MATHEMATICAL THINKING AND LEARNING, *2*(4), 269–307

A Framework for Characterizing Children’s Statistical Thinking

Graham A. Jones, Carol A. Thornton,
Cynthia W. Langrall, and Edward S. Mooney
Department of Mathematics
Illinois State University

Bob Perry
Faculty of Education and Languages
University of Western Sydney
Macarthur, Australia

Ian J. Putt
School of Education
James Cook University
Townsville, Australia

Based on a review of research and a cognitive development model (Biggs & Collis, 1991), we formulated a framework for characterizing elementary children’s statistical thinking and refined it through a validation process. The 4 constructs in this framework were describing, organizing, representing, and analyzing and interpreting data. For each construct, we hypothesized 4 thinking levels, which represent a continuum from idiosyncratic to analytic reasoning. We developed statistical thinking descriptors for each level and construct and used these to design an interview protocol. We refined and validated the framework using data from protocols of 20 target students in Grades 1 through 5. Results of the study confirm that children’s statistical thinking can be described according to the 4 framework levels and that the framework provides a coherent picture of children’s thinking, in that 80% of them exhibited thinking that was stable on at least 3 constructs. The framework contributes do-

---

Requests for reprints should be sent to Graham A. Jones, Department of Mathematics 4520, Illinois State University, Normal, IL 61790–4520. E-mail: jones@math.ilstu.edu

main-specific theory for characterizing children’s statistical thinking and for planning instruction in data handling.

In response to the critical role that information and data play in our technological society, there have been international calls for reform in statistical education at all grade levels (Australian Education Council, 1994; National Council of Teachers of Mathematics, 1989, 2000; School Curriculum and Assessment Authority & Curriculum and Assessment Authority for Wales, 1996). These calls for reform have advocated a more pervasive approach to the study of statistics, one that includes describing, organizing, representing, and interpreting data. This broadened perspective has created the need for further research on the learning and teaching of statistics, especially in the elementary grades, where there has been a tendency to focus narrowly on graphing rather than on broader topics of data handling and data analysis (Shaughnessy, Garfield, & Greer, 1996).

Although some elements of students’ statistical thinking and learning have been investigated (Bright & Friel, 1998; Cobb, 1999; Curcio, 1987; De Lange, van Reeuwijk, Burrill, & Romberg, 1993; Gal & Garfield, 1997; Konold, Pollatsek, Well, & Gagnon, 1997; Lehrer & Romberg, 1996; Mokros & Russell, 1995), research on students’ statistical thinking is emergent rather than well established. Existing research on students’ statistical thinking has not yet developed the kind of detailed cognitive models of students’ learning that researchers like Cobb et al. (1991) and Resnick (1983) deemed necessary to guide the planning and development of mathematics curriculum and instruction. In essence, although research has developed coherent frameworks to describe students’ thinking for mathematical topics such as whole numbers (Carpenter & Moser, 1984), rational numbers (Lamon, 1993; Mack, 1990), geometry (van Hiele, 1959/1985), and probability (Jones, Langrall, Thornton, & Mogill, 1997), this has not happened for students’ statistical thinking. We addressed that void in this study by developing and validating a framework for characterizing elementary school (Grades 1–5) students’ statistical thinking in a way that will be appropriate for informing assessment and instruction.

AIMS OF THIS RESEARCH

The aims of this research were threefold. First, through synthesis of existing research on statistical thinking, in this study we sought to develop an initial framework for characterizing how students think in statistical situations. This framework was situated in a general cognitive development model (Biggs & Collis, 1991). Second, in this study, we set out to generate assessment protocols based on this initial framework. Third, based on observations and insights of children’s statistical thinking, we sought to refine and validate the initial framework using the assessment protocols and case-study analysis.

THEORETICAL CONSIDERATIONS

The thesis of this study is that, for students to exhibit statistical thinking, there is a need for them to understand data-handling concepts that are multifaceted and develop over time. Shaughnessy et al. (1996) maintained that data handling incorporates "organizing, describing, representing, and analyzing data, with a heavy reliance on visual displays such as diagrams, graphs, charts, and plots" (p. 205). In generating our initial statistical thinking framework (initial framework; see Table 1), we adopted the four constructs identified by Shaughnessy et al. in a slightly modified form: describing, organizing and reducing, representing, and analyzing and interpreting data.

Our initial framework was also formulated on the basis that elementary children will exhibit four levels of statistical thinking in accord with Biggs and Collis's (1982, 1991) general developmental model, Structure of the Observed Learning Outcome (SOLO). These levels of statistical thinking, described in our framework as *idiosyncratic, transitional, quantitative,* and *analytical,* are elaborated in the following.

The Four Constructs

Each of the four major constructs in our initial framework is described and amplified through reference to research on students' statistical thinking. We also use research on each construct to identify the kinds of tasks that were employed to assess children's statistical thinking in this study.

Describing Data

Our first construct incorporates what Curcio (1987; Curcio & Artz, 1997) called *reading the data*. Curcio noted that reading the data means extracting information explicitly stated in the display, recognizing graphical conventions, and making direct connections between the original data and the display. Hence, describing data refers to explicit reading of the data and should not be confused with Curcio's terms *reading between the data* and *reading beyond the data*. These latter notions are incorporated in our fourth construct, analyzing and interpreting data, and are discussed later.

Major studies that have addressed children's thinking vis-à-vis describing data include Beaton et al. (1996) and Pereira-Mendoza and Mellor (1991). Results from these studies indicate that elementary students, especially at the upper grade levels, can read data displays. Pereira-Mendoza and Mellor, for example, obtained correct descriptions of bar graphs from 95% of the fourth graders in their study, and Beaton et al. found that over 75% of the third- and fourth-grade students in the Third International Mathematics and Science Study (TIMSS) could read data di-

TABLE 1
Initial Statistical Thinking Framework

Construct	*Level 1: Idiosyncratic*	*Level 2: Transitional*	*Level 3: Quantitative*	*Level 4: Analytical*
Describing data displays (D)	In reading the data, demonstrates little awareness of the data and graphing conventions (e.g., title, axis labels)	In reading the data, demonstrates some awareness of data and graphing conventions	In reading the data, generally demonstrates complete awareness of the data and graphing conventions	In reading the data, demonstrates complete awareness of the data and graphing conventions
	Recognizes when different displays represent the same data by giving a justification based on subjective judgments	Recognizes when different displays represent the same data by giving a justification based on cosmetic features	Recognizes when different displays represent the same data by establishing partial relations between the displays	Recognizes when different displays represent the same data by establishing precise numerical relations between the displays
	Considers irrelevant or subjective features when evaluating the effectiveness of different displays of the same data set	Focuses only on one aspect when evaluating the effectiveness of different displays of the same data set	Focuses on more than one aspect when evaluating the effectiveness of different displays of the same data set	Provides a coherent and comprehensive explanation when evaluating the effectiveness of different displays of the same data set
Organizing and reducing data (O)	Does not attempt to group data into classes	Groups data into classes on the basis of criteria that the student may not be able to explain	Groups data into classes and can explain the basis for this grouping	Can group the data into classes in more than one way and can explain the basis for these different groupings
	Does not attempt to order data	Orders data but may not be able to explain the value of doing this	Orders data and is able to explain the value of doing this	Can order the data in different ways and can explain the basis for these different orderings

	Gives a subjective description or is not able to describe data in terms of "typicality"	Describes "typicality" of data using invented measures that are partially valid	Describes "typicality" of data using the mode or invented measures that are valid	Describes "typicality" of data in terms of common measures of center such as the median or the mean
	Gives a subjective description or is not able to describe data in terms of spread	Describes spread of data using invented measures that are partially valid	Describes spread of data using invented measures that are valid	Describes spread of data in terms of a common measure such as the range
Representing data (R)	Constructs an idiosyncratic or invalid display when asked to complete a partially constructed graph associated with a given data set	Constructs a display that is valid in some aspects when asked to complete a partially constructed graph associated with a given data set	Constructs a valid display when asked to complete a partially constructed graph associated with a given data set	Constructs a valid display and justifies the procedure when asked to complete a partially constructed graph associated with a given data set
	Produces an idiosyncratic display that does not represent the data set	Produces a display that represents the data but does not attempt to reorganize the data	Produces a display that shows some attempt to reorganize the data	Produces multiple valid displays some of which reorganize the data
Analyzing and interpreting data (A)	Makes no response or gives an irrelevant response to questions on reading "between the data" and "beyond the data"	Compares data within the display or data set ("reading between the data")	Makes inferences concerning the variables beyond the scope of the data present ("reading beyond the data")	Makes inferences (about two or more data sets or displays) that involve reading between the data or reading beyond the data
	Makes no response or an irrelevant response to the question, "What does the display *not* say about the data?"	Makes a relevant but limited response to the question, "What does the display *not* say about the data?"	Makes multiple relevant responses to the question, "What does the display *not* say about the data?"	Makes a comprehensive contextual response to the question, "What does the display *not* say about the data?"

rectly from a bar graph. Further, approximately 40% of the TIMSS third-grade students and 70% of the fourth-grade students could read data from a pictograph.

Based on our definition of *describing data* and a synthesis of the aforementioned research, we identified the key elements in describing data as follows: (a) reading data displays, (b) showing awareness of elemental graphing conventions (e.g., title, axis labels), (c) recognizing when different displays represent the same data, and (d) evaluating different displays of the same data. In accord with these key elements, we generated clusters of questions or tasks like the following to assess children's thinking on this construct:

- What does this picture tell you?
- Do you think these pictures represent the same data?
- Which one of these pictures would be more useful?

We recognize that a question like "What does this picture tell you?" is open-ended. However, we wanted to assess the limits of children's thinking in relation to describing data, and we were aware that we could use follow-up questions to probe and focus their thinking.

Organizing and Reducing Data

This construct incorporates mental actions such as ordering, grouping, and summarizing data (Moore, 1997). As such, it also involves reducing data using notions of center and spread.

Research on organizing data at the elementary school level is virtually nonexistent (Cobb, 1999, p. 5). However, based on their interviews with middle school students, Bright and Friel (1998) reported that students experience great difficulty in making connections between ungrouped visual displays and grouped visual displays of the same data.

Most of the available research on elementary schoolchildren's thinking in data reduction has focused on their understanding and use of measures of center, particularly the mean. A seminal study by Strauss and Bichler (1988) involving 8-, 10-, 12-, and 14-year-olds examined students' understanding of seven properties of the arithmetic mean. With respect to the two younger groups, these researchers found that approximately one half of the 8-year-olds and almost all the 10-year-olds understood that the average is located between extreme values. Further, Strauss and Bichler found that nearly all of the students in these two groups realized that the average is influenced by values in the data set and that the average does not necessarily equal one of the actual data values. However, less than 15% of these students understood that a value of zero must be taken into account when finding the mean.

In their study of fourth-, sixth-, and eighth-grade students' concept of average in problem-solving situations, Mokros and Russell (1995) found that younger students in their study interpreted *average* as the value that occurred with greatest frequency (mode). Mokros and Russell suggested that students who consider the average as a mode are at a point in which they see the data set as individual values rather than as an entity that can be represented by a single value. Bright and Friel (1998) claimed that students use a mode to describe the typicalness of data in a display because the mode is easily identifiable in a graph. Some of the younger students in Mokros and Russell's study considered the average to be a data point roughly centered within the data, that is, these students had begun to conceptualize the notion of median. The idea of average as a mathematical point of balance, a vital characteristic of the mean, did not emerge in the thinking of younger students in the Mokros and Russell study. Further, Mokros and Russell found that younger students did not use an algorithmic procedure to determine the average even when they were aware of the procedure.

Based on our definition of *organizing and reducing data* and our review of research, we identified the key elements of this construct as follows: (a) grouping and ordering data, (b) recognizing that information may be "lost" in a reorganization of data, (c) describing data in terms of representativeness or typicality, and (d) describing data in terms of spread. In accord with these key elements, we generated clusters of questions or tasks like the following to assess children's thinking on this construct:

- How would you organize this data in another way?
- Can you tell in this graph, with a different organization of the data, how many Beanie Babies® Mary had?
- About how many friends came to visit?
- What's the typical number of friends who came to visit?
- What is the average number of friends who came to visit?
- Which of these sets of scores have the greatest spread, or do they have the same spread?

Note that, in determining children's understanding of center (average), we consistently asked three variations of the same question: (a) one that asked "about how many?" (b) one that used the term *typical,* and (c) one that used the term *average.* We were aware that young children might not recognize the term *average,* and we thought that these variations might reveal the full extent of children's knowledge of center.

Representing Data

Constructing representations and visual displays that exhibit different organizations of a data set is central to this construct in our initial framework. It also involves certain elemental conventions that are associated with the presentation of visual displays.

Younger elementary students have difficulty creating visual displays of data sets. In the TIMSS study, only 30% of third-grade students and 55% of fourth-grade students could complete a bar graph using data from a table (Beaton et al., 1996). In the National Assessment of Educational Progress (NAEP; Zawojewski & Heckman, 1997), 60% of the fourth-grade students could complete a bar graph or pictograph using given data, but there was less success (33% and 42%, respectively) when students had to organize the data needed to complete a bar graph or a pictograph. In an exploratory study, Curcio and Folkson (1996) asked kindergarten children to represent the number of "*p* words" in a literature story. Four types of representations emerged: Some children wrote the *p* word every time they saw it, others wrote numerals as they counted the *p* words, others used tally marks, and some did not represent anything on paper except the final number.

Based on our definition of *representing data* and a synthesis of the research previously discussed, we identified the key elements in representing data: (a) completing a partially constructed data display and (b) constructing displays to represent different organizations of a data set. In accord with these key elements, we generated questions or tasks like the following to identify children's thinking on this construct:

- Complete this graph that has been started by your teacher.
- How would you organize and present this data in another way?

Again, we wanted to determine the limits of children's knowledge in representing data and, in particular, to ascertain whether they would show any tendencies to represent different organizations of the same data.

Analyzing and Interpreting Data

This construct incorporates recognizing patterns, trends, and exceptions in data and making inferences and predictions from the data. It includes what Curcio (1987) referred to as "reading between the data" and "reading beyond the data" (p. 384). For Curcio (1989), reading between the data requires students to compare quantities and use mathematical concepts and operations to combine and integrate data. Reading beyond the data requires students to predict or infer from the data by tapping their existing schemata for information that is neither explicitly nor implicitly stated in the visual representation (Curcio, 1989, p. 6).

Elementary school students have difficulty analyzing and interpreting data. Putt et al. (1999) found that 80% of the first- and second-grade students interviewed gave responses that were superficial, incomplete, or inconsistent when they attempted to analyze data from a line plot and a bar graph. In the NAEP assessment (Zawojewski & Heckman, 1997), less than one half of the fourth-grade students could compare data in a table or interpret data in a bar graph. Similarly, Pereira-Mendoza and Mellor (1991) found that, although fourth-grade students in their study were successful at literal reading of bar graphs (over 95% success rate), they were less successful at interpreting (52% success rate) and predicting (less than 20% success rate). These researchers found that most interpretation errors could be traced to computation and reading errors, whereas prediction errors frequently arose when students were misled by the arrangement of the data or when information had to be interpolated. Fourth graders in the Pereira-Mendoza and Mellor study also had difficulty thinking beyond the data (Curcio, 1987). For example, when predicting the allowance for a child in a year beyond that presented on a graph, most students at this grade were unsuccessful because they considered the graphs to be complete.

Based on our definition of *analyzing and interpreting data* and Curcio's (1987; Curcio & Artz, 1997) research, the key elements for this construct were (a) comparing and combining data (reading between the data) and (b) extrapolating and predicting from the data (reading beyond the data). Consistent with these key elements, we generated clusters of questions or tasks like the following to assess children's thinking on this construct:

- Which day had the highest number of visitors? (compare)
- How many friends came to visit during the week? (combine)
- About how many friends would you expect to visit during the next 4-week month? (predict)
- What *doesn't* this picture tell you? (extrapolate)

In the last question, we claim that the children have to discriminate between conclusions that can and cannot be inferred from the data. We have labeled this task as extrapolation because we believe that it requires children to think beyond the given data set or display.

The Thinking Levels

Our previous research in formulating frameworks for number sense (Jones, Thornton, & Putt, 1994; Jones et al., 1996) and probability (Jones et al., 1997; Tarr & Jones, 1997) demonstrated that students' thinking in number sense and probability was consistent with the SOLO model of Biggs and Collis (1991). Consequently, in formulating our framework for this study, we hypothesized that children's statistical thinking across the four constructs could be characterized in terms of the SOLO model. This model incorporates five modes of functioning: sensorimotor (from birth), ikonic (from around 18 months), concrete symbolic (from around 6 years), formal (from around 14 years), and postformal (from around 20 years); within each mode, four levels of response (prestructural, unistructural, multistructural, and relational) represent shifts in the complexity of students' reasoning. According to Biggs and Collis, each of the five modes of functioning emerges and develops in a way that incorporates the continuing development of earlier modes. The modes that are most applicable to students in our research are the ikonic and concrete symbolic.

Based on Biggs and Collis (1991), we postulated four levels of statistical thinking (see Table 1) that corresponded to the four levels of cognitive thinking identified in the SOLO model. Hence, we hypothesized that students at Level 1 would exhibit the characteristics of the prestructural level in the sense that they would be engaged in the task but would be distracted or misled by irrelevant aspects. In fact, Biggs and Collis suggested that students at the prestructural level use both the ikonic mode and the concrete-symbolic mode, and hence, exhibit intuitive thinking that involves myths and idiosyncratic imaging. In translating this into statistical thinking, we hypothesized that students exhibiting Level 1 thinking would focus on irrelevant features when describing, representing, and analyzing data. More specifically, given their penchant for idiosyncratic imaging, we expected them to have little regard for the conventions of reading and representing data, to use subjective judgments in describing and summarizing data, and to substitute their own data for the given data when representing and analyzing data.

We hypothesized that students exhibiting Level 2 statistical thinking would exhibit the characteristics of Biggs and Collis's (1991) unistructural level in the sense that they would engage a task in a relevant way but would generally pursue only one aspect of it. We also expected that students at this level would function more consistently in the concrete-symbolic mode and, as such, would have the potential to represent their thinking using quantitative language and symbols. There would also be potential for conflict because students at this level tend to regress to ikonic intuitions and, hence, exhibit glimpses of prestructural reasoning (Watson, Collis, & Moritz, 1997). In translating the characteristics of the unistructural level into our framework, we hypothesized that students exhibiting statistical thinking at Level 2 would be in transition between idiosyncratic and quantitative thinking about data. That is, students at this level would try to represent their ideas through quantitative thinking but would generally focus on just one aspect of the data and would, at times, regress to idiosyncratic thinking. More specifically, in attempting to describe, organize, represent, and analyze data, we expected them to demonstrate some awareness of the conventions of reading and representing data, to begin to formulate invented measures for center and spread, albeit not always valid, and to try to use quantitative thinking to analyze data. Notwithstanding this move to more quantitative thinking, their representation and analysis of data would be incomplete and narrowly focused.

We hypothesized that students exhibiting Level 3 statistical thinking would exhibit the characteristics of Biggs and Collis's (1991) multistructural level in the sense that they would generally use the concrete-symbolic mode and focus on more than one relevant feature of the task. However, they would not necessarily integrate their thinking on these different features. In translating the characteristics of the multistructural level into our framework, we hypothesized that students exhibiting Level 3 statistical thinking would use informal quantitative thinking and begin to focus on more than one aspect of a data exploration task. More specifically, in attempting to describe, organize, represent, and analyze data, they would demonstrate a more complete awareness of the conventions of reading and representing data, begin to formulate more valid measures of center and spread, and tend to provide multiple responses when representing and analyzing data. However, in accord with Biggs and Collis, they would not always integrate different aspects of their analysis of data or relate their analysis to the original data context.

We hypothesized that students exhibiting Level 4 thinking would exhibit the characteristics of Biggs and Collis's (1991) relational level and would consistently function in the concrete-symbolic mode. They would not only focus on several relevant features of a task but would also make connections among these various features. In translating the characteristics of the relational level into our framework, we hypothesized that students exhibiting statistical thinking at Level 4 would demonstrate analytical and quantitative reasoning about data in the sense that they would provide multiple and logical perspectives and explanations of data situations. More specifically, in attempting to describe and organize data, they would demonstrate greater coherence in reading and organizing data and more adaptability in using and differentiating among several valid measures of center and spread. When representing and analyzing data, they would not only examine and connect multiple perspectives of the data but would also relate their analysis to the original context.

In formulating and in later refining the framework (Table 1), we attempted to achieve stability or homogeneity for a student's thinking across the four constructs. That is, we constructed the framework on the expectation that when a student's statistical thinking was at Level 2 for describing data, it would also be at Level 2 for organizing and reducing data, representing data, and analyzing and interpreting data. Stability across constructs at a particular level was consistent with our situating the framework in the SOLO model. In essence, we expected that when a student exhibited prestructural thinking in one construct, they would exhibit such thinking in all constructs. From a practical perspective, this stability feature was also intended to make the framework more user friendly when monitoring children's statistical thinking during instruction.

METHODOLOGY

The formulation and validation of the framework was an evolutionary and iterative process. We began the process by using previous research on students' statistical thinking and the SOLO general cognitive development model (Biggs & Collis,

1991) to formulate an initial framework that described children's statistical thinking according to the four cognitive levels of SOLO and across four key constructs (see Table 1). In validating the framework, our intent was to assess and refine the initial framework, to examine the stability and growth of the target students' profiles, and to characterize and interpret the four levels of statistical thinking based on the target students' thinking.

The Validation Process

The process employed to validate the initial framework was similar to that used in earlier studies (Jones et al., 1997; Jones et al., 1994; Jones et al., 1996). It involved seven components: (a) constructing the statistical thinking protocol based on the initial framework, (b) interviewing 20 target students (4 each from Grades 1–5) using the statistical thinking protocol, (c) analyzing the target students' responses to the statistical thinking protocol, (d) refining the initial descriptors of the four levels of statistical thinking, (e) determining target students' statistical thinking levels for each construct on the refined framework, (f) examining the stability of target students' thinking over the four constructs, and (g) illuminating the distinguishing characteristics of each thinking level. Qualitative methods were used to analyze the case-study data generated from the target students' responses to the statistical thinking protocol.

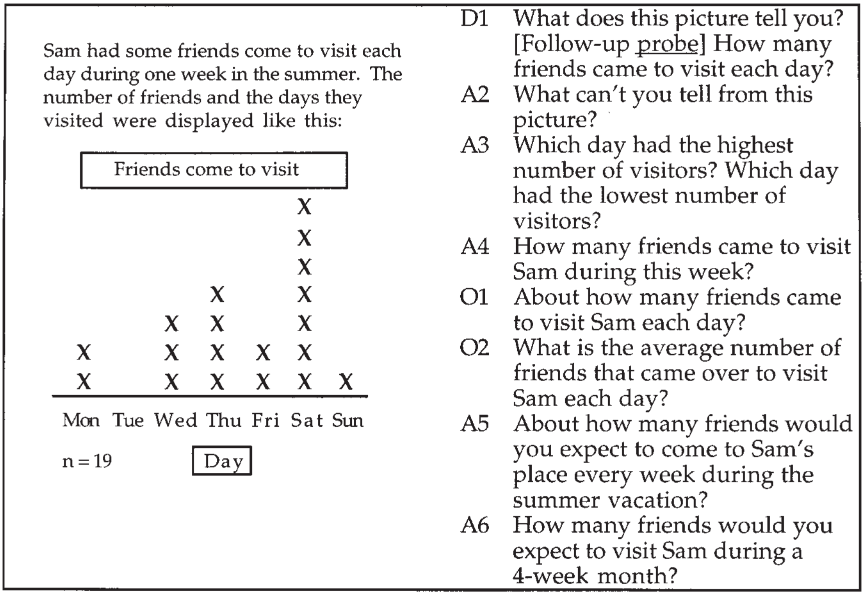
Participants

Students in Grades 1 through 5 from a midwestern U.S. school formed the population for this study. This school population was representative of a broad spectrum of socioeconomic and cultural backgrounds. Consistent with earlier validation studies (e.g., Jones et al., 1996), 20 target students, 4 from each of the five grades, were purposefully selected (Miles & Huberman, 1994) from this population based on mathematics achievement records. Two students were selected from the middle 50%, and 1 was selected from both the upper and lower quartiles at each grade level. Pseudonyms are used for all target students.

Data Collection and Validation Instrument

Data for the study were generated from the 20 target students' responses to the statistical thinking protocol. Near the beginning of the school year, two members of the research team each took responsibility for administering the protocol to 10 of the target students, with equal numbers from each grade level. Student responses were audiotaped and transcribed, and student artifacts such as drawings and graphs were collected.

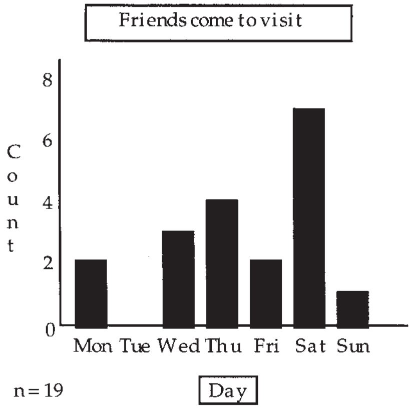
The protocol was linked to the key constructs of the initial framework and enabled the researchers to investigate children's statistical thinking across all four levels of this framework. The protocol comprised data explorations from three contexts: Sam's friends, Beanie Babies®, and a beanbag game. Each of these data explorations incorporated tasks, open-ended questions, and a series of probes on these questions. Eight questions were associated with describing data (D), nine with organizing and reducing data (O), three with representing data (R), and nine with analyzing and interpreting data (A). Ideally, we would have liked to include more questions on representing data, but this was not possible given the time-consuming nature of such tasks, especially with the younger children in the study. Sample questions for each of the data exploration contexts are presented in Figures 1 through 5. Each question is labeled *D, O, R,* or *A* according to the framework construct it assessed, and questions within the same cluster are followed by the same number.



Sam had some friends come to visit each day during one week in the summer. The number of friends and the days they visited were displayed like this:

D1 What does this picture tell you? [Follow-up probe] How many friends came to visit each day?
A2 What can't you tell from this picture?
A3 Which day had the highest number of visitors? Which day had the lowest number of visitors?
A4 How many friends came to visit Sam during this week?
O1 About how many friends came to visit Sam each day?
O2 What is the average number of friends that came over to visit Sam each day?
A5 About how many friends would you expect to come to Sam's place every week during the summer vacation?
A6 How many friends would you expect to visit Sam during a 4-week month?

FIGURE 1 Line plot of Sam's friends and questions. D = describing data; A = analyzing and interpreting data; O = organizing and reducing data. From "Young Students' Informal Statistical Knowledge," by I. J. Putt, G. A. Jones, C. A. Thornton, C. W. Langrall, E. S. Mooney, and B. Perry, 1999, *Teaching Statistics, 21,* p. 75. Copyright 1999 by *Teaching Statistics.* Reprinted with permission.

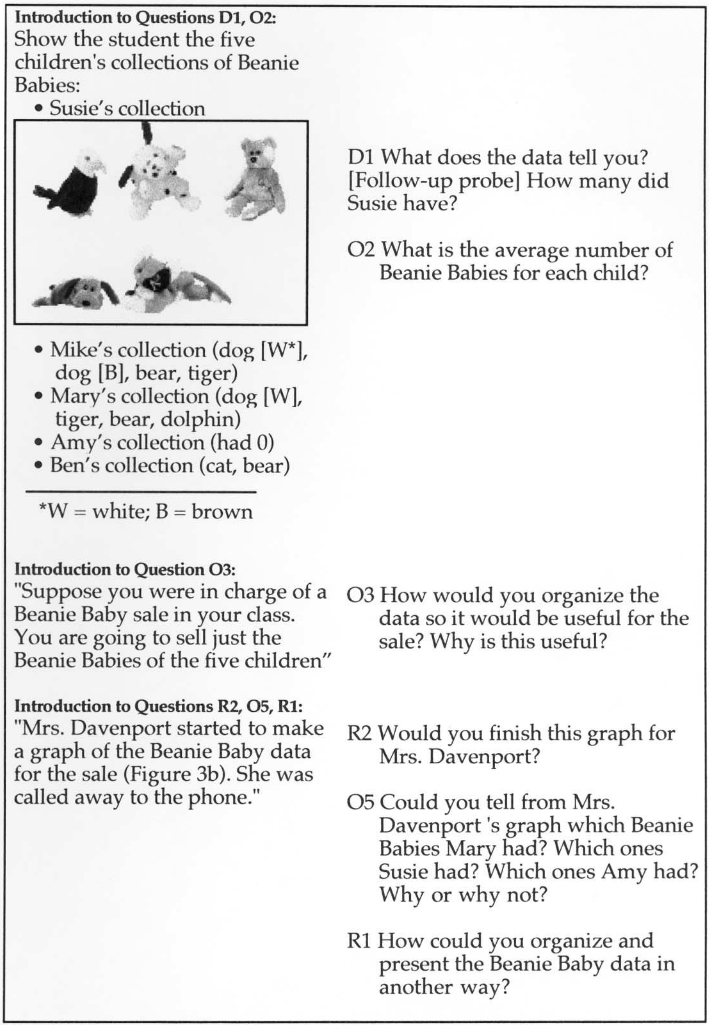


D1 What does this picture tell you?
D2 How are the two pictures alike? Do you think the pictures represent the same data?
D3 Which one of these pictures would be more useful?
R1 Can you draw either of these graphs in another way?

FIGURE 2 Bar graph of Sam's friends and questions. D = describing data; R = representing data. From "Young Students' Informal Statistical Knowledge," by I. J. Putt, G. A. Jones, C. A. Thornton, C. W. Langrall, E. S. Mooney, and B. Perry, 1999, *Teaching Statistics, 21,* p. 75. Copyright 1999 by *Teaching Statistics.* Reprinted with permission.

The double coding procedure described by Miles and Huberman (1994) was used to code the interview protocols. In this procedure, two or more researchers independently code data and then clarify their differences until consensus is reached. For this particular study, Graham A. Jones, Carol A. Thornton, and Cynthia W. Langrall used the initial framework descriptors as criteria (Table 1) to independently code all questions of each student's protocol. Questions were coded according to the construct and level of thinking exhibited by each student. The three researchers then met to compare and negotiate students' thinking levels on each question, to refine the framework descriptors, and to construct a refined framework. Following the adoption of the refined statistical thinking framework, the three researchers again independently coded all questions of each students' protocol and established levels of thinking for each target student on each of the questions. Rater reliability among the three researchers on this second coding procedure was 85%. Final variations were then clarified until consensus was reached on each target student's thinking level for each question. In subsequently determining a target student's dominant level of thinking for a construct, the primary strategy was to identify the student's modal level of thinking for all questions associated with that construct. In cases in which there was more than one mode, the dominant level of thinking for the construct was determined by identifying the median level on the relevant questions. The modal level was preferred because our intent was to capture a students' dominant level of thinking on each construct.

During the coding process described previously, the researchers produced within-case displays to describe and synthesize the statistical thinking of individual target students on the protocol questions associated with the four constructs. Based on these individual within-case displays, we generated a clustered matrix (Construct × Level) to display and amplify the key statistical thinking patterns associated with each level and each construct of the framework (Miles & Huberman,



Introduction to Questions D1, O2:
Show the student the five children's collections of Beanie Babies:

- Susie's collection

D1 What does the data tell you? [Follow-up probe] How many did Susie have?

O2 What is the average number of Beanie Babies for each child?

- Mike's collection (dog [W*], dog [B], bear, tiger)
- Mary's collection (dog [W], tiger, bear, dolphin)
- Amy's collection (had 0)
- Ben's collection (cat, bear)

*W = white; B = brown

Introduction to Question O3:
"Suppose you were in charge of a Beanie Baby sale in your class. You are going to sell just the Beanie Babies of the five children"

O3 How would you organize the data so it would be useful for the sale? Why is this useful?

Introduction to Questions R2, O5, R1:
"Mrs. Davenport started to make a graph of the Beanie Baby data for the sale (Figure 3b). She was called away to the phone."

R2 Would you finish this graph for Mrs. Davenport?

O5 Could you tell from Mrs. Davenport 's graph which Beanie Babies Mary had? Which ones Susie had? Which ones Amy had? Why or why not?

R1 How could you organize and present the Beanie Baby data in another way?

FIGURE 3 The Beanie Baby® questions. D = describing data; O = organizing and reducing data; R = representing data.

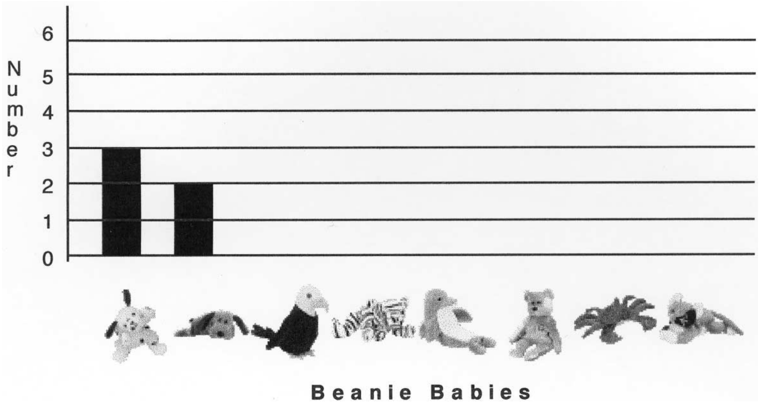
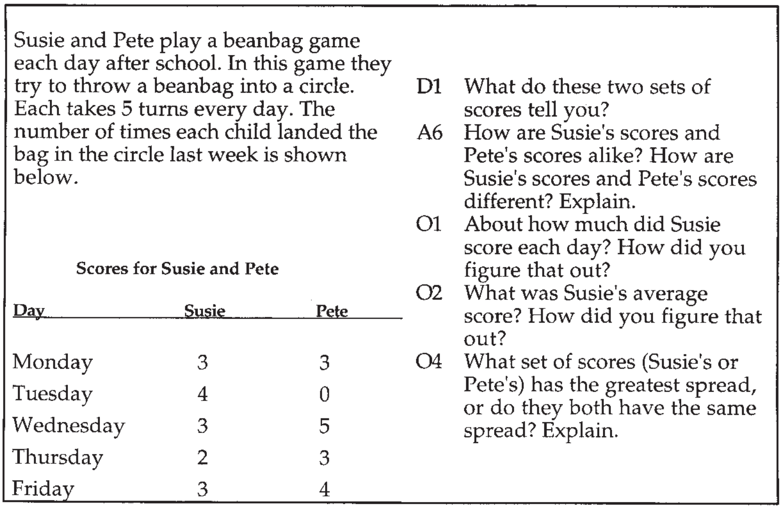


FIGURE 4 Mrs. Davenport’s incomplete graph.



Susie and Pete play a beanbag game each day after school. In this game they try to throw a beanbag into a circle. Each takes 5 turns every day. The number of times each child landed the bag in the circle last week is shown below.

Scores for Susie and Pete

Day	Susie	Pete
Monday	3	3
Tuesday	4	0
Wednesday	3	5
Thursday	2	3
Friday	3	4

D1 What do these two sets of scores tell you?
A6 How are Susie's scores and Pete's scores alike? How are Susie's scores and Pete's scores different? Explain.
O1 About how much did Susie score each day? How did you figure that out?
O2 What was Susie's average score? How did you figure that out?
O4 What set of scores (Susie's or Pete's) has the greatest spread, or do they both have the same spread? Explain.

FIGURE 5 The beanbag game and questions. D = describing data; A = analyzing and interpreting data; O = organizing and reducing data.

1994, p. 127). These patterns were used to further refine the framework descriptors and to generate the summaries that illuminated the framework descriptors.

DATA ANALYSIS AND RESULTS: VALIDATING THE FRAMEWORK

The results of the validation process are presented in three parts. The first part describes refinements made to the framework after the collection and analysis of target student data. In the second part, the profiles and stability of target students' thinking across the four constructs of the refined framework are examined. Finally, summaries, exemplars, and interpretations of target students' thinking are presented to illuminate the four levels of statistical thinking in the refined framework.

Refinements to the Framework

The refined framework is presented in Table 2. In this refined framework, we italicized changes to the descriptors of the initial framework (Table 1) for each of the four constructs: describing data displays, organizing and reducing data, representing data, and analyzing and interpreting data. The refinements for each of these constructs are described next.

In relation to describing data displays, we sharpened the descriptors for the first two rows to better reflect students' literal reading of data displays ("What does the picture tell you?"), their awareness of elemental graph conventions, and their ability to recognize when two different displays represent the same data. One major change was to remove the Level 4 descriptor in row 1. Based on our data, we believe that students' literal descriptions of the data and awareness of graph conventions can be contained within three levels of thinking rather than four. In essence, literal descriptions appear to be complete at the level of quantitative thinking and do not require analytic reasoning.

With respect to organizing and reducing data, refinements were made to all four rows of descriptors (Table 2). In the case of row 1, dealing with grouping, the major change was to the Level 1 descriptor to better capture the students' idiosyncratic or irrelevant groupings. The emphasis given to ordering the data in the initial framework was reduced because it is not a key process for these children, and it appears to occur only as a part of grouping. Although the Level 4 descriptor on grouping was not changed because data at this level were limited to one student, we believe that it may need to be enhanced to reflect students' thinking on more complex grouping tasks. Row 2 presents new descriptors dealing with loss of information when data are reorganized. In formulating the initial framework, we had anticipated that such descriptors would not be necessary because they were part of grouping. Target-student data showed that this was a problematic process, so we developed descriptors that reflect a similar gradation in thinking to that in the first row (grouping). In row 3, dealing with typicality, we attempted to capture more precisely the children's conceptual focus on centers for Levels 2 and 3 and also the degree of completeness of their reasoning. Row 4 involves the notion of spread, and the refined descriptors identify students' attempts to invent measures of spread for Levels 2 and 3.

In refining the descriptors for representing data, we were aware that parts of our data for this construct were more limited because activities involving representing data are so time-consuming with young children. With reference to the row 1

TABLE 2
Refined Statistical Thinking Framework

Construct	*Level 1: Idiosyncratic*	*Level 2: Transitional*	*Level 3: Quantitative*	*Level 4: Analytical*
Describing data displays (D)	In reading the data *literally, gives a description that is unfocused and includes idiosyncratic or irrelevant information;* has no awareness of graphing conventions (e.g., title, axis labels)	In reading the data *literally, gives a description that is hesitant and incomplete* but demonstrates some awareness of graphing conventions	In reading the data *literally, gives a confident and complete description* and demonstrates awareness of graphing conventions	In reading the data *literally, gives a confident and complete description* and demonstrates awareness of graphing conventions
	Does not recognize when two different displays represent the same data OR *indicates some recognition but uses idiosyncratic or irrelevant reasoning*	Recognizes when two different displays represent the same data but uses a justification *based purely on conventions*	Recognizes when two different displays represent the same data by establishing *partial correspondences between data elements in the displays*	Recognizes when two different displays represent the same data by establishing precise numerical correspondences between data elements in the displays
	Considers irrelevant or subjective features when evaluating the effectiveness of two different displays of the same data	Focuses only on one aspect when evaluating the effectiveness of two different displays of the same data	Focuses on more than one aspect when evaluating the effectiveness of two different displays of the same data	Provides a coherent and comprehensive explanation when evaluating the effectiveness of different displays of the same data
Organizing and reducing data (O)	Does not group or *order* the data or *gives an idiosyncratic or irrelevant grouping*	G*ives a grouping or ordering that is not consistent* OR groups data into classes using criteria they cannot explain	Groups or *orders* data into classes and can explain the basis for grouping	Groups or *orders* data into classes in more than one way and can explain the basis for these different groupings
	Does not recognize when information is lost in reduction process	*Recognizes when data reduction occurs but gives a vague or irrelevant explanation*	*Recognizes when data reduction occurs and can explain the reasons for the reduction*	*Recognizes that data reduction can occur in different ways and gives complete explanations for the different reductions*

	Is not able to describe data in terms of representativeness or “typicality”	*Gives hesitant and incomplete descriptions of data in terms of “typicality”*	*Gives valid measures of “typicality” that begin to approximate one of the centers (mode, median, or mean); reasoning is incomplete*	Describes “typicality” of data in terms of common measures of center such as the median or the mean
	Cannot describe data in terms of spread; *gives idiosyncratic or irrelevant responses*	*Invents a measure, usually invalid, in an effort to make sense of spread*	Uses an invented measure or description that is valid, *but the explanation is incomplete*	Uses the range or an invented measure that has the same meaning as the range
Representing data (R)	Constructs an idiosyncratic or invalid display when asked to complete a partially constructed graph associated with a given data set	Constructs a display that is valid in some aspects when asked to complete a partially constructed graph associated with a given data set	Constructs a valid display when asked to complete a partially constructed graph associated with a given data set; may *have difficulty with ideas like scale or zero categories*	Constructs a valid display when asked to complete a partially constructed graph associated with a given data set; *works effectively with scale and zero categories*
	Produces an idiosyncratic or invalid display that does not represent or *reorganize the data set*	*Produces a display that is partially valid* but does not attempt to reorganize the data	Produces a *valid* display that shows some attempt to reorganize the data	Produces multiple valid displays, some of which reorganize the data
Analyzing and interpreting data (A)	Makes no response or an *invalid or* irrelevant response to the question, “What does the display *not* say about the data?”	Makes a relevant but *incomplete* response to the question, “What does the display *not* say about the data?”	Makes multiple relevant responses to the question, “What does the display *not* say about the data?”	Makes a comprehensive contextual response to the question, “What does the display *not* say about the data?”
	Makes no response or gives an invalid or incomplete response when asked to “read between the data”	*Gives a valid response to some aspects of “reading between the data” but is imprecise when asked to make comparisons*	*Gives multiple valid responses when asked to “read between the data” and can make some global comparisons*	*Gives multiple valid responses when asked to “read between the data” and can make coherent and comprehensive comparisons*
	Makes no response or gives an invalid or incomplete response when asked to “read beyond the data”	*Gives a vague or inconsistent response when asked to “read beyond the data”*	*Tries to use the data and make sense of the situation when asked to “read beyond the data”; reasoning is incomplete*	*Gives a response that is valid, complete, and consistent when asked to “read beyond the data”*

Note. Italic type indicates changes to the descriptors of the initial statistical thinking framework (except in the question, “What does the display *not* say about the data?”).

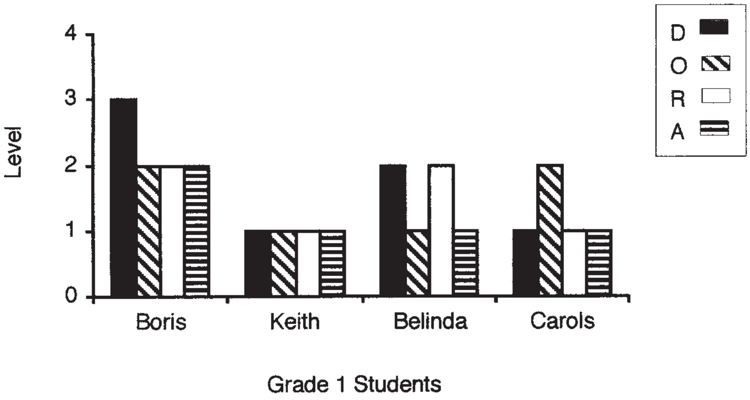
descriptors focusing on the completion of a display, the demands of our task with respect to scale were only sufficient to measure thinking at the first three levels. Hence, our refinement to the Level 4 descriptor is an extrapolation from the Level 3 descriptor rather than a descriptor that is validated by our data. The changes in row 2 are minor, reflecting sharper descriptions of students' thinking at Levels 1, 2, and 3.

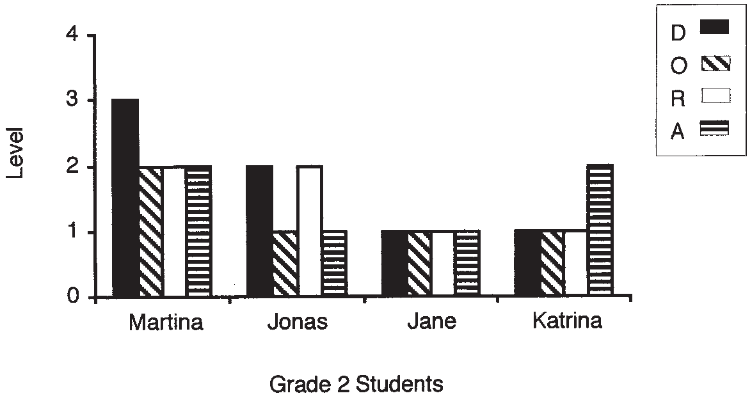
With respect to analyzing and interpreting data, we generated separate rows to describe children's thinking in relation to reading between the data and reading beyond the data (Curcio, 1987, p. 384). Row 2 (which was originally part of row 1) now focuses completely on reading between the data (Curcio, 1987, p. 384), and we have attempted to capture gradation of children's thinking across the four levels. Similarly, row 3 (which was also originally part of row 1) now focuses completely on reading beyond the data. In making changes to Level 4 descriptors, we were cognizant of the fact that only one student consistently responded at that level; hence, the Level 4 descriptors for reading between the data and beyond the data are largely extrapolations from Level 3. Finally, the original row 2, which focuses on what the display does not say about the data, is largely unchanged but has been more conveniently located as row 1.

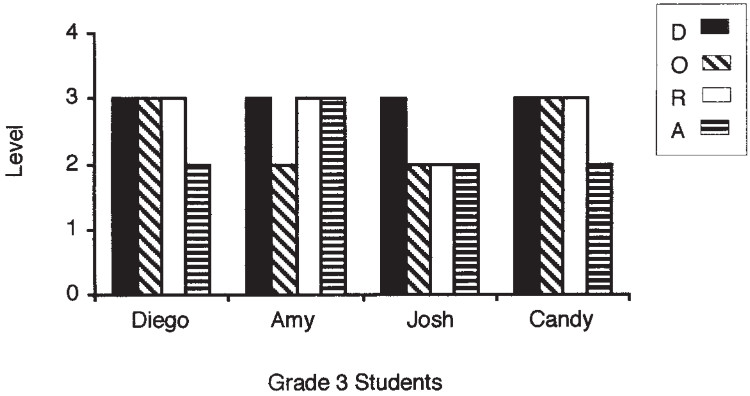
Profiles and Stability of Thinking Across the Four Constructs

An important aspect of the validation process was to assess the stability of each student's thinking across all four constructs. In essence, we wished to measure the extent to which we had been able to translate the SOLO levels (Biggs & Collis, 1991) into levels of statistical descriptors that were stable with respect to students' thinking across constructs. By *stable* we mean, for example, that if a student's statistical thinking level is at Level 2 on describing data displays, it should also be at Level 2 on organizing and reducing data, representing data, and analyzing and interpreting data. From the theoretical considerations associated with the SOLO model, we claim that stability across all four constructs is important because it enhances the coherence of the framework, and greater coherence is seen to make the framework more viable to curriculum builders and teachers when developing instructional programs in data exploration for the elementary school.

The profiles of the 20 target students, 4 each from Grades 1 through 5, are presented in Figure 6. These profiles reflect the target students' thinking levels based on the refined framework (Table 2). Although only 6 students (Keith, Grade 1; Jane, Grade 2; Jai, Grade 4; Sean, Grade 4; Joachim, Grade 5; and Fred, Grade 5) had completely consistent profiles on all four constructs, 16 of 20 (80%) of the target students were consistent (i.e., responded at the same thinking level) on at least three of the four constructs.

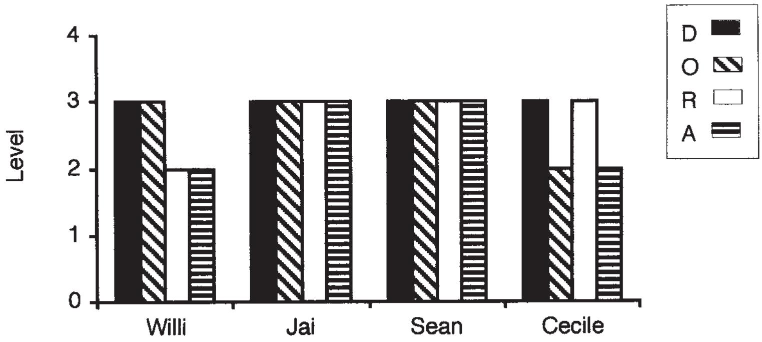




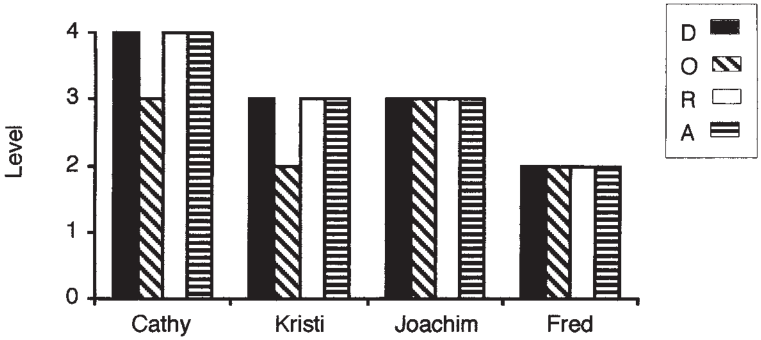


(continued)

FIGURE 6 Target students: Statistical thinking profiles. D = describing data; O = organizing and reducing data; R = representing data; A = analyzing and interpreting data.



Grade 4 Students



Grade 5 Students

FIGURE 6 (Continued) Target students: Statistical thinking profiles. D = describing data; O = organizing and reducing data; R = representing data; A = analyzing and interpreting data.

In Grades 1 and 2, students typically exhibited Level 1 or Level 2 thinking. Only Boris and Martina reached Level 3 thinking on one construct, describing data displays. In Grades 3 and 4, the dominant pattern of thinking exhibited by five of the eight students (Diego, Amy, Candy, Jai, and Sean) was Level 3. One of the other students (Josh) was a typical Level 2 thinker, whereas the other two students exhibited Level 3 thinking on some constructs and Level 2 thinking on others. It is interesting to note that all eight target students in Grades 3 and 4 exhibited Level 3 on describing data, but only three out of the eight were at Level 3 on analyzing and interpreting data. The thinking of the Grade 5 students was more diverse. Cathy demonstrated Level 4 thinking on all constructs except organizing and reducing data. Kristi and Joachim were typical Level 3 thinkers, and Fred exhibited Level 2 thinking that was more indicative of the younger target students.

The profiles also revealed some interesting trends with respect to students' thinking on analyzing and interpreting data and describing data displays. For two students (Diego, Grade 3, and Candy, Grade 3), the level of statistical thinking on analyzing and interpreting data lagged behind the thinking exhibited on all other constructs. For four students (Belinda, Grade 1; Jonas, Grade 2; Willi, Grade 4; and Cecile, Grade 4), the level of statistical thinking on analyzing and interpreting data was lower than that exhibited on two of the other constructs. This trend suggests that analyzing and interpreting data may be more challenging for some students than the other constructs. It may also indicate that elementary children have fewer and less intensive experiences in interpreting and predicting.

On the other hand, the thinking levels of three target students (Boris, Grade 1; Martina, Grade 2; and Josh, Grade 3) on describing data displays were higher than their thinking levels on all other constructs. For four students (Belinda, Grade 1; Jonas, Grade 2; Willi, Grade 4; and Cecile, Grade 4), their levels of statistical thinking on describing data displays were higher than those exhibited on two of the other constructs. Although this trend may be endemic to these target students, it does suggest that some students were more facile with describing data than with the other constructs. Although we tried to accommodate trends like this when refining the descriptors of the framework, the inconsistencies associated with students' thinking on analyzing and interpreting data and on describing data remained and contributed to the lack of stability in target student profiles.

Analysis of Statistical Thinking at Each Level

We summarized and analyzed the responses and statistical reasoning for each of the students who served as target students, and we produced a clustered matrix (Construct × Level) to illuminate and explicate the thinking patterns that are described in the refined framework (Table 2). This analysis provides greater insights into the development of students' thinking for each of the four constructs across the four thinking levels. Although all students at a particular thinking level on a construct tended to generate the same kind of statistical thinking and misconceptions, for ease in reading we have used, as far as possible, exemplars of target students who were at the same level on all four constructs.

Level 1 Thinking

Students exhibiting Level 1 thinking tended to adopt a narrow and often idiosyncratic perspective in relation to data exploration. When describing data displays, their responses were generally unfocused and invariably included irrelevant or incomplete information. For example, when Jane (Grade 2) was shown the bar graph of Sam's friends and asked, "What does this picture tell you?" (Figure 2, D1), she said, "How many blocks there are." Jane made no attempt to relate the display to the context or to give particular details about any of the blocks. In a similar way, when asked whether the line plot and the bar graph represent the same data (Figure 2, D2), Jane said, "No, because this one has writing [a reference to the story line on top of the line plot] and this one doesn't." In essence, she did not recognize that the two displays represented the same data because she focused only on some cosmetic features of the displays.

When organizing and reducing data, Level 1 thinkers did not usually group the categorical data in a systematic manner, and they were not able to characterize data in terms of typicality (average) or spread. For example, when asked how she would organize the Beanie Baby® data (associated with the five children's collections of Beanie Babies®) so it would be useful for a sale (Figure 3, O3), Jane did not reorganize the Beanie Babies® at all. She simply put each child's Beanie Babies® in a bag and began to talk about the Beanie Babies® that the children had in common. Typical of Level 1 thinkers, she did not focus on the question posed but addressed her own agenda. Again, in relation to a series of questions on average, we asked Jane, "About how many friends came to visit Sam each day?" (Figure 1, O1 and O2). She responded, "7," and then said, "I guessed." She was trying to make sense of the question, and, in fact, used the largest number in the data set rather than an average. However, she did not recognize that 7 could not be typical because most days had less than seven friends visiting. In fact, when Jane responded to questions about the average, she sometimes picked the largest number and sometimes listed the complete set of data values. It is interesting to note that, when asked about spread (Figure 5, O4), she again chose the largest number in the data set to indicate the spread. In essence, Level 1 students do not seem to exhibit any prior knowledge about center and spread.

With respect to representing data, students exhibiting Level 1 thinking did not normally complete or construct a valid graph of the data. They often constructed an idiosyncratic graph that had little or no relation to the original data. For example, when Jane was asked (in relation to the bar graph and line plot on Sam's friends; Figure 2, R1), "Can you draw either of these graphs in another way?" she drew what appeared to be a pictograph (Figure 7). Her explanation was typically idiosyncratic: "This is like a bunch of baby snakes on top of each other and [some are] more, [some are] less, and there is one lesser." Although her pictograph bore no relation to the data shown in either the line plot or the bar graph, she did try to find a new symbol (snakes) to generate her graph. It is interesting to note that Jane demonstrated Level 2 thinking in relation to the completion of Mrs. Davenport's graph about the Beanie Babies® (Figure 3, R2, and Figure 4). A more typical Level 1 response was given by Keith (Grade 1). His graph bore no relation to the Beanie Baby® data that was given. He established his own personal Beanie Baby® database by using the ages of each Beanie Baby® rather than the number of each

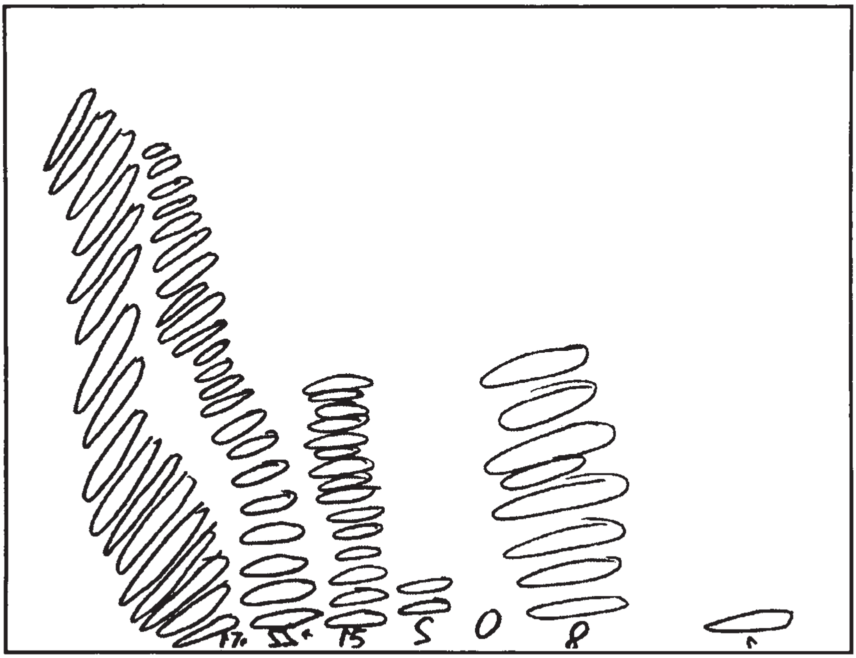


FIGURE 7 Jane's graph. From "Young Students' Informal Statistical Knowledge," by I. J. Putt, G. A. Jones, C. A. Thornton, C. W. Langrall, E. S. Mooney, and B. Perry, 1999, *Teaching Statistics, 21,* p. 76. Copyright 1999 by *Teaching Statistics.* Reprinted with permission.

Beanie Baby® in the combined children's collection. In essence, he introduced his own idiosyncratic extension to the data by talking about their birthdays.

When asked to analyze and interpret data, Level 1 thinkers did not seem to be able to read between the data (Curcio, 1987, p. 384) or read beyond the data. For example, when Jane was asked which day had the lowest number of visitors (Figure 1, A3), she said "Sunday, because it had 1." In responding to this between-the-data (ordering) question, she ignored the fact that Tuesday had 0 visitors, a response that was typical of other Level 1 students. When asked, "About how many friends would you expect to come to Sam's place every week during the summer vacation?" (Figure 1, A5), she said, "Four, because during the summer I had 4 people over." In responding to this beyond-the-data question, she made no attempt to predict. Rather, she based the response on her own personal data set.

Interpretation of Level 1 thinking. Students exhibiting Level 1 thinking were consistently limited to idiosyncratic reasoning that was often unrelated to the given data and frequently focused on their own personal data banks. As hypothesized, the data confirm that the thinking of these Level 1 students reflects what

Biggs and Collis (1991) termed the *prestructural level* of the learning cycle, in that they were engaged in the learning task but were distracted or misled by irrelevant aspects. Note, for example, how consistently Jane gave responses that were unrelated to the context or dealt with peripheral or cosmetic features. This fixation with such features seems to be consistent with Lyons's (1977) treatise on surface features, in which he claimed that surface features can distract the novice from getting to the deep structure of a display in which mathematical relations are embedded. Biggs and Collis also claimed that prestructural thinkers function largely in the ikonic mode of representation and, hence, may indulge in intuitive thinking that involves myths and mental imaging. Such a theoretical position is consistent with the idiosyncratic thinking that Jane displayed, especially in relation to the vivid exemplar in which she constructed a graph with the snakes that bore no relation to the original data.

Level 2 Thinking

Students exhibiting Level 2 thinking attempted to make sense of the data and showed evidence of using quantitative reasoning. However, their responses were typically hesitant and incomplete, and they were often satisfied with just a single response where multiple responses or perspectives would have been appropriate. The connections Level 2 students made between representations of the data and the context of the data seemed to be tenuous at best.

When describing data displays, Level 2 thinkers did not appear to connect the visual display with the original data and had difficulty making use of conventions. For example, when Fred (Grade 5) was asked, "What does this picture [bar graph] tell you?" (Figure 2, D1), he said, "It tells you how many friends came over and well, it tells you the day. And like it's colored more." Although Fred linked the bar graph to the original context about Sam's friends, he did it only in a global way, making no reference to the number of friends that came each day. Even when probed, he was reluctant to give precise information about the number of friends who visited. Again, when asked whether the line plot and the bar graph represented the same data (Figure 2, D2), Fred's response was unfocused: "Yes, a little. Both have the day and how many friends came." Although Fred was trying to make sense of the two displays and may have recognized that they represented the same data, his reasoning was imprecise and largely based on identifying identical labels and information in the titles of the graphs. When evaluating the effectiveness of different displays (Figure 2, D3), Fred focused on only one aspect: "The line plot because the crosses are easy to count."

When organizing and reducing data, Level 2 thinkers tended to produce groupings that overlapped or were incomplete. If they did produce a viable grouping, they generally could not explain it. In relation to the question on organizing the Beanie Baby® data for a sale (Figure 3, O3), Fred started to reorganize the data for the five children by different animals. He put the four bears in a group and the two tigers in a group but then had trouble when he reached the two different kinds of dogs and the two different breeds of cats. He said, "I'm not sure what to do with these." In other words, Fred was able to do a partial reorganization of the data, but his thinking broke down because he did not appear to have a clear criterion or rationale on which to build his reorganization.

When dealing with measures of center and spread, these students (exemplified by Fred) used invented measures that were partially valid. Fred's responses were typical. In relation to the question, "What is the average number of friends that came over to visit Sam each day?" (Figure 1, O2), Fred said, "Between 7 and 0. It's somewhere there, but I don't know." When asked, "What was Susie's average score?" (Figure 5, O2), Fred responded, "It is mostly 3 or 4." In the first question, Fred did not give a value but located the average between the two extremes; in the second question, he tried to focus on the mode (which was actually 3). In both instances, he was striving for an invented measure and seemed to be trying to construct his own meaning for average.

With respect to the question on spread (Figure 5, O4), Fred's response was more characteristic of Level 1 thinking. He listed the two sets of data for Susie and Pete but did not attempt to invent a meaning for spread as he had in the questions on average. A more typical Level 2 response to the question on spread was given by Keith: "Pete, because he has a 0 and a 5" (the two extremes in the data). In this response, Keith appeared to be aware that Pete's data had greater extremes and that this may be a factor in determining the spread. With respect to data reduction, Level 2 students usually recognized that data could be lost in moving between visual displays. However, they did not appear to be able to explain their thinking. For example, in relation to the question, "Could you tell from Mrs. Davenport's graph which Beanie Babies® Mary had?" (Figure 3, O5, and Figure 4), Fred said, "No. They should have put the names on [the graph]." Although his "no" response to the question was correct, he did not explain his response in terms of the bar graph as it was presented. Rather, he thought that the bar graph should be changed to include Mary's name.

With respect to representing data, Level 2 thinkers attempted to complete the Beanie Baby® graph (Figure 3, R2, and Figure 4) but generally produced a display that was not completely valid. Fred actually generated a valid Level 3 response to this question, but Jonas's response was more reflective of Level 2 thinking. He completed the graph, showing only three bears (rather than four) and one cat instead of two. Typically, such students were not always accurate in using the scale that was incorporated in the uncompleted graph. When asked to represent the Beanie Baby® data in another way (Figure 3, R1), Fred's response was characteristic of Level 2 thinkers. He tried to reorganize the data by combining the two different sets of dogs and the two different sets of cats, but then he ran out of ideas. Although his response was incomplete, it did illustrate a partial reorganization of the data.

When asked to analyze and interpret data, Level 2 thinkers normally provided relevant but limited responses to questions involving analysis and interpretation. For example, in relation to the question, “What can’t you tell from this picture?” (Figure 1, A2), Fred said, “The names of the people that came.” He was able to generate this one valid response but was not able to give an additional response even when probed.

Fred had no difficulty with questions that assessed reading between the data. For example, in relation to the question, “How many friends came to visit Sam during this week?” (Figure 1, A4), Fred said, “19. I added them all.” However, when confronted with a reading beyond the data question that required him to predict how many friends would visit another week (Figure 1, A5), Fred said, “18 or 19. [About] the same [as the first week], but not some weeks. They may be on vacation.” Although this seemed at first to exhibit Level 3 thinking, Fred’s response to the follow-up question about the number of friends visiting in a 4-week month (Figure 1, A6) demonstrated the inconsistency of his thinking. He said, “About 20, I guess. That’s all.” In examining Fred’s follow-up response, we were not sure whether he overlooked the fact that the prediction was for 4 weeks (for which 72 or 76 would have been more viable predictions) or whether he did not recognize that there was a connection between the two predictions. Such inconsistencies between these two predictions were common in Level 2 thinkers.

Interpretation of Level 2 thinking. Students exhibiting Level 2 thinking are beginning to recognize the importance of quantitative thinking, and they even use numbers to invent measures for center and spread, albeit not always valid measures. As expected, the data confirms that the statistical thinking of these Level 2 students was consistent with what Biggs and Collis (1991) called the *unistructural level* of the learning cycle. This is evident in the fact that students exhibiting Level 2 thinking were constantly trying to make sense of the data, but their perspective was often single-minded, and they seldom connected their representations or analysis of the data to its context. Note, for example, how consistently Fred focused on only one aspect of the tasks and how often he gave incomplete responses. His answers were generally confined to global considerations (the title or the label) rather than an examination of relations. Moreover, because his thinking was often restricted to one aspect, he had difficulty with tasks like the one that asked him to reorganize the Beanie Babies® for a sale. Fred’s grouping criterion (categories of animals) seemed clear but broke down because he was not flexible enough to accommodate two different kinds of dogs and two different kinds of cats.

Level 3 Thinking

In contrast to students exhibiting Level 2 thinking, Level 3 thinkers used quantitative reasoning more effectively and were more spontaneous and more complete in their responses. They consistently attempted to make sense of the data and generally achieved this, sometimes using their own valid invented measures.

When describing data displays, Level 3 thinkers normally used conventions and labels meaningfully and, as a result, were able to give accurate descriptions of the display. For example, when Sean (Grade 4) was asked, "What does this picture [line plot] tell you?" (Figure 1, D1), he responded, "It tells you 2 visited on Monday, 0 visited on Tuesday, 3 on Wednesday, … and 1 on Sunday." Sean's valid response was given without being probed. However, he did omit direct reference to the "friends" aspect of the context, a link that seems to be more characteristic of Level 4 thinkers. Similarly, when asked whether the bar graph and the line plot represented the same data (Figure 2, D2), Sean replied: "They tell the same story. This one [pointing to the line plot] has Xs and this one [pointing to the bar graph] has bars. Yes, they tell you exactly the same thing—the same numbers each day." Although Sean's response was not completely coherent, he did seem to recognize the key relation between the graphs (i.e., that the numbers are the same for each day). However, he did not establish precisely that the numbers were the same for each day by giving at least one example to demonstrate his point. When evaluating the effectiveness of different displays, Sean was able to give two reasons to support his answer. In relation to Question D3 (Figure 2), Sean said, "The bar graph [is best], because on this one [the line plot] you have to count, and on this one [bar graph] you just read across: '7 friends on Saturday, 2 on Monday.'"

When organizing and reducing data, Level 3 thinkers were usually able to group the categorical data into consistent and nonoverlapping classes and were able to explain the criteria for their classes. For example, when asked to organize the Beanie Baby® data for the sale (Figure 3, O3), Sean readily grouped them by types of animal and said, "Dogs over here, cats over here, bears here. … It [the grouping] would tell people that the dogs are here, the cats are there."

When dealing with measures of center, Level 3 thinkers tended to use quantitative reasoning that enabled them to approximate one of the centers. For example, when asked the average number of friends that came to visit Sam each day (Figure 1, O2), Sean replied, "About 3 or 4. This one has 3, this has 4, this has 7. So if you take 3 away from that [the 7] and give it to the day with 0, you have about 4." Sean did not go on with it, but he appeared to be trying to balance the numbers in a manner that would approximate the mean. We observed that some other students thinking at Level 3 used modal reasoning or tried to split the data values into two equal parts so as to locate a median. Although their explanations were valid, they were incomplete and somewhat inchoate. Level 3 students also tended to recognize that data reduction occurred in some visual displays. For example, when asked if he could tell which Beanie Babies® in Mrs. Davenport's graph belonged to Mary (Figure 3, O5, and Figure 4), Sean emphatically said, "No," and explained, "'Cause it doesn't say Mary had this one; it doesn't have the names on it."

With respect to representing data, Level 3 thinkers could complete a graph and, in contrast to Level 2 students, used the scale provided in a precise manner. Sometimes they had trouble with categories that had zero frequency. For example, although there were no crabs in the Beanie Baby® data, some Level 3 students included a tiny block for the crab when completing Mrs. Davenport's graph (Figure 3, R2, and Figure 4). In essence, they wanted to achieve a degree of completeness, even when it was not appropriate. When asked to construct a graph in another way, Level 3 thinkers tended to produce a display that validly represented the data and also incorporated features that distinguished it from the original graph. For example, when asked to draw the line plot or (vertical) bar graph in another way (Figure 2, R1), Sean drew a valid and complete horizontal bar graph, essentially rotating the original bar graph through 90°. Although the change was minimal, his thinking represented an advance over Level 2 thinkers, who tended to produce identical graphs or graphs that were only partially valid.

When asked to analyze and interpret data, Level 3 thinkers generally provided multiple and meaningful responses. For example, when Sean was asked, "What can't you tell from this picture [the line plot]?" (Figure 1, A2), he responded, "Who visited, what their names were, and whether they came for the whole day." Students who exhibited Level 3 thinking, like those at Level 2, were able to respond correctly to questions that focused on reading between the data. However, students at Level 3 showed a greater sense of being able to make predictions in situations that involved reading beyond the data than their Level 2 counterparts. For example, in response to the questions that asked for predictions about the number of friends visiting in any other week (19 came the first week) and during a 4-week month (Figure 1, A5 and A6), Sean responded, "Between 15 and 10, because I don't think 19 would come every week." Then, in relation to the 4-week month, he said, "Maybe 55—3 weeks of 15 and 1 week of 10. 'Cause 19 [a week] is too high." In making this response, Sean used 19 as a basis and made what appeared to be reasonable refinements. Moreover, when predicting the 4-week month, he used a consistent approach, basing his prediction on his reasoning to the previous question.

Interpretation of Level 3 thinking. Students exhibiting Level 3 thinking consistently use quantitative reasoning as the basis for their statistical judgments and have begun to build valid conceptions of center and spread. The data confirms that the statistical thinking of these Level 3 students reflected what Biggs and Collis (1991) called the *multistructural level* of the learning cycle. This is evidenced in the fact that Level 3 students took a broader and more flexible approach when exploring data and seemed to be able to represent and analyze data from multiple perspectives. Moreover, although they consistently focused on more than one relevant feature of a task, they did not usually connect their thinking on different features. Note, for example, how Sean gave a precise description of the numbers for each day but did not relate this to the context of Sam's friends. It was almost as if the numbers and the days were context free. Except for a cursory reference to the title of the graphs, similar context-free reasoning occurred when Sean explained why the line plot and the bar graph represented the same data. Sean was also able to initiate the complex reasoning needed to balance data values in finding a mean, but his reasoning was incomplete because he could not integrate the mental action of balancing with the need to isolate a single center. Instead he said, "It's about 3 *or* 4," because he was not able to isolate the number about which he wished to forge a balance. Typical of Level 3 thinkers, Sean was able to state more than one piece of information that could not be ascertained from the graph and also showed consistency with respect to his two predictions on how many friends would visit Sam after the original week and in a 4-week month (Figure 1, A5 and A6). In summary, he displayed multistructural reasoning (Biggs & Collis, 1991) in the sense that he was able to make multiple responses to tasks but did not always integrate his thinking.

Level 4 Thinking

Given the fact that only one student, Cathy (Grade 5), consistently exhibited Level 4 responses, some caution needs to be exercised in illuminating the Level 4 descriptors of the framework. Cathy's thinking showed coherence and completeness that was not evident in the thinking of Level 3 students. Her responses were also more analytical, in that she recognized relations in the data and was able to make connections between the data and the context in which the data were situated.

Based on Cathy's thinking in describing data displays, we concluded that Level 4 students recognize the role of conventions and use these conventions to identify relations in much the same way as Level 3 students. By way of contrast, when Cathy was asked whether two different displays represented the same data, she was able to establish more precise contextual correspondences between the displays than her Level 3 counterparts. For example, note how Cathy did this when asked whether the bar graph and the line plot represent the same data (Figure 2, D2): "Well, like on Tuesday, no one came; on Saturday, 7 came on both of these graphs; they are all the same number [of friends] that come over on the other days. And they are all friends that came." Even Cathy's reference at the end to "all friends that came" suggests that she felt the explanation needed a global perspective as well as precise correspondences. Moreover, when Cathy (C) was asked by the interviewer (I) to evaluate the effectiveness of the line plot and the bar graph (Figure 2, D3), she gave a powerful and comprehensive response:

C: Well, it depends on what age you are. But for my age it would probably be this one [pointing to the bar graph].

I: Why do you think that?
C: Well, because it gives us more of a challenge. Because you have to make sure you are going to the right number. Also, because of stocks and stuff, all kinds of things are graphed like that. But when you are younger, you have little Xs, so it is easier to tell how many people.

With respect to organizing and reducing data, Cathy's responses suggest that Level 4 thinkers should be able to group complex data sets into consistent and nonoverlapping classes and be able to develop criteria for their groupings. None of the questions in our protocol was complex enough to assess this kind of thinking; however, Cathy's response to Item R1 (Figure 3), which is presented in the next section on representing data, approximated Level 4. When dealing with measures of center, Level 4 students were expected to demonstrate a conceptual understanding of the mean or the median. Although none of the target students consistently exhibited Level 4 thinking on questions relating to average, Cathy made a Level 4 response when asked about the average number of Beanie Babies® for each child (Figure 3, O2).

C: Well, if you take this [Beanie Baby®] and give it to Amy, and this 1 from here and give it to Amy, and this 1 and give it to Amy, and then 1 to Ben from here, then the average would probably be 3.
I: Why?
C: Because, well, they would all have 3. That shares them out.

In essence, Cathy balanced the data by sharing the Beanie Babies® among all of the five students. Although Cathy's approach could not always be used to generate the mean, her thinking did exhibit conceptual understanding of the mean.

Similarly, with regard to spread, none of the target students gave a response that exhibited Level 4 thinking. However, Cathy approximated a Level 4 response, one that showed some conceptual understanding of spread. In answering the question about the spread in Susie and Pete's scores (Figure 5, O4), she said: "Susie's numbers are closer together. They're 1 apart. It usually increases by 1 or decreases by 1. But Pete is spread differently. His goes 3, 0, 5, 3, 4—up and down. They [Susie and Pete's scores] are about the same [spread], but Pete goes further." Although Cathy did not use an invented measure that had the same meaning as the range, she did appear to recognize that Pete had bigger and smaller numbers and that Susie's scores were more densely distributed.

With respect to representing data, Cathy gave a response that showed Level 4 thinking when she was asked to reorganize data and present it in another way. For example, in responding to R1 (Figure 3), she reorganized the Beanie Baby® data by habitats: jungle (tiger), farm (cats and dogs), water (dolphin), air (eagle), and then drew a valid bar graph. It is worth noting that she was the only student who was able to completely reorganize the data and graph it in a meaningful way. The fact that she could manage this reorganization suggests that she may have been capable of reorganizing even more complex data had she been given the opportunity.

When asked to analyze and interpret data, Cathy's responses suggest that Level 4 thinkers are able to provide comprehensive responses in the sense that they establish relations and situate these relations within the original context. There was also evidence that she was able to generate alternative interpretations of the data. For example, when Cathy was asked, "What can't you tell from this picture [line plot]?" (Figure 1, A2), she replied, "How long they stayed, what hour they came; if they stayed overnight and were counted twice." Not only did Cathy identify several aspects that the graph did not reveal, she raised an alternative interpretation about the data; that is, that some of the people may have been counted twice. Her analytic abilities were further evidenced when she was asked to predict the number of Sam's friends who would visit in any other week (19 came the first week) and during a 4-week month (Figure 1, A5 and A6). She responded, "Well, since there was 19 this week, that's a good idea [estimate]. Probably 15, because Saturday had so many people. It might not have that many next time because a lot of people go on vacation." Then, in relation to the question on the 4-week month, she said, "Four by 15—that's 60. Kind of a lot, but 15 per week." Cathy demonstrated valid and consistent reasoning when dealing with questions like this that focused on reading beyond the data. She used the original data to make predictions and then justified her predictions with plausible arguments.

Interpretation of Level 4 thinking. None of the target students exhibited Level 4 thinking across all constructs; however, Cathy gave us frequent glimpses of Level 4 thinking. Based on the limited sampling provided by Cathy, we claim that students who exhibit Level 4 thinking use both analytical and numerical reasoning in explorations of data. She also showed evidence of being able to make connections between different aspects of data. In essence, Cathy's thinking mirrors what Biggs and Collis (1991) called the *relational level,* in the sense that she constructed meaningful and coherent responses by integrating several relevant aspects of the tasks. She not only recognized multiple aspects of a data task, as Level 3 thinkers do, but she also went beyond Level 3 thinkers by making connections among those different aspects of the tasks. For example, in contrast to Sean (Level 3), note how Cathy related her precise correspondences between the line plot and the bar graph (Figure 2, D2) to the context of Sam's friends to show that both graphs represented the same data. There was a sense of coherence, not evident in Level 3 responses, when Cathy integrated the corresponding relations in the graphs and the context within which they were situated. This ability to integrate several facets of her thinking was even more evident when she compared the effectiveness of the line plot and the bar graphs from the perspectives of a younger and older child (Figure 2, D3).

Again, when she predicted the number of friends who would visit Sam in subsequent weeks and in a 4-week month (Figure 1, A5 and A6), her thinking demonstrated the integrated and relational reasoning characterized by Biggs and Collis.

DISCUSSION

In responding to the need for research on students' thinking that will inform instruction (Cobb et al., 1991; Fennema et al., 1996; Resnick, 1983), in this study, we formulated and validated a framework for characterizing elementary children's statistical thinking. The formulation and validation of the framework was an evolutionary and iterative process in which we began by using previous research on students' statistical thinking and the SOLO general cognitive development model (Biggs & Collis, 1991) to formulate the initial framework. This initial framework described children's statistical thinking according to the four cognitive levels and across four key constructs: describing data, organizing and reducing data, representing data, and analyzing and interpreting. The validation process was initiated by constructing a protocol incorporating 29 tasks that assessed a wide range of thinking across four constructs. These tasks were used to assess the statistical thinking of 20 target students, 4 purposefully chosen from each of Grades 1 through 5. Subsequently, we undertook a case-study analysis of each child's protocol to refine the initial framework, examine the stability and growth of the students' profiles, and illuminate the characteristics of each of the four thinking levels.

We made a number of changes to the framework descriptors during the refinement phase of the validation process (see Table 2). Some of the refinements to framework descriptors (e.g., those associated with the completion of an unfinished graph under the construct representing data) resulted from the fact that the demands of the protocol task, with respect to scale, were only sufficient to measure thinking at the first three levels. Other changes resulted from the fact that a cognitive ceiling was reached prior to Level 4 (e.g., in the first descriptor on describing data displays). In addition, we believe that the changed descriptors now appearing under the construct of analyzing and interpreting data better represent the students' thinking in relation to tasks characterized as reading between the data and reading beyond the data (Curcio, 1987). Notwithstanding these modifications, it is evident that further refinements will need to be made to the framework descriptors, especially with respect to constructs, such as organizing and representing data, that are time-consuming and difficult to assess with young children. More specifically, further research is needed to sharpen the framework's descriptions in relation to grouping and representing data (especially more complex numerical data), average, and spread.

In formulating and refining the framework, we attempted to achieve stability (consistency) for a student's thinking across the four constructs by translating the four SOLO cognitive levels (Biggs & Collis, 1991) into four statistical thinking levels for each construct. That is, when a student's statistical thinking was at Level 2 on one construct (e.g., describing data), we formulated the framework on the basis that the student's thinking would also be a Level 2 on the other three constructs. Stability was important because we wanted the framework to have overall coherence so that it would be more user-friendly for teachers when they design instructional programs and monitor a student's statistical thinking. Although 80% of the target students exhibited consistent thinking levels on at least three of the four constructs and 30% had completely consistent thinking patterns across the four constructs, there was still "static" (irregularities) in the system, a phenomenon that has been observed in previous studies (Jones et al., 1997; Jones et al., 1996; Tarr & Jones, 1997).

Given the fact that the research base on young children's statistical thinking is emergent rather than well developed, this static may well have resulted from our inability to translate precisely from SOLO cognitive levels to corresponding statistical thinking levels for each construct. There may also be limitations within the SOLO model itself. For example, in a recent study, Watson and Moritz (1998) suggested that there may be two cycles of structural levels in the concrete-symbolic SOLO modes of representation rather than one as was assumed in this study. Apart from these considerations associated with the SOLO model, our data suggest that the static may be related to the fact that thinking levels on analyzing and interpreting data seemed to lag behind other constructs, especially for some third- and fourth-grade students. There was also a tendency for thinking levels on describing data to be higher than on other constructs. This was especially apparent for the three students (Boris, Martina, and Josh) whose thinking on describing data was higher than on any other constructs. Although both of these findings are consistent with previous research findings (Beaton et al., 1996; Pereira-Mendoza & Mellor, 1991; Zawojewski & Heckman, 1997) that students have more success on questions involving describing data than on questions involving analyzing and interpreting data, the results raise a pertinent issue with respect to refining the framework and improving its stability. Responding to this issue, F. R. Curcio (personal communication, April 21, 1999) questioned whether the construct describing data displays should be subsumed under the construct of analyzing and interpreting data. In suggesting this aggregation of the two constructs, she claimed that a student's response to a question like "What does the picture tell you?" (e.g., Figure 1, D1) may well determine whether the student is describing the data display or analyzing and interpreting the data display. In essence, she was saying that the nature of a student's response rather than the question itself may well determine which of the two constructs is appropriate. We believe that we were able to distinguish between students' thinking when describing data literally and when analyzing or interpreting data. However, we agree that this issue needs further investigation especially because the constructs describing data displays and analyzing and interpreting data may have been instrumental in producing static in the framework.

The validation analysis also added to our theoretical understanding of statistical cognition by revealing more in-depth insights into students' statistical thinking at each level and across all four constructs. In particular, the analysis confirmed that the statistical thinking of young children can be situated in the more general neo-Piagetian development model developed by Biggs and Collis (1991). Our evidence showed that it was possible to build and refine four levels of statistical thinking, idiosyncratic (Level 1), transitional (Level 2), quantitative (Level 3), and analytical (Level 4), that are domain-specific manifestations of the four structural levels of the Biggs and Collis concrete-symbolic learning cycle. This conclusion is consistent with the findings of previous studies on elementary school students' thinking in other mathematical areas: number sense (Jones et al., 1996) and probability (Jones et al., 1997; Watson & Moritz, 1998). As a consequence, this research not only extends the mathematics education literature but also provides confirmatory evidence of the viability of using more general developmental models, such as SOLO, to generate domain-specific frameworks of children's mathematical thinking. Notwithstanding the overall integrity of these conclusions, we recognize that some caution must be exercised in relation to the Level 4 descriptors and analysis. Because only one student in this study consistently exhibited this level of thinking, further research is needed to confirm and sharpen the characteristics of analytical thinking (Level 4) in the framework.

The refined framework developed in this study has implications for curriculum design and instruction in data exploration. Because it provides a coherent picture of children's thinking across four key constructs, we believe that curriculum designers and teachers can use the framework to build goals and tasks that are within the scope of students' statistical thinking yet incorporate the broad range of statistical ideas inherent in the four constructs. More specifically, in providing cognitive knowledge to inform instruction, we maintain that the framework offers a very accessible means of building instructional sequences or hypothetical learning trajectories (Simon, 1995). That is, the framework provides a valuable tool for the teacher in planning learning goals, designing learning tasks, and predicting the kind of learning and thinking that will occur as those tasks are played out. As an example, tasks like those of this study could be used as focus problems in instruction and also as problems to assess students' knowledge at various stages of the instructional process. Our claim that the framework can play a critical role in informing instruction is based on our use of similar frameworks to develop and implement instructional sequences in number sense (Jones et al., 1996) and probability (Johnson, Jones, Thornton, Langrall, & Rous, 1998; Jones, Langrall, Thornton, & Mogill, 1999). In using the framework to inform instruction and monitor students' thinking, we are not suggesting that growth in each student's statistical thinking across the four constructs will necessarily be uniform, nor are we suggesting that all students will follow an ordered learning progression through the levels of the framework. What we are suggesting is that teachers can use the level descriptors as broad guidelines for designing instruction and for assessing children's prior and emergent knowledge.

Although the students in this study represented a range of cultural and socioeconomic backgrounds, the size of the target student sample may limit the extent to which conclusions about the framework can be applied to more culturally diverse populations of elementary school students. In accord with Shaughnessy's (1992) call for cross-cultural research in probabilistic and statistical thinking, we recommend further research to investigate whether the framework is appropriate for students from other cultural and linguistic backgrounds. In addition, teaching experiments (Cobb, 1999) are needed to evaluate the viability of using the framework for informing instruction in regular classroom situations, that is, to assess the ease and effectiveness with which classroom teachers can use the statistical thinking framework to enhance student learning in data handling. Such research would also provide further opportunities for fine-tuning the framework and for tracing students' statistical thinking across levels. In essence, further research is needed to continue the iterative process of framework building so that it will be more effective for supporting instructional programs that build on students' prior knowledge, foster their thinking, and monitor their understanding.

REFERENCES

Australian Education Council. (1994). *Mathematics: A curriculum profile for Australian schools.* Carlton, Australia: Curriculum Corporation.

Beaton, A. E., Mullis, I. V. S., Martine, M. O., Gonzalez, E. J., Kelly, D. L., & Smith, T. A. (1996). *Mathematics achievement in the middle school years: IEA's Third International Mathematics and Science Study (TIMSS).* Chestnut Hill, MA: Boston College, Center for the Study of Testing, Evaluation, and Educational Policy.

Biggs, J. B., & Collis, K. F. (1982). *Evaluating the quality of learning: The SOLO taxonomy (Structure of the Observed Learning Outcome).* New York: Academic.

Biggs, J. B., & Collis, K. F. (1991). Multimodal learning and intelligent behavior. In H. Rowe (Ed.), *Intelligence: Reconceptualization and measurement* (pp. 57–76). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Bright, G. W., & Friel, S. N. (1998). Graphical representations: Helping students interpret data. In S. P. Lajoie (Ed.), *Reflections on statistics: Learning, teaching, and assessment in Grades K–12* (pp. 63–88). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Carpenter, T. P., & Moser, J. M. (1984). The acquisition of addition and subtraction concepts in grades one through three. *Journal for Research in Mathematics Education, 15,* 179–202.

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

Cobb, P., Wood, T., Yackel, E., Nicholls, J., Wheatley, G., Trigatti, B., & Perlwitz, M. (1991). Assessment of a problem-centered second-grade mathematics project. *Journal for Research in Mathematics Education, 22,* 3–29.

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

Curcio, F. R. (1989). *Developing graph comprehension.* Reston, VA: National Council of Teachers of Mathematics.

Curcio, F. R., & Artz, A. F. (1997). Assessing students' statistical problem-solving behaviors in a small-group setting. In I. Gal & J. B. Garfield (Eds.), *The assessment challenge in statistics education* (pp. 123–138). Amsterdam: IOS Press.

Curcio, F. R., & Folkson, S. (1996). Exploring data: Kindergarten children do it their way. *Teaching Children Mathematics, 6,* 382–385.

De Lange, J., van Reeuwijk, M., Burrill, G., & Romberg, T. (1993). *Learning and testing mathematics in context. The case: Data visualization.* Madison, WI: University of Wisconsin, National Center for Research in Mathematical Sciences Education.

Fennema, E., Carpenter, T. P., Franke, M. L., Levi, L., Jacobs, V. R., & Empson, S. B. (1996). A longitudinal study of research to use children's thinking in mathematics instruction. *Journal for Research in Mathematics Education, 27,* 403–434.

Gal, I., & Garfield, J. B. (1997). *The assessment challenge in statistics education.* Amsterdam: IOS Press.

Johnson, T. M., Jones, G. A., Thornton, C. A., Langrall, C. W., & Rous, A. (1998). Students' thinking and writing in the context of probability. *Written Communication, 15,* 203–229.

Jones, G. A., Langrall, C. W., Thornton, C. A., & Mogill, A. T. (1997). A framework for assessing and nurturing young children's thinking in probability. *Educational Studies in Mathematics, 32,* 101–125.

Jones, G. A., Langrall, C. W., Thornton, C. A., & Mogill, A. T. (1999). Students' probabilistic thinking in instruction. *Journal for Research in Mathematics Education, 30,* 487–519.

Jones, G. A., Thornton, C. A., & Putt, I. J. (1994). A model for nurturing and assessing multidigit number sense among first grade children. *Educational Studies in Mathematics, 27,* 117–143.

Jones, G. A., Thornton, C. A., Putt, I. J., Hill, K. M., Mogill, A. T., Rich, B. S., & Van Zoest, L. R. (1996). Multidigit number sense: A framework for instruction and assessment. *Journal for Research in Mathematics Education, 27,* 310–336.

Konold, C., Pollatsek, A., Well, A., & Gagnon, A. (1997). Students analyzing data: Research of critical barriers. In J. B. Garfield & G. Burrill (Eds.), *Research on the role of technology in teaching and learning statistics: Proceedings of the 1996 ISAE round table conference* (pp. 151–167). Voorburg, The Netherlands: International Statistical Institute.

Lamon, S. (1993). Ratio and proportion: Children's cognitive and metacognitive processes. In T. P. Carpenter, E. Fennema, & T. A. Romberg (Eds.), *Rational numbers: An integration of research* (pp. 131–156). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

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

Lyons, J. (1977). *Noam Chomsky* (Rev. ed.). New York: Penguin.

Mack, N. K. (1990). Learning fractions with understanding: Building on informal knowledge. *Journal for Research in Mathematics Education, 21,* 16–32.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis.* Thousand Oaks, CA: Sage.

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

Moore, D. S. (1997). *Statistics: Concepts and controversies* (4th ed.). New York: Freeman.

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

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

Pereira-Mendoza, L., & Mellor, J. (1991). Students' concepts of bar graphs: Some preliminary findings. In D. Vere-Jones (Ed.), *Proceedings of the Third International Conference on Teaching Statistics* (Vol. 1, pp. 150–157). Voorburg, The Netherlands: International Statistical Institute.

Putt, I. J., Jones, G. A., Thornton, C. A., Langrall, C. W., Mooney, E. S., & Perry, B. (1999). Young students' informal statistical knowledge. *Teaching Statistics, 21,* 74–78.

Resnick, L. B. (1983). Toward a cognitive theory of instruction. In S. G. Paris, G. M. Olson, & W. H. Stevenson (Eds.), *Learning and motivation in the classroom* (pp. 5–38). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

School Curriculum and Assessment Authority & Curriculum and Assessment Authority for Wales. (1996). *A guide to the national curriculum.* London: Author.

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

Shaughnessy, J. M., Garfield, J., & Greer, B. (1996). Data handling. In A. J. Bishop, K. Clements, C. Keitel, J. Kilpatrick, & C. Laborde (Eds.), *International handbook of mathematics education* (Pt. 1, pp. 205–238). Dordrecht, The Netherlands: Kluwer.

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

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

Tarr, J. E., & Jones, G. A. (1997). A framework for assessing middle school students' thinking in conditional probability and independence. *Mathematics Education Research Journal, 9,* 39–59.

van Hiele, P. M. (1985). The child's thought and geometry. In D. Fuys, D. Geddes, & R. Tischler (Eds.), *English translation of selected writing of Dina van Hiele-Geldof and Pierre M. van Hiele* (pp. 243–252). New York: Brooklyn College, School of Education. (Original work published 1959; ERIC Document Reproduction Service No. ED 287 697)

Watson, J., Collis, K. F., & Moritz, J. (1997). The development of chance measurement. *Mathematics Education Research Journal, 9,* 60–81.

Watson, J., & Moritz, J. (1998). Longitudinal development of chance measurement. *Mathematics Education Research Journal, 10,* 103–127.

Zawojewski, J. S., & Heckman, D. S. (1997). What do students know about data analysis, statistics, and probability? In P. A. Kenney & E. A. Silver (Eds.), *Results from the sixth mathematics assessment of the National Assessment of Educational Progress* (pp. 195–223). Reston, VA: National Council of Teachers of Mathematics.