Hourly Pay (Task 7, Appendix B) was then designed to challenge the notion that a mode was always an appropriate representative value by presenting a distribution where the mode was isolated from the larger body of data.

Retrospective analysis occurred after completion of the teaching experiment and involved re-analyses of audio and video records of the students' work. The retrospective analysis sought to find patterns in participants' learning processes, to examine the impact of the instructional materials, and in turn investigate whether the hypotheses developed over the course of the teaching experiments were correct. Whereas the purpose of the analyses carried out of each individual teaching episode was to develop participants' understanding, the retrospective analysis functioned to find empirical evidence to support or reject assertions and in turn to place the study findings within a broader theoretical framework. The retrospective analysis was carried out twice for each student. The first retrospective analysis consisted of examining the records in the order in which the clinical interviews and teaching episodes occurred. The purpose of this analysis was to reexamine participants' mathematical activity using a historical lens. The primary teaching episode analyses were influenced by the knowledge arising from the clinical interview and teaching episodes that had preceded it. In the retrospective analysis these episodes were reexamined in light of all we knew, in other words we had knowledge of episodes that followed the one being examined, resulting in providing a different perspective than that held originally. The first pass over the data resulted in the generation of new insights into the data and the corresponding construction of new questions about students' understandings. The purpose of the second retrospective analysis of the data was to find confirming and disconfirming evidence related to the learning trajectories constructed for individual students. Accordingly the second pass through the data was structured around reexamining the proposed learning trajectories with a view to modifying the trajectories in light of the collective picture presented by the retrospective analysis.

4.4. Tasks

Each of the teaching episodes was characterized by the presentation of specific tasks designed to elicit models of participants' understanding of certain statistical phenomena. Tasks were categorized into four groups. The 11 clinical interview tasks were presented at the initial and final clinical interviews and were broad in scope so as to provide an extensive description of individual participant's statistical understanding. Clinical interview tasks were aimed at developing a picture of the extent to which participants were familiar with graphical methods (Task 1, Appendix A), measures used to describe distributions of data (Task 2a, Appendix A), typical values (Task 2b, Appendix A), data comparison strategies (Task 3, Appendix A), measures of central tendency (Task 4, Appendix A), and understandings of the mean (Task 5, Appendix A). The second and third group of tasks were used in the teaching episodes: the second group was designed to examine approaches to describing data (Tasks 1 and 2a, Appendix B) and the third group explored the defining features of student's self-constructed typical values (Tasks 2b and 3–10, Appendix B). These tasks were more in-depth than clinical interview tasks with many of them probing the responses produced in the clinical interview. The final clinical interview had several tasks in addition to those presented in the initial clinical interview (for example Task 6, Appendix A). These additional tasks probed the understandings that had been developed over the course of the teaching experiment.

The data and tasks used in the teaching episodes were derived from four sources. For one subgroup of tasks, the data were generated during statistical investigations carried out by the participants themselves (Task 11, Appendix B). The benefits of engaging students in reasoning about data they themselves have collected results in experiences and activities that empower students, actively engage students in doing statistics, and provide authentic experiences of statistical activities that have real-world applications. The second group of tasks was emergent tasks which were constructed by the researchers in response to the meanings generated by students in the teaching experiment. These tasks were used over the course of the teaching experiment in an attempt to uncover unanticipated approaches and ways of thinking held by students. The third source of tasks was tasks used by permission of researchers at Vanderbilt (McGatha, Cobb, & McClain, 2002). The fourth source was tasks constructed by the researchers prior to initiation of the study, and it was generated as a result of patterns emerging from the pilot study.

The majority of tasks were administered to all participants at some stage in the teaching experiment. However, across participants, no two teaching episodes were the same, as episodes were developed to respond to participants' mathematical activity in previous episodes. Tasks were presented in several forms (e.g. graphical, tabular, etc.) and students were asked to describe distributions and determine typical values for the distributions. All data sets were presented predominantly on line plots or on stem and leaf plots chosen due to their structure facilitating the display of individual data values, the visibility of which was considered important for the construction of typical or representative values.

Representative Distributional Indices	Category 8			Measures employing use of the mean in combination with other strategies	Measures employing use of the mean in combination with other strategies	Measures employing use of the mean in combination with other strategies
	Category 7		Measures employing use of the mean	Measures employing use of the mean	↑	↑
	Category 6		*Measures employing use of the median*	↑	*Measures employing use of the median*	*Measures employing use of the median*
	Category 5		Measures which reflect an adjustment from the mode	Measures which reflect an adjustment from the mode	Measures which reflect an adjustment from the mode	
	Category 4		Measures which specify a typical value located within a suitable range or interval of data values		Measures which specify a typical value located within a suitable range or interval of data values	↑
	Category 3		Measures which specify a range within which a typical value occurs		Measures which specify a range within which a typical value occurs	
Non-Representative Distributional Indices	Category 2	Typical values which employ exclusive use of the mode	Typical values which employ exclusive use of the mode		Typical values which employ exclusive use of the mode	Typical values which employ exclusive use of the mode
	Category 1	Typical values that do not reflect the data				
		Savannah's developmental trajectory	**Hope's** developmental trajectory	**Yan's** developmental trajectory	**Rachael's** developmental trajectory	**Jason's** developmental trajectory

Fig. 2. Profiles of students' individual learning trajectories when constructing typical values for distributions of data.

5. Results

Results are organized by first describing classes of strategies participants employed when indexing data. We then coordinate the teaching activity (i.e. tasks and questions), that mediated between categories of strategy use and across different conceptions of center and spread, with students' constructions (Simon, 1995; Steffe & Thompson, 2000). Following the establishment of this frame, we provide its' embodiments in the form of the detailed cases of the five participants, highlighting commonalities and idiosyncrasies among them.

5.1. *Coordinating student learning and teacher activity*

Fig. 2 outlines the learning trajectories of each student when constructing distributional descriptors. The individual learning trajectories incorporate two dimensions of cognitive growth utilized by microgeneticists: path and variability. The trajectories encompass *path of change* as they depict the sequence of approaches used while gaining competence in representing distributions. Each trajectory, then, represents movement through a set of ordered categories (Agresti, 1996), where a category is considered to represent an array of approaches utilized to represent a body of data. *Variability of change* refers to differences among children; while all children moved through the categories in the same order, they did not all enter the teaching experiment at the same point. Each category incorporates all approaches in the category preceding it; and movement through categories assumes acquisition of additional approaches. For example, students who utilize Category 2 approaches when constructing representative indices have a considerably smaller number of distributional indices at her disposal compared to a student situated at Category 5 who has a repertoire of several approaches at her disposal. In other words, the model (Fig. 2) represents movement through a set of ordered categories consequently demonstrating acquisition of a greater *number* of representative indices rather than implying greater *sophistication* in approaches. As a result student A (Hope, for example), who is situated at a category of higher order than student B (in this case Savannah), is likely to generate a more representative value due to the variety of approaches at their disposal to index features of any given distribution.

Elements of the model presented in Fig. 3 are comparable to components of the Hypothetical Learning Trajectory posited by Simon (1995). The "creating" and "refining" activities in Fig. 3 represent the instructional activities component of the learning trajectory. The creating activities were designed to support the development of understandings related to concepts closely tied to the study (e.g. the mean) whereas the refining activities were developed to stimulate movement from particular unrepresentative approaches. These activities all contribute to reaching the learning goal of constructing representative indices that reflect properties of distributions of data. Movement through the ordered categories, then, represents the

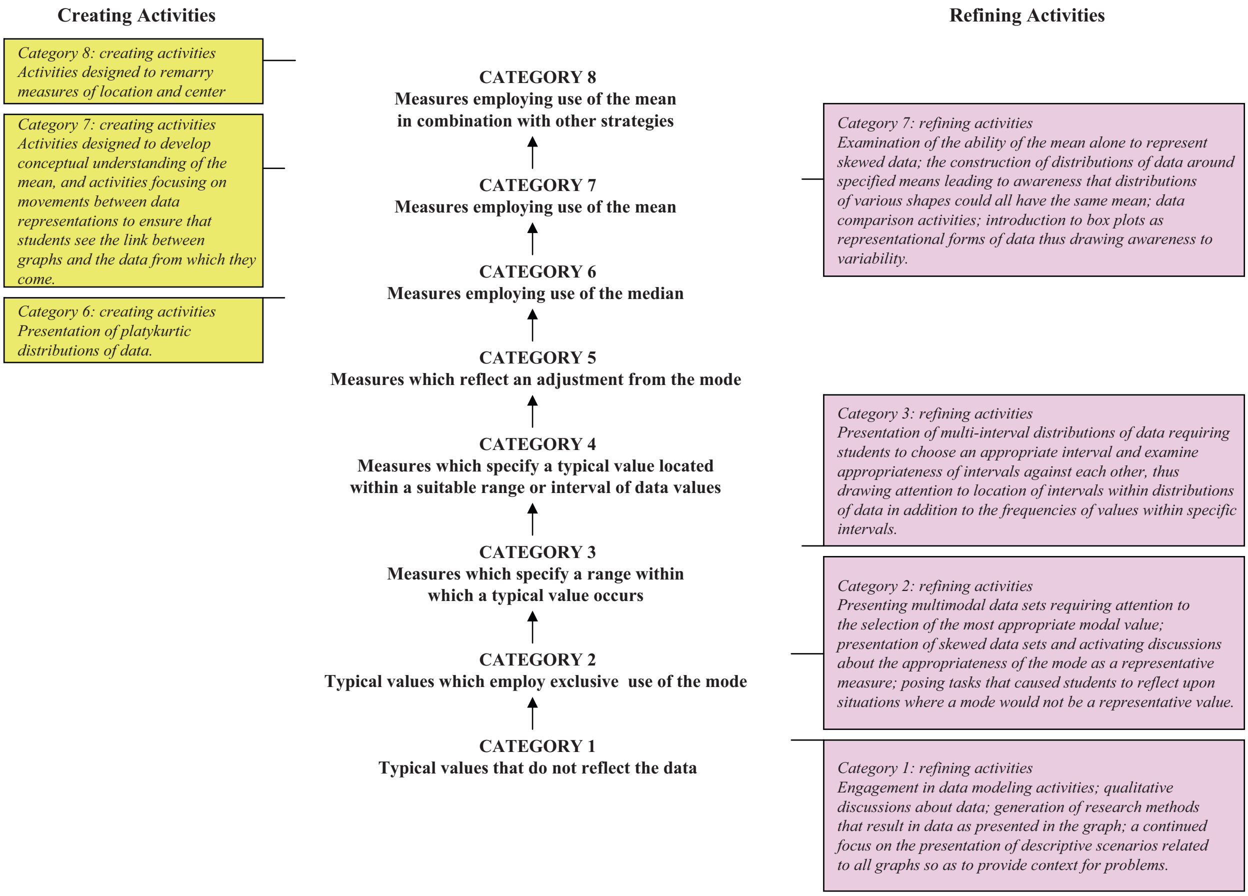

Fig. 3. Sources of change stimulating student activity.

learning process component, as each category (with the exception of Category 1) consists of a measure that has the potential to index a particular property of a distribution of data.

Movement through the set of ordered categories was motivated by presenting students with a variety of tasks to challenge their current understandings and stimulate them to develop new measures that better describe the new data. Teacher activity was pivotal in supporting and triggering student learning, and it involved interactions with the students that were intended to engage them in mathematical activity related to the tasks. Through engagement in mathematical activity, the goal was that students' conceptual constructions would be made explicit and developed. These manipulations by the instructor/researcher constitute the *source of change* of students' constructions of typicality (e.g. Siegler & Crowley, 1991). In Fig. 3 we make explicit the effect that teacher activity had on children's changing constructions of typicality through the presentation of a framework that coordinates student activity (as presented in their learning trajectories) with this teacher activity, through the delineation of types of tasks and activities presented by the teacher that stimulated students' changes in thinking for different distributions. As mentioned earlier, activities are divided into two classes: creating activities and refining activities. Refining activities were used when the researchers observed a student consistently using an approach that resulted in poorly representative measures. Tasks were designed to address the pattern of mathematical behavior. These tasks were designed specifically to draw students' attention to deficiencies related to their approaches and stimulate movement to the use of more representative measures. In the majority of occasions, however, the researchers did not know what these more representative measures would be, and neither was it our intention to suggest a specific alternative approach. Examination of Fig. 3 shows that these refining activities were used to stimulate movement from the categories of approaches identified as 1 (Task 11, Appendix B), 2 (Tasks 2, 6, and 7, Appendix B), 3 (Tasks 4 and 9, Appendix B) and 7 (Tasks 12 and 13, Appendix B). The intention, then, of refining activities was to highlight inadequacies in approaches with the hope that this spotlight would motivate generation of more representative indices. The purpose was not to direct students toward a specific approach; consequently the focus was on moving students *away from* a specific approach rather than *directing toward* an alternative approach.

Throughout the teaching experiment many of the students' approaches and strategies emerged as a result of interaction with a variety of rich data sets. However, in some instances it became increasingly evident that certain (useful) approaches would not materialize unless purposefully seeded by the teacher–researcher. Tasks designed with these goals in mind were referred to as creating activities. For illustration purposes, in Fig. 3 creating activities are presented alongside the category to which they are related. However, the purpose of these tasks was not to influence or direct students along a particular trajectory; it was rather to provide opportunities for students to construct the understandings necessary to utilize the measure as a representative index. One such instance involved constructing a conceptual approach to the mean. Several teaching episodes engaged students in tasks designed to support them in constructing understanding of the mean as a *fair share* and as *a balancing point or center of gravity of a distribution*. Further illustration of how creating and refining tasks were employed throughout the teaching experiments is detailed in the case-by-case analyses.

5.2. Constructing typical values to describe distributions of data

Overall, students' notions of typical values fell into one of two classes: Representative and non-representative. Table 1 outlines the various strategies used by students to come up with a number that summarizes a set of numbers.

5.2.1. Non-representative indices

5.2.1.1. Indices that do not reflect the data. The first and most primitive distributional indices were those that did not reflect the data set. These responses were idiosyncratic and were proffered by students who were not familiar with data. Students who constructed these indices were presented with Category 1 refining activities (see Fig. 3) which required experiences in actively collecting and analyzing data to come to an understanding of the meaning of data values (Lehrer & Romberg, 1996). The transcript of the Gummi Bear problem (Task 1, Appendix B) illustrates how Savannah depended on her own experiences of the problem context and consequently ignored the presented distributions of data.

T: How many gummi bears would you say are in a packet of candy?
S: Ah. like . . . maybe 25 or no. probably 26.
T: Why do you say 26 gummi bears? . . .. . .
T: And did those students (pointing to the graph) get as many as 26?
S: No, but . . . I think you'd get more. It really depends on . . . just . . . I always get more.

5.2.1.2. Employment of the mode exclusively. The second non-representative measure was exclusive use of the mode. Admittedly, the mode is a representative value for certain distributions, although it is not to these situations that we refer. We refer to instances wherein the student applied the mode as a blanket measure regardless of the distributional shape. Students who used the mode extensively considered the most frequently occurring value to be most representative of any given distribution, despite the presence of other factors such as gaps in the data, skew and the presence of outliers. The following transcript (Task 2, Appendix B) illustrates Hope's exclusive reliance on the mode in determining a value to index the distribution.

T: Generally what was the number of medals won by a country participating in the Olympics?
H: One.
T: Why one?
H: Because most countries got one.
T: Do you think one is representative of every country though? Does it represent all of them?
H: Yeah. Cause it's there the most.

Students who demonstrated a preference for the mode generally needed to reflect on the capacity of the mode to represent all the values of the distribution (see Category 2 refining activities, Fig. 3).

5.2.2. Representative indices

Six different classes of strategies used by the participants were classified as representative. Each class of strategies was representative for some distributional form but not necessarily for all forms.

5.2.2.1. Specifying a range. The first representative strategy was that which specified a range within which a typical value occurred. The relevance of variability in determining indices of distribution is evident in this approach. Examination of these ranges or intervals highlighted that they bounded clusters of data, the density of values within these regions being higher than densities within comparable intervals elsewhere in the distribution. In many of the cases the ranges encompassed the mode of the data set – in these situations this strategy reflects the *modal clump* described in the literature. However, the range did not always encompass the mode. In the following transcript (see Task 3, Appendix B), Hope the 7th grader utilizes ranges in her constructions.

T: Based on that graph how long do you think a 5th grader can hold their breath?
H: Well in the range of 9 to 12 seconds or 8 to 12 seconds . . . around there
T: why?
H: Cause most of the students in the class were around there

5.2.2.2. Specifying a value within a range. A second closely related representative measure involved locating a specific typical value within a 'suitable' range or cluster of data. The specified ranges normally encompassed the largest cluster or 'clump' of data values. At times the value fell within the modal clump and at other times it did not. This strategy represents an attempt to coordinate location and variation within a distribution of data. In the following transcript (see Task 4,

Table 1
Strategies used to construct typical values.

Type of approach	Specific strategy
Non-representative approaches	Employment of indices that do not reflect the data • Values do not fall within the range of the data values • Justification of typical value: ○ relates to the students' own knowledge of the problem context ○ makes no reference to the presented data values • Student lacks experience constructing and using data
	Employment of exclusive use of the mode • Value reflects the mode of the distribution • The most frequently occurring value is considered most typical of any given distribution • Features of the distribution such as gaps, skew and outliers are not considered when constructing the typical value • Student normally does not equate the meaning of the term "typical" with the concept of representativeness
Representative approaches	Specifying a range within which a typical value occurs • Reluctance to designate a specific typical value • Nomination of a range of values within which an unspecified typical value occurs • Range encompasses a large corpus of data and normally excludes outliers • The appointed range reflects an attempt at incorporating greater representativeness in determinations of typicality
	Specifying a typical value located within a suitable range or interval of data values • Natural progression from the stage of nominating a range alone • A range/interval is specified and then a typical value identified that is located within the range • An interval is considered suitable if it: ○ is located in the middle of the data set, or ○ represents the largest interval of data in the distribution • The typical value normally reflects the mode of the range/interval • Little attention is paid to features of the data such as gaps, outliers and skew
	Adjustment from the mode • Utilizes the mode as a benchmark from which to begin the process of constructing a typical value • Adjusts the typical value from the mode in the direction of specific features of the distribution (outliers, clumps of data etc.). Indicates increasing awareness to the issue of representativeness when choosing typical values
	Employing use of the median • Occurred less frequently than other methods • Use of the median is motivated by distributional shape and employed when the: ○ distribution contains an ambiguous mode ○ data are non-interval, or ○ when existing intervals are equal in size
	Employing use of the mean • The mean is utilized as a typical value when student possesses ○ A rich conceptual understanding of the mean ○ Understanding of mean as representing the balancing point of a distribution
	Employing use of the mean in combination with other strategies • Mean is calculated and a value chosen based on comparison of the mean and either: ○ median or mode ○ an eyeball estimate of typicality • Strategy developed by students as a result of working with a wide variety of distributional shapes • Students normally possess understanding of the mean as some form of middle, but do not have understanding of the mean as a balance point of the distribution

The degree to which a strategy employs notions of representativeness is determined, in part, by the shape of the distribution. For example, use of the mode embodies representativeness for categorical data and for distributions with relatively few "clumps" of a reasonable size. It may not be representative for flat distributions of data with many clumps of proportional size.

Appendix B) Rachael's determination is based on identification of a cluster of data and location of a specific value within the cluster.

T:	What is the typical height of a student in Miss Murphy's class?
R:	48 inches.
R:	Well or 43.. it's in that bunch between 42 and 44. No, I think 48.
T:	why?
R:	It has one of the most and clustered around it is like a whole bunch of like similar ones
T:	And what about 43 then?
R:	Well look at all the people that are around 48 inches.

5.2.2.3. Adjustment from the mode. The third representative strategy emerged when students found the mode and then adjusted to account for the presence of outlying groups of values. The following task (see Task 2, Appendix B) carried out with Yan illustrates how he adjusts from the modal value in an attempt to reflect other data in the data set.

T: What is the general number of medals won by a country?
Y: Oh eh probably 9.. Actually no not 9. *Yan laughs.* 6
T: 6. Why 6?
Y: Well like most got 1 . . . but a lot got more than one. And because like most of the people are below 6 but 1 or 2 people are above it so yeah.

5.2.2.4. Employing use of the median. A fourth representative measure was the use of the median. This method occurred less frequently than others and was motivated by the shape of the distribution. The median signified a representative measure particularly in situations where students were presented with skewed sets of data or data with an ambiguous mode. Hope used the median only when presented with a platykurtic distribution that had no clusters of data (Task 10, Appendix B). Similarly, Rachel used the median only on occasions when the mode was ambiguous and when data were not clustered (line plot 1 of Task 5, Appendix B). The following transcript (Task 4, Appendix B) illustrates Jason's use of the median as a strategy to determine a representative value. In the following situation, the distribution consisted of two modes. In determining the representative value Jason identified the modes, and then calculated the median of the data and determined which mode was closest to the median value. This mode was then designated as the representative value.

T: What is the height of a student in Ms. Murphy's class?
J: 48 inches . . . or 43 inches
T: Okay.
J: But 48 inches is the mode and it is closest to the median so I would say 48 inches and . . . no it wouldn't be close to the median . . . so . . . let me see
Jason calculates the median by counting half of the X's
J It is it's [48] the median and it's one of the modes . . . so 48 is the height

5.2.2.5. Use of the mean. In situations where the data set resembled a normal distribution, using the mean as an index of distribution resulted in the construction of representative values. However use of the mean as an index of graphically represented distributions did not occur, for the majority of the students, until the latter end of the teaching experiment, when understanding the mean as a balancing point was sufficiently developed. The following transcript illustrates Rachael's initial use of the mean as a measure to index the distribution of values (see Task 6, Appendix B).

R: This person is not a very good student . . . Two weeks he got 18 out of 25 and 4 weeks he got 25 out of 25. I would just average it. Like get the mean.
T: Okay. Why?
R: Cause it counts the high ones and the low ones. That's what they do with grades anyway they get a grade point average . . . I want to know what he got anyway in this class because if he can do so badly and still get a good grade I might consider doing that.

5.2.2.6. Employing the mean in combination with other indices. The strategy of <u>mean-in-combination</u> dealt with non-normal and skewed data sets. This developed out of a realization that emerged that the mean had poor stability when used with small and skewed data sets. The following example (see Task 7, Appendix B) illustrates Jason choosing measures of central tendency to determine a value to represent the data set.

T: What do you think is the typical hourly wage paid to people living in that community?
Jason begins to count the X's on the graph (he is determining the median)
J: I calculated 18.5 for the median. Yeah. Um . . . Most got 6. For the mode.
J: 692 divided by 40. *Jason calculates the mean correctly*
T: Your mean was?
J: 17.3. I am just going to say.. *Jason crosses out the mode of 6*
T: You've gotten rid of your mode. You got a mean of 17.3. And a median of 18.5.
J: How about 17? *Jason writes 17.* 17. That is the hourly rate.
T: Why.. so in this case you did the three of them. The mode, median and you did the mean. Now last week you always did two, why did you decide to do three this time?
J: Cause there is such a wide range of numbers.
T: Okay. When you found the answers.. you decided not to consider the mode. Why not?
J: Um.. just so.. it is such an outcast.

5.3. Case-by-case analysis

5.3.1. Case 1: Savannah

Savannah initially provided distributional indices based on her own experience of the phenomenon under investigation (Category 1). Once this pattern of constructing idiosyncratic and non-representative measures was identified, the researchers re-examined all instances in which the strategy had occurred and developed a hypothesis to account for the observed mathematical behavior. The hypothesis proposed that Savannah possessed limited understanding of the meaning of data; as a result she drew on her own experiences of the problem context when constructing representative values. A goal was identified as the need to instigate a move from constructing idiosyncratic measures of typicality to constructing measures that reflected the presented data. Activities and tasks were designed to realize the goal (Category 1 refining activity, Fig. 3) and several teaching episodes were devoted almost exclusively to data collection activities (Task 11, Appendix B for example), accounting for the greater number of teaching episodes carried out with Savannah. Examples of data modeling activities involved: measuring the length of leaves, counting the numbers of raisins in boxes of raisins, the number of candies in packets, recording heart rates, etc. As a result of engaging in data modeling activities Savannah began to attend to the data, and midway through the teaching experiment started to pay attention to externally constructed data sets. Following extensive data modeling activities Savannah began to construct values using a second strategy: the provision of modal values (Category 2). The use of modes was a welcome addition to Savannah's repertoire as it indicated a move from idiosyncratic measures and an increasing understanding of the meaning of data. Over time, it became apparent that Savannah was consistently using the mode as a blanket measure of representativeness. However, regardless of the shape of the distribution presented to Savannah, she consistently chose the mode as the value that represented the data set. Many attempts were made to draw Savannah's attention to the importance of a typical value being representative (Category 2 refining activities, Fig. 3). It was hoped that Savannah would realize that the mode, while representing an observable landmark in the data, was not always appropriate to use as a representative value.

Retrospective analysis of the work carried out with Savannah implies that she may not have been cognitively ready for many of the experiences presented to her. Analysis of her performance on the tasks, and given that she never constructed modal clumps, indicated that she did not consistently engage in multiplicative reasoning in data analysis contexts. Evidence of multiplicative reasoning was verified in the clinical interview; however further instances throughout the teaching episodes occurred only sporadically. This finding is supported by the research literature that has shown multiplicative reasoning to develop slowly across the early elementary grades (Clark & Kamii, 1996; Steffe, 1992).

5.3.2. Case 2: Hope

Hope's initial attempts to index distributions were modal in nature (Category 2). While the mode was an appropriate representative measure for normal distributions of data, Hope was also observed using modes when presented with skewed distributions. Hope was presented with a variety of non-normal distributions to test the hypothesis that she was not attending to the distributional shape. Hope continued to suggest the modal value as representative of the distributions. Once this pattern was confirmed, we proposed hypotheses to account for the pattern of behavior. We hypothesized that Hope was attending to frequencies without considering representativeness, and we set the goal to draw her attention to the limited capability of particular modal values to adequately represent all values within a data set. A number of activities were designed to bring about this goal (Category 2 refining activities, Fig. 3). Hope continued to choose modal values as typical until she was asked to consider whether her typical values were representative of all the data points. Confronting the notion of representativeness resulted in Hope's altering her approach to choosing typical values and employing a variety of strategies in coming to a determination of typicality.

Once Hope disposed of using modal values she employed the use of ranges of data to index a data set (Category 3). These ranges normally encompassed the largest interval or "cluster" of data values. It resolved the dilemma of representativeness in that the range was more representative than modal values because ranges encompassed a greater number of the values present in the data set. Hope moved from providing large ranges of data values as representative indices to choosing a specific interval within which the typical value would lie (Category 4). Her decision regarding the most appropriate interval to choose was based on a number of factors: the number of data values lying within the interval and the position of the interval relative to other intervals. Once the interval was chosen she tended to select the value in the middle of the interval as the typical value.

Following a variety of experiences constructing typical values for distributions of data it became evident that Hope did not have the mean in the repertoire of strategies available to her. While Hope possessed the ability to calculate the mean when asked (evident from the clinical interview), she did not utilize the mean as a representative value when presented with graphical representations of data. Our goal then for Hope was the development of conceptual understanding of the mean and in turn to develop understanding of the mean as a representative value, through the use of creating activities (Category 7 creating activities, Fig. 3). The transition to employing the mean as a measure of typicality required many experiences working with a variety of tasks over several teaching episodes. This understanding finally came about as a result of activities in which Hope was presented with raw data sets that required summarization (i.e. the data were presented in lists as opposed to being presented graphically). She readily employed the mean in summarizing these data sets, and she slowly began to realize that it could also be used to index a graphical distribution of data. It was following these activities that Hope employed the mean as a measure to index distributions of data. Had she not developed some understanding of the function of the mean in the context of other activities, she may never had spontaneously utilized it when working with graphical distributions of data.

Hope used the median (Category 6) as a typical value in two situations. When platykurtic data were presented on line graphs she calculated the median value as the typical value. However she was reluctant to base her determination on the median alone and preferred to verify her answer by calculating the mean; and in cases where a discrepancy existed she chose the mean rather than the median. The second occasion when she employed the median was when the data were presented on stem and leaf plots. The presentation of data on line plots, wherein the actual numerical values of the data are present, may have mirrored traditional contexts of identifying the median from a list of ordered data.

5.3.3. Case 3: Yan

In the initial clinical interview Yan displayed a more sophisticated understanding of representativeness than did any of the other students. He demonstrated an understanding of the typical values as representative of all values in the data set, by the adjustment of the typical value from the mode in the direction of the greater body, or tail, of data (Category 5). Outliers and tails in the data were central to influencing his construction of distributional indices, as they were the features of the data to which he attended when adjusting from the mode. In fact, Yan termed the data lying at a distance from the corpus of data as “influences.” He displayed an understanding that these values were weighted, in that they exerted a greater pull on a central value than did a value close to the main body of data. Hence his measures of typicality resembled an attempt to visually balance the data.

Yan imposed one restriction when constructing eyeball estimates of representative values: That the typical value be literally represented in the data set (see Task 9, Appendix B).

T: Typically how many times do you think that Jett got reprimanded per day?
Y: Probably 7 or 14
T: If I said to you somebody else did it and they picked 10. Do you think that
Y: (interrupts) No No. Because there aren’t any even on 10. *Yan laughs*
T: Okay so that is not something you’d do
Y: Yeah no way

Yan was the only student in the clinical interview to employ the mean as a measure to represent data presented on a line plot. Interestingly, in later teaching episodes on many occasions the mean value was a value not represented in the data set, however Yan did not display any discomfort choosing the non represented calculated mean value as the typical value contrary to his behavior in the teaching episode described above when making an eyeball estimate. The researcher presented him with his thoughts on the same subject (Task 10, Appendix B).

T: Typically what does he score in his exam?
Y: . . .. 7 or 8. Because that’s the average.. mean. Yeah 7.
T: Oh okay and how did you get that (value)?
Y: By looking at them I guess and balancing the set.
T: Another week I asked you for a typical score and you told me that it couldn’t be a value that wasn’t there. Do you remember that? *Researcher presents the data set to which she is referring.*
Y: No. Well yeah it can. Because.. wait um . . .. I don’t think it can because if you do the average it could be like 2.5 and you don’t have 2.5 so.. yeah it can yeah it can yeah.
T: It can? Because?
Y: The average mightn’t be a value but can still be the right answer. So I suppose you could pick a value that is not actually there.. like a middle value or something. Yeah.. as long as the value looked right you could pick it.

Throughout the teaching experiment Yan utilized a mean-in-combination approach – always examining the data visually when constructing a typical value. The variability and center of the distributions were features that he attempted to coordinate in his constructions of distributional indices. He did not calculate a mean value on each occasion when asked to construct a distributional index, rather, he used mean values when he was having difficulty deciding on a typical value to choose. He always commented on the presence of outliers but did not always adjust typical values due to their presence. Usually, when the outlier was at a distance from a large body of data, he contemplated whether the outlying value should be central to his deliberations or not. Yan described this as the tendency of the outlier to ‘hurt the whole graph’ as is demonstrated in the transcript below (see Task 3, Appendix B).

T: So the question is typically how long do you think a 5th grader can hold their breath?
Y: Eh 11 seconds yeah
T: Why?
Y: Cause it is like the most. Of all of them. So the typical is 11.
T: What about the person at 20 seconds?
Y: It must be the teacher. *Yan laughs*
T: So you think your score of 11 represents them all?
Y: Eh well.. they’re not quite a lot when you look at the graph, but Um I don’t think it really hurts the whole graph

5.3.4. Case 4: Rachael

Rachael’s preliminary measures of typicality reflected the modal values of the distributions (Category 2). Rachael chose to disregard the presence of outliers referring to them as “non typical” and paid no attention to the skew of the data set when choosing typical values. This pattern was similar to that observed with Hope; therefore Rachael was presented with Category 2 refining activities. Rachael continued to choose modes as distributional indices until asked to consider the representativeness of the modal values. Referring to the notion of representativeness drew her attention to the necessity of her constructed values to embody in some fashion, the collective values of the data set. This was followed by a shift to the reporting of ranges within which the typical value would lie (Category 3). These ranges encompassed the modal value and were sufficiently wide so as to encompass a sizeable portion of the data. The shift to this approach represented an increasing awareness of the variability of a distribution. While the use of these ranges was a more appropriate strategy, Rachael continued to use ranges for an extended period of time and demonstrated reluctance to use other measures that were within her repertoire of approaches. Thus it was observed that the focus on variability was enduring and *replaced* other approaches (Category 2) rather that becoming an additional strategy. Our hypothesis was that the construction of broad ranges was partly in response to previous activities aimed at drawing awareness to the limitations of modal values (Category 2 refining activities) in representing distributions of data, and reflected an attempt on Rachael’s part not to choose a single value as representative. Our goal then was to design tasks that supported the incorporation of greater specificity in constructed measures of typicality and to support the coordination of measures of variability with measures of center in construction of typicality. A description of the types of tasks is presented in Fig. 3 (see Category 3 refining activities).

A retrospective analysis of Rachael’s data that was carried out at the end of teaching episode 4 indicated that Rachael had yet to use the median or mean as a representative value for the data set. Our hypothesis was that the use of the median was absent primarily due to Rachael's focus on frequencies. Up until this stage representative values were constructed based on frequencies alone (exclusive use of the mode), or some coordination of frequency with variation (adjustment from the mode or points within intervals). Rachael had also used ranges but had tended to locate these ranges around frequencies of values. Use of the median as a representative value was set as a goal, and we engaged in the demonstration of scenarios wherein the median would be an expedient value. Admittedly, the result was the construction of highly manipulated platykurtic data sets (Category 6 creating activities, for example task 5 Appendix B). When such data were evenly dispersed Rachael chose the median value. When data were unevenly dispersed she adjusted her median value in the direction of outlying values. This was the first sign of her increasing awareness of outlying values, and an indication that the concept of balance could be used fruitfully in discussing the arithmetic average.

As Rachael developed an understanding of the function and utility of the mean, she realized that it could be used to describe the distributions of data. She began to use the mean initially to index properties of distributions with outliers. Despite her insistence that the outliers were "not relevant," she started to calculate the mean as a technique to double check her eyeball estimates ensuring the outlying value was not pulling the mean upwards.

Rachael slowly converged on using one general strategy in constructing distributional indices. She started by making an eyeball estimate of the center of the data set, and then she calculated the mean in an attempt to verify her estimate. Through a process of guess-and-check she came to the realization that, in symmetric distributions of data, the mean represented all the values of the distribution and she realigned her strategies in accordance with this knowledge. When presented with non-normal and skewed distribution she initially calculated the mean and compared it to the mode of the data set (Category 8). As she was presented with more skewed data sets she soon realized that the mode could be quite different from the mean of the data set, and she began to refer to the mode less frequently in relation to typical values.

5.3.5. Case 5: Jason

Jason's initial distributional indices were modal (Category 2). As Jason was presented with skewed data sets, he was questioned about the suitability of the mode in embodying all the values of the given distribution. Hence situations requiring Jason to reflect on the representativeness of the mode (Category 2 refining activities) facilitated him in realizing that the mode was not representative and he subsequently suggested that the mean would be more representative of the values (see Task 2, Appendix B).

T: What is the typical number of medals won by a country?
J: One. 'Cause that has the most number of medals for it
T: Do you think that the value represents the number of medals won by every country?
J: No. Because those countries have 25 out here and some countries have 8 and 9.
T: Okay.. how could you explain or give a value that would represent everything? Could you do that?
J: Eh.. if you took an average* maybe.
(*note: conversation with John following this task revealed that the term 'average' implied 'mean' for Jason)

This use of the mean as a representative value occurred at the end of teaching episode 2. Jason's use of the mean was somewhat surprising as we had expected that he would construct ranges or modal clumps given the direction of Hope and Rachael's trajectories. We were also aware that Jason had not demonstrated conceptual understanding of the mean in the initial clinical interview, thereby making his response somewhat unexpected.

As time progressed Jason developed a strategy of calculating two measures of central tendency, and using these values as a means to determine the representative or typical value for the distribution of data (Category 8). Initially Jason consistently used the mode and one other measure (either median or mean) to construct the typical value. As Jason came to realize the disparity between the mode and the other measures of central tendency (except in cases of normal distributions) he began to calculate all three measures of central tendency and use a comparison of all three measures. This strategy of using all three measures persisted across several tasks.

Analyses of Jason's strategies indicated that he tended not to refer to the variability of data sets when constructing distributional indices. One hypothesis was that Jason's overemphasis on the use of the mean (and other measures of central tendency) was as a result of the time we devoted to developing conceptual understanding of the mean. Another explanation for the prominence of the mean is that the mean is a historically relevant practice in school mathematics, thus students see it as a valuable measure. Therefore the activities in the final teaching episodes endeavored to help Jason remarry the use of measures of variation with measures of center (Category 8 creating tasks). Hence the final teaching episodes consisted of presenting Jason with tasks designed to draw his awareness to the limitations of measures of central tendency as exclusive distributional indices. It was as a result of constructing different data sets each having the same means that Jason came to the realization that the mean, used in isolation, does not divulge an ample amount of information about the distribution (Tasks 12 and 13, Appendix B). Jason constructed multiple data sets having identical means, medians and modes and concluded that these measures describe the center of the distribution but do not render information regarding the range of the distribution. In one such activity Jason was presented with sports statistics from a newspaper article stating that a prominent basketball player had scored an average of 80 points in his past 5 games. Jason was asked to construct a distribution of data to reflect possible scores in the past 5 games. He was then asked to another possible distribution using different values. Finally he was presented with the actual statistics and asked to compare them with his constructed distributions.

Despite Jason's affinity toward using the mean as a representative value these activities supported him in developing an awareness of the limitations of the mean as a representative measure. It was at this junction that Jason started to use the range in addition to measures of central tendency to index distributions of values.

In summary, students' thinking progressed from earlier categories which focused on idiosyncratic notions, to notions of the mode, to more sophisticated adjustments to take account for outliers, to indices of center like the median and mean. This progression stemmed from and capitalized on initial orientations of students to attempt to make sense of variability. The use of outliers and data sets with ambiguous modes proved helpful in stimulating children to move to quantitative indices of distribution, but over-emphasis on the importance of the mean proved detrimental. In our cases explicit attempts to reconnect – to remarry the mean as an index of center (i.e. location) with ranges, clumps, and other indices of variability was critical in establishing a more flexible, analytical understanding.

6. Conclusion

The central contribution of this study is the development of an initial plausible model of students' learning trajectories when constructing indices that represent typicality. Other research has provided insight into measures of typicality constructed by students at *one static point in time.* The model presented in this study provides researchers and teachers with a preliminary snapshot of the *sequence* in which students develop representative indices, thus providing us with some insight as to which statistical skills are building blocks for others. More importantly, however, is the concurrent provision of tasks and instructional strategies that occasion the development of students' reasoning in more or less fruitful directions toward a flexible understanding of distribution and its associated indices of center and spread. If we consider these learning trajectories hand-in-hand with the allied activities and tasks that challenged and altered children's thinking we then have a plausible first model of how we can stimulate and promote the decision making processes used to develop and choose appropriate indices for the great variety of forms data sets take on.

Examining the profiles of student's individual learning trajectories illustrates some interesting features. The pilot study students, and Hope, Rachael and Jason's developmental trajectories were initiated by the use of the mode as a typical value. This less representative measure was the primary approach utilized by 6 of the 8 students. Interestingly, these students represented a span of 4 school grades and an age range of 5 years 6 months. Hence, despite the maturation of the older students and their greater experiences in school mathematics and statistics, they did not demonstrate greater sophistication in their strategies in the initial clinical interview. This implies that notions of representativeness may not develop unless specifically addressed in instruction. In the main study, students generally seemed satisfied with the notion of the mode as representing a typical value even when distributions were ambiguous. In many cases this approach was demonstrated across tasks and teaching episodes, and a change in approach was not evident until students' attention was drawn to the notion of representativeness. When asked a question such as "is your value representative of all the values in the data set," students re-evaluated their strategies and moved to constructing a value that in some way represented the greater body of data. Hence it seems that for these students the term "typicality" does not embody the notion of "representativeness."

An interesting approach emerged once students attempted to embody representativeness within their typical values. Many moved to choosing a "modal range" in which the mode was situated within a large clump or cluster of data. This represented a move from "mode as typical" to "mode as representative" of values. This strategy of situating representative values within clusters or clumps of data has been used by 3rd graders (Konold & Higgins, 2003), 4th graders (Lehrer & Schauble, 2002), 7th graders (Cobb, 1999) and in situated nursing practices (Noss, Pozzi, & Hoyles, 1999). The location of modes within these clusters of data represented an intermediate stage of coordinating typicality with variability. This lends to the argument that variation and clumping are psychological primitives of variability and center because they are tied to specific data, and are not general forms like mean and variance. Konold and Higgins (2003) caution that the use of modal clumps may, at times, merely represent the construction of an extended mode, in other words identifying the most common values. While this certainly is a possibility, it is unlikely that participants in this study possessed this limited conceptualization, as the associated justifications for construction and use of these modal clumps as representative values made reference to densities, therefore indicating attempts to locate variability and center.

The establishment of learning trajectories provides a framework from which we can interpret individual students' understanding of distribution, and it furnishes direction and guidance in how to facilitate the development of a rich conceptual understanding of distribution. Examination of the trajectories highlights the similarities in approaches used by our students of various ages and grade levels and forms a reasonably coherent picture of how individual statistics may contribute to children's notion of the structure of data.

Savannah was the only student who persisted in focusing on individual cases for a large part of the teaching experiment, and she displayed difficulty constructing a generalized picture of the group. The remainder of the students did not demonstrate, to any great extent, difficulties reasoning about the aggregate, as have been evidenced in other studies (Hancock et al., 1992). One possibility may be that the questions relating to the distributions required students to focus on global features of the data. Another difference was that there was little use of the midrange (locating a value mid way between the extreme values) evidenced in the study. This strategy should not be discounted as an approach used by children as other studies have identified use of the midrange fruitfully as an average (Konold & Higgins, 2003). Students in our study were aware of the extreme values of the distribution, but they tended to funnel their energies into locating values within a specified range encompassing a cluster of values. The order of values was also not a feature often attended to, indicated by the low frequency of constructed typical values that were medians. Despite the demonstrated facility of students in this study in identifying the median from a list of data values, these same students tended not to utilize the median as a representative value until presented with extremely contrived distributions. This finding reflects the outcomes of work with 7th graders carried out by Bakker (2004) who described the median as "conceptually even harder than we had expected" (p. 257). In Bakker's study he also found that in situations where facility with the median was evident, the generated understandings of the median were heavily dependent on the representation and contexts that students were used to. For example, students who could identify the median using a particular mini-tool (where the focus was on median as two equal groups) were unable to find the median of a list of numbers, leading to the conclusion that the difficulty of the median is heavily tied to the representations used.

The trajectories revealed in this study extend the body of work in this field in that they provide detail, both contextual and mathematical, regarding key transitional thinking and the kinds of tasks and prompts that support advancement through the model, thus helping children advance from idiosyncratic to analytical forms of statistical thinking. This category of (teaching) detail may prove useful for the design of instructional sequences that are tailored to the particulars of understanding center and spread, representativeness of measures, and a flexible means of describing and manipulating complex data sets.

Acknowledgement

We would like to thank Dr. Finbarr Sloane for his comments on this manuscript.

Appendix A. Clinical interview tasks

Task 1: Exercise and heart rates

The following values are the heart rates of a class of 6th graders following exercising for 30 s: 72, 73, 73, 73, 74, 76, 77, 78, 78, 79, 79, 79, 80, 80, 81, 81, 82, 83, 83, 83, 83, 84, 85, 85, 85, 87, 88, 89. Draw a line plot of the data.

Task 2: A walk in the jungle (picture of snakes are presented with the task)

Here are the pictures of 27 snakes commonly found in South East Asia. The numbers written on the bottom right hand corner of the pictures are the lengths, in feet, of the fully-grown adults. The lengths have been graphed on the line plot below.

						X				
						X				
						X	X			
						X	X			
			X			X	X	X		
	X		X	X	X	X	X	X	X	
	X	X	X	X	X	X	X	X	X	
3	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*

a) What does the graph tell you about the lengths of snakes in South East Asia?

b) If you were walking in the jungle and came upon a snake, what length would you expect the snake to be?

Task 3: Books, books, and more books

Susan and Anne had a contest to see who could read the most books over the 9 week summer holidays. The following are the number of books that each person read. Who do you think read the most books? Why do you think so?

Susan:	*18*	*16*	*13*	*15*	*5*	*13*	*11*	*0*	*8*
Anne:	*20*	*10*	*6*	*8*	*15*	*6*	*17*	*10*	*7*

Task 4: Test scores

The following are the scores that Kate achieved on her last 11 math tests. The highest score that Kate can get in any one test is a score of 10. Her scores are: 9, 3, 8, 7, 10, 5, 9, 4, 7, 8, 7

Calculate the mean, median and mode of Kate's test scores.

Task 5: Rabbit lifespan

John has a pet rabbit that is 5 years old. He wonders if his rabbit is old compared to other rabbits. At the pet store, he finds out that the mean age for a rabbit is 7 years.

a) What does the mean tell John about the life span for a rabbit?

b) If there were ten rabbits, what are the possible life spans for each rabbit, if the mean life span is 7 years?

c) What additional information would help John predict the life span of his rabbit?

Task 6: Smoking cigarettes

The World Health Organization (WHO) is going to launch a campaign highlighting the dangers of smoking. One recent research study found that in several countries men smoke significantly more cigarettes per day than do women. The organization is concerned with whether, in the United States, it should aim its advertising campaign at both men and women or at men alone. They hired an organization to carry out a survey of the number of cigarettes that men and women smoke per day. If they find that men smoke more per day then the WHO will aim its campaign at men alone. The following are the results of the survey carried out in Phoenix on the first day of the study. 25 men and 25 women were surveyed.

			X						X				X		X	X			X			X		
			X	X	X	X	X		X	X		X	X	X	X	X	X		X	X	X	X	X	
1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*	*18*	*19*	*20*	*21*	*22*	*23*	*24*	*25*

Number of cigarettes smoked per day by ***men***

									X										X					
							X		X				X			X			X			X		
			X	X	X		X		X	X	X	X	X		X	X	X	X	X	X	X	X		
1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*	*18*	*19*	*20*	*21*	*22*	*23*	*24*	*25*

Number of cigarettes smoked per day by ***women***

Would you recommend that the WHO aim its advertising campaign at both men and women or at men alone? Why?

Appendix B. Teaching episode tasks

Task 1: Number of 'Gummi Bears' in a packet

A class of students were interested in examining the number of Gummi Bears in a packet. They each counted the number of bears in a packet and put their results on a line plot.

a) What does the graph tell you about the number of Gummi bears students counted?

b) From examining the graph, generally how many Gummi Bears would you expect to find in a packet of Gummi Bears? Why? Here are their results:

			X							
	X		X							
	X		X							
X	X	X	X	X	X		X			
X	X	X	X	X	X	X	X	X		
7	8	9	10	11	12	13	14	15	16	17

Task 2: The Winter Olympics problem [Adapted from Landwehr and Watkins (1986)]

The 1984 Winter Olympics were held in Sarajevo, Yugoslavia. The data were organized into a line plot.

a) Describe the data.

b) What is the typical number of medals won by a country?

X																								
X																								
X	X		X				X																	
X	X	X	X	X	X		X	X				X											X	X
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25

Task 3: How long can you hold your breath?

Students in a fifth grade classroom wanted to investigate the typical length of time a fifth grade student could hold his/her breath. They each found a partner. One member of each pair of students held his/her breath while the other person recorded the length of time on a watch. They then repeated the process for the other person. The student's recorded their results. The student's constructed a line plot of their results. If a new student were to enter the class, how long do you think he/she could hold his/her breath? Why do you choose this number?

				X										
		X		X	X									
		X	X	X	X									
	X	X	X	X	X	X		X					X	
7	8	9	10	11	12	13	14	15	16	17	18	19	20	21

Task 4: Heights of the Basketball Team

On school sports day, Miss Murphy's 5th grade basketball team played a game of basketball against Mr. Cody's basketball team. Miss Murphy's team won the game.

Jamie, a student in Mr. Cody's class, believes that the game was not fair. She believes that students in Miss Murphy's class are taller than students in Mr. Cody's class. So, when choosing a basketball team, Miss Murphy has a lot of tall students from which to choose. Jamie decided to measure the height of each student in both classes and construct a line plot for each class. Examine the graphs.

a) What is the height of a student in Miss Murphy's class?

b) What is the height of a student in Mr. Cody's class?

c) Do you agree with Jamie that the students in Miss Murphy's class are taller than the students in Mr. Cody's class?

Height of Miss Murphy's students

			X					X										
			X				X	X	X	X								
		X	X				X	X	X	X								
X		X	X	X			X	X	X	X			X	X			X	X
40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58

Height of Mr. Cody's students

				X				X				X						
				X				X				X						
				X	X			X	X			X	X					
X	X		X	X	X	X	X	X	X	X	X	X	X		X	X		
40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58

Task 5: Blowing paper clips

Examine the values for the other two students, John and Jackie. The data for John and Jackie are in the table below. The data has also been graphed on a line plot. How far can John and Jackie blow the paper clip?

Distances (in centimeters) that John blew the paper clip:	*34, 36, 28, 2, 38, 20, 42, 32, 46, 24*
Distances (in centimeters) that Jackie blew the paper clip:	*18, 4, 8, 10, 2, 0, 20, 12 6, 14*

Distance John can blow a paper clip

	X									X		X		X		X	X	X	X		X		X	
0	2	4	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34	36	38	40	42	44	46	48

Distance Jamie can blow a paper clip

X	X	X	X	X	X	X	X		X	X														
0	2	4	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34	36	38	40	42	44	46	48

Task 6: Performance in statistics

Students in a college statistics class are asked to assess their performance in the class over the past semester. Each student has the scores that they received on each of the 17 weekly quizzes they completed over the course of the semester. One student, however, has difficulty assessing his general performance for the class. The following is a line plot of his scores. Based on the information in the graph, what do you think is the grade that the student should give himself? Why?

																								X
																								X
																	X				X		X	X
X	X											X				X	X		X	X	X	X	X	X
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25

Task 7: Hourly pay

A class of 8th graders wanted to carry out a survey of the hourly wages that people in their community were paid. So, the students went to the local mall and asked 40 people how much money they were paid per hour in their jobs. The 8th graders graphed the data on a line plot. What do you think is the typical hourly wage paid to people living in that community?

	X																			
	X																			
	X										X		X	X	X		X			
	X									X	X	X	X	X	X	X	X	X		
	X									X	X	X	X	X	X	X	X	X	X	
	X									X	X	X	X	X	X	X	X	X	X	
5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25

Amount of money paid per hour worked

Task 8: Paces to the gym

The last time we met you and I investigated how many paces it would take you to walk to the water fountain. In this problem the students did something similar. A class of students investigated the question of how many paces it takes to travel from their class to the cafeteria. They measured the distance by counting the number of paces each student walked. Every step made on the right foot counted as one pace. Here are the results:

		X										
		X										
		X										
		X	X									
	X	X	X					X				
X	X	X	X	X	X			X	X			
X	X	X	X	X	X	X		X	X			X
26	27	28	29	30	31	32	33	34	35	36	37	38

Describe the set of data. What is the typical number of paces it takes to travel to the cafeteria? Explain your reasoning.

Task 9: Number of times the teacher calls on you per day

The following data set represents the number of times a student, Jett, was reprimanded by his teacher over a 18 day period. Each of the occasions involved the teacher calling Jett's name and asking him to either return to his seat or to pay attention to his work. Typically how many times do you think that Jett gets reprimanded per day?

		X							X		
	X	X	X					X	X	X	
X	X	X	X	X			X	X	X	X	X
5	6	7	8	9	10	11	12	13	14	15	16

Task 10: Statistics exam

The following are the scores that Mike got on his last 7 math exams. Each exam is scored out of 26. Typically what does Mike score in his exams?

	X		X		X		X		X		X			X											
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26

Task 11: Low fat chips

Nearly two-thirds of adults in the United States are overweight, and 30.5% are obese, according to data from the 1999–2000 National Health and Nutrition Examination Survey (NHANES). The detrimental health effects of obesity and being over weight has resulted in many food manufacturers producing low fat options of their food products. I have here a selection of 23 6oz bags of regular and low fat chips. Nutritional information for each packet is listed on the back of the bag. Represent the fat content for each bag on a line plot. Using your line plot, determine the amount of fat you would expect to eat if you ate any one of these bags of chips.

Task 12: Number of pets in a household

> *A classroom teacher gave the following homework assignment to her class of 5th graders: "A group of 8 students were surveyed to examine how many pets they had in their households. The mean number of pets was found to be 6. Draw a possible distribution of data values that contains values for each of the 8 students and that has a mean of 6 pets."*

The next day before class started Amy and Anna compared their results and found that each of them had drawn different distributions of data. Could this happen? Can you draw two different distributions each having 8 values but both having a mean of 6? How would you now describe the two distributions so that your classmates would understand that even though the means are the same, the distributions do not actually look alike?

Task 13: Wimbledon semi-final

The average speed of the five aces produced by the winner of the Wimbledon women's semi-final was 80mph.

a) What does the mean (80 mph) tell us about the set of data values?

b) Can we construct the data set so that it exactly represents the 5 speeds?

Construct a possible data set.

Can you construct a different data set?

References

Agresti, A. (1996). *An introduction to categorical data analysis*. New York: Wiley.

Bakker, A. (2004). *Design research in statistics education; On symbolizing and computer tools*. Utrecht, the Netherlands: CD Beta Press.

Bakker, A., Biehler, R., & Konold, C. (2005). Should young students learn about boxplots? In G. Burrill (Ed.), *Curricular development in statistics education*. International Association for Statistical Education (IASE) Roundtable, Lund, Sweden, 28 June to 3 July, pp. 163–173, http://www.stat.auckland.ac.nz/~iase/publications.

Bakker, A., & Gravemeijer, K. P. E. (2004). Learning to reason about distribution. In D. Ben-Zvi, & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 147–168). Dordrecht, the Netherlands: Kluwer Academic Publishers.

Ben-Zvi, D., & Arcavi, A. (2001). Junior high school students' construction of global views of data and data representations. *Educational Studies in Mathematics*, *45*, 35–65.

Ben-Zvi, D., & Sharett-Amir, Y. (2005). How do primary school students begin to reason about distributions? In K. Makar (Ed.), *Reasoning about distribution: A collection of current research studies. Proceedings of the fourth international research forum on statistical reasoning, thinking, and literacy (SRTL-4)* University of Auckland, New Zealand, 2–7 July 2005,. Brisbane: University of Queensland.

Cai, J. (1998). Exploring students' conceptual understanding of the averaging algorithm. *School Science and Mathematics*, *98*, 93–98.

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.

Clark, F. B., & Kamii, C. (1996). Identification of multiplicative thinking in children in grades 1–5. *Journal for Research in Mathematics Education*, *27*(1), 41–51.

Clement, J. (2000). Analysis of clinical interviews: Foundations and model viability. In A. Kelly, & R. Lesh (Eds.), *Handbook of research design in mathematics and science education* (pp. 547–590). Mahwah, NJ: Erlbaum.

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

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., et al. (2007). *Guidelines and assessment for instruction in statistics education (GAISE) report: A pre-K-12 curriculum framework*. Alexandria, VA: American Statistical Association.

Friel, S. N., Curcio, F., & Bright, G. (2001). Making sense of graphs: Critical factors influencing comprehension and instructional implications. *Journal for Research in Mathematics Education*, *32*(2), 124–158.

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

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

Jones, G. A., Langrall, C. W., Thornton, C. A., Mooney, E. S., Wares, A., Perry, B., et al. (2001). Using students statistical thinking to inform instruction. *Journal of Mathematical Behavior*, *20*, 109–144.

Konold, C., & Higgins, T. (2003). Reasoning about data. In J. Kilpatrick, G. Martin, & D. Schifter (Eds.), *A research companion to NCTM's standards*. Reston, VA: NCTM.

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., Robinson, A., Khalil, K., Pollatsek, A., Well, A., Wing, R., et al. (2002). *Students' use of modal clumps to summarize data*. Paper presented at the Sixth International Conference on Teaching Statistics, Cape Town, South Africa.

Landwehr, J. M., & Watkins, A. E. (1986). *Exploring data: Quantitative literacy series*. Dale Seymour Publications.

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

Lehrer, R., & Schauble, L. (2002). Inventing data structures for representational purposes: Elementary grade students' classification models. *Mathematical Thinking and Learning*, *2*, 51–74.

McClain, K., & Cobb, P. (2001). Supporting students' ability to reason about data. *Educational Studies in Mathematics*, *45*, 103–129.

McGatha, M., Cobb, P., & McClain, K. (2002). An analysis of students' initial statistical understandings: Developing a conjectured learning trajectory. *Journal of Mathematical Behavior*, *16*(3), 339–355.

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 Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

Noss, R., Pozzi, S., & Hoyles, C. (1999). Touching epistemologies: Meanings of average and variation in nursing practice. *Educational Studies in Mathematics*, *40*, 25–51.

Piaget, J. (1963). *The child's conception of the world*. New Jersey: Littlefields, Adams.

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

Reading, C., & Pegg, J. (1996). Exploring understanding of data reduction. In L. Puig, & A. Guttierrez (Eds.), *Proceedings of the 20th conference of the international group for the psychology of mathematics education* Vol. 4, Valencia, Spain, (pp. 187–194).

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

Siegler, R. S. (1996). *Emerging minds: The process of change in children's thinking*. New York: Oxford University Press.

Siegler, R. S., & Crowley, K. (1991). The microgenetic model. A direct means for studying cognitive development. *American Psychologist*, *46*(6), 606–620.

Simon, M. A. (1995). Reconstructing mathematics pedagogy from a constructivist perspective. *Journal for Research in Mathematics Education*, *26*, 114–145.

Steffe, L. P. (1992). Schemes of action and operation involving composite units. *Learning and Individual Differences*, *4*, 259–309.

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: Erlbaum.

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.

Watson, J. M., & Kelly, B. A. (2002). *Can grade 3 students learn about variation?* Paper presented at the sixth international conference on teaching statistics, Cape Town, South Africa.