Mathematics Education Research Journal 2003, Vol. 15, No. 1, 42-58

Children's Representation and Organisation of Data

Steven Nisbet
Griffith University

Graham Jones, Carol Thornton, Cynthia Langrall and Edward Mooney
Illinois State University

This study investigated how children organised and represented data and also examined relationships between their organisation and representation of data. Two protocols, one involving categorical data and the other involving numerical data, were used to interview 15 students, 3 from each of Grades 1 through 5. Although there were differences between Grade 1 students and the rest, the study suggested that numerical data was significantly harder for children to organise and represent than categorical data. Children beyond Grade 1 could make connections between organising and representing data for categorical data but their connections for numerical data were more tenuous. The process of reorganising numerical data into frequencies was not intuitive for the children in this study but they showed greater readiness in recognising and interpreting data that had already been reorganised as a frequency representation. Given this latter result, a pedagogical approach that asks students to make links between raw data and a frequency representation of it may prepare students to create and construct their own frequency representations.

Recent reforms in mathematics education (Australian Education Council, 1994; National Council of Teachers of Mathematics, 2000) have called for more extensive and earlier introduction to data handling in the elementary school. This broader perspective has highlighted the need for further research on the learning and teaching of data handling at the elementary school level—an area that is underrepresented in the mathematics education research literature (Shaughnessy, Garfield, & Greer, 1996). Although there is evidence that research on students' statistical thinking and learning is beginning to emerge (Bright & Friel, 1998; Curcio, 1987; Friel, Curcio & Bright, 2001; Greer, 2000; Lehrer & Romberg, 1996; Mokros & Russell, 1995; Moritz, 2002; Watson & Moritz, 2000), this research has largely focused on students' statistical thinking beyond early elementary. In fact, even though graphing has been part of the elementary school curriculum for the last two decades, there is almost no research on how young children organise and represent raw data.

The present study, albeit exploratory, attempted to address this gap by examining how children organise data when they are in the process of representing data. More specifically, this study asked the following questions: (a) How is children's statistical thinking in organising and representing data related to grade level? (b) What is the effect of the type of data (categorical versus numerical) on children's statistical thinking in organising and representing data? (c) Is children's statistical thinking in representing data related to their thinking in organising data? Is this relationship influenced by whether the data is categorical or numerical? (d) Is students' performance in recognising organisations of data related to their performance in creating organisations of data?

Theoretical Considerations

This study is based on a Statistical Thinking Framework (Framework) (Jones et al., 2000) that characterises children's statistical thinking according to four cognitive levels and across four key constructs: describing data, organising and reducing data, representing data, and analysing and interpreting data. The present study is concerned with two of the constructs—(a) representing data, and (b) organising and reducing data. *Representing data* involves the construction of visual representations of data, for example, bar graphs, tally graphs, tables and line plots. Of particular importance in this study were representations that incorporated different organisations or groupings of data. *Organising and reducing data* incorporates mental actions such as grouping, ordering and summarising data (Moore, 1997). We were especially interested in children's organisations of data that were generated whilst they were representing the given data. That is, we were interested in children's preferred organisations of data that arose as part of their actions in representing the data. In addition we also examined children's reactions to a representation and organisation of data generated by an imaginary child.

The Jones et al. (2000) Framework provides expectations of the levels and kinds of thinking children engage in when representing and organising sets of data. For *representing data*, children exhibiting Level 1 thinking (Idiosyncratic) produce idiosyncratic or incomplete displays of the data set. (The authors of the Framework regard both idiosyncratic and incomplete responses as baseline indicators of children's thinking in relation to data representation.) Children at Level 2 (Transitional) produce displays that represent the data but do not attempt to reorganise it. Children at Level 3 (Quantitative) produce displays that show some attempt to reorganise or regroup the data. Children at Level 4 (Analytical) produce multiple valid displays, some of which reorganise the data. With respect to *organising and reducing* data, the Framework reveals that children exhibiting Level 1 (Idiosyncratic), that is baseline thinking, give idiosyncratic groupings or do not group or order the data at all. Children at Level 2 (Transitional) give groupings or ordering that are not consistent, or group data into classes using criteria that they cannot explain. Children at Level 3 (Quantitative) order or group data into classes, and can explain the basis for their groupings. Children at Level 4 (Analytical) order or group the data into classes in more than one way, explain the basis for their different groupings or orderings of the data.

In validating this framework, Jones et al. (2000) noted limitations in the way that they had assessed young children's organisation of data. In particular, they were concerned that their protocol question dealing with the organisation of the Beanie Baby data had focused the children's thinking in a particular way (by animal categories) rather than allowing the children to create they own organisations of the data. They advocated further research into the cognitive processes used by children when they were confronted with tasks that provided opportunities for them to represent and organise data in multiple ways. This study is a response to their recommendation.

Two other studies (Lehrer & Schauble, 2000; Nisbet, 1998) also provided a stimulus for this research. Using a task involving categorical data similar to the transport protocol presented in this study, Nisbet examined the representations generated by 114 teacher education students. These students created 11 different types of representations ranging from lists (ungrouped and grouped) to various types of tables, pictographs, line plots and bar graphs. Their representations also revealed the complete range of thinking levels identified by the Jones et al. (2000) Framework. Lehrer and Schauble worked with elementary school children in Grades 1, 2, 4 and 5. They examined how these children developed and justified models to categorise and differentiate (by grade level) drawings made by children in the same grade levels as themselves. The Grades 1 and 2 children were reluctant to use attributes to classify the drawings by grade level and tended to justify their categorisations based on idiosyncratic perceptions of a drawer's grade levels or age. By way of contrast, the Grades 4 and 5 children developed category systems that used dimensional attributes associated with the drawing characteristics shown for arms, fingers, necks, feet, and other body parts. The results suggest that as grade level increases children demonstrate increasing sophistication in their strategies for organising data.

Method

Participants

Students in Grades 1 through 5 from a Midwestern U.S. elementary school were the population for the study. The school population was representative of a broad spectrum of socioeconomic and cultural backgrounds. The sample for the study comprised 15 children, three being selected from each of the five grades. At each grade level children were purposefully sampled (Miles & Huberman, 1994), based on their previous mathematics achievement—one high (upper quartile), one middle (middle quartiles), and one low (lower quartile)—in order to increase the representativeness of the sample However, the study did not attempt to make comparisons among children categorised as high, middle or low on mathematics achievement.

Children in these grade levels had undertaken instruction in data exploration. This instruction focussed largely on learning to construct particular kinds of graphs such as bar graphs, circle graphs, and line graphs, using data that was provided by the teacher or the textbook and had already been organised. However, the children had little experience in dealing with raw data, and even less experience in creating their own representations and organisations of data.

Interview Protocols

The first author interviewed all children in the sample using two researcher-designed Statistical Representation Protocols. These were presented in two separate sessions of approximately 20 minutes each, which were separated by a week approximately. Both protocol sessions were audio-taped, and the children were provided with paper and felt pens to draw their pictures or graphs.

For each child, Protocol 1 was administered in the first session, and Protocol 2 in the second session. We maintained this order so that any learning effects would be consistent across the sample. Protocol 1 (Figure 1), involved data on how a class of 10 students in a rural school travelled to school. The data was *categorical* with each student listed by name and mode of transport. Each child in the study read the story and associated data once, followed by a second reading with the researcher. In the case where a child was not able to read the information, the researcher read it the first time, and the child and researcher read it a second time together. Following the reading of the story and data, the researcher probed the child with simple questions like, "How does Brendan come to school?" to ensure that the child was familiar with the story and the data. The researcher then focused on the major task that involved the child in drawing a picture or graph that showed how the students in the story got to school. After giving the child time to construct a picture or graph, the researcher asked the follow-up questions presented in Figure 1.

Story:

Some children were talking about how they came to school. This is what they said.

Alice comes by bus.
Brendan comes by car.
Cathy rides a bike.
Denis comes by bus.
Elouise walks to school.
Francis comes by bus.
Gail comes by car.
Herby walks.
Ilsa comes by car.
Jack comes by bus.

Task:

I'd like you to draw a picture or graph that shows how the children in this class get to school.

Follow-up questions:

Can you tell me what you have drawn here?
Why did you draw it that way?
If someone came into the room and saw your graph, what would they learn from it?
Is there anything else they would learn?
What title could you write at the top?

Figure 1. How some children came to school.

Protocol 2 (Figure 2) involved the number of pet fish belonging to a group of 10 students. The fish data were *numerical* with each student listed by name and number of fish. The procedures for reading the story and associated data were exactly the same as in Protocol 1. As before, following the reading of the story and data, the researcher probed each child with simple questions such as, "How many fish does Bruce have?" Then the researcher focused on the major task that involved the child in drawing a picture or graph that showed the number of pet fish these students had. After giving the child time to complete her picture or graph, the researcher asked the follow-up questions presented in Figure 2. The first set of questions related to the child's graph. After those questions had been asked, the researcher showed a graph of the fish data (Figure 3) drawn by an imaginary child Mary and asked the second set of follow-up questions. By

Story: Ten children from the Pet Fish Club met at school one day. They were talking about how many fish they had.
Amy had 4 fish.
Bruce had 2 fish.
Cary had 5 fish.
Don had 4 fish.
Ember had 7 fish.
Francio had 3 fish.
Gary had 0 fish. (They all died.)
Hugo had 2 fish.
Izra had 9 fish.
Janita had 4 fish.
Task: I'd like you to draw a picture or graph that shows something about the numbers of pet fish these children have.
Follow-up questions (Set 1):
Can you tell me what you have drawn here?
Why did you draw it that way?
If someone came into the room and saw your graph, what would they learn from it?
Is there anything else they would learn?
What title could you write at the top?
Follow-up questions (Set 2):
Another person called Mary has drawn this graph.
Can you read what it says?
What do you think Mary was doing when she drew this graph?
What do these crosses mean?

Figure 2. Pet Fish Club.

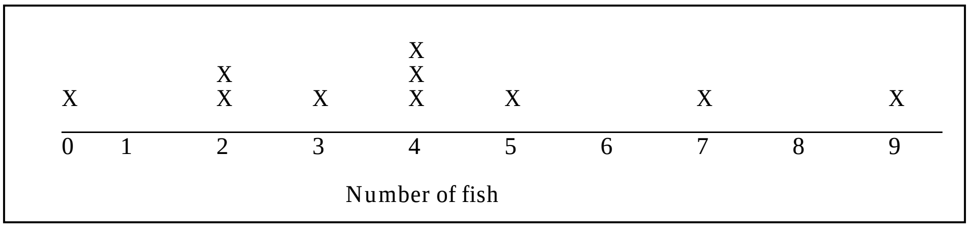

Figure 3. Mary's graph.

including Mary's graph (with its inherent organisation of the data) we provided an opportunity for children to compare their organisation and representation of the data with an alternative that may have been different from theirs. In the case of children who had not able to reorganise the raw data, it enabled us to determine if they could recognise and interpret a reorganisation of the data. In the case of children who were able to reorganise the data, it enabled us to determine whether they recognised commonalities and differences between their representation and that of Mary.

Data Sources and Analysis

The sources of data for this study comprised transcripts of the audio-taped interview sessions, and students' artifacts in the form of pictures or graphs.

Adopting a double-coding procedure (Miles & Huberman, 1994), the first two authors used these sources to prepare summaries of each student's responses to Protocols 1 and 2. More specifically, they coded the children's representations according to the following characteristics: level of thinking on the Framework (Levels 1 to 4), type of representation (bar graph, table, line plot, tally plot), manner of organising the data (categories, groupings), and use of conventions (labeling axes, and scale). Following a similar procedure to Jones et al. (2000), the two researchers independently assessed thinking levels for each of the 15 students on the two constructs and the two protocols, that is, 60 assessments of levels by each researcher. The level of agreement between researchers for the independent phase was 90% (54 out of 60 assessments). Following the independent assessments, the researchers then clarified points of difference until consensus was reached.

During the coding and levels analysis the two researchers produced "within-case displays" to synthesise the statistical thinking of each child on each construct (representation & organisation) and each protocol (categorical versus numerical). Based on these individual within-case displays, the researchers generated a clustered matrix (Miles & Huberman, 1994, p. 127) that showed key statistical patterns for the sample of children by grade, thinking level, and protocol. These patterns were used to report and interpret the results of the study.

The Wilcoxon Signed Ranks Test (Siegel & Castellan, 1988) was used to test differences between the students' thinking levels when *representing* categorical versus numerical data. The same test was also used to assess differences between the children's thinking levels in relation to the *organisation* of categorical versus numerical data. Spearman rank correlations coefficients were calculated to find the degree of consistency between students' levels of thinking when *representing* categorical and numerical data versus their corresponding thinking levels for *organising* these two types of data. A Spearman rank correlation was also calculated to assess the degree of consistency between students' ability to actually organise data and their ability to recognise another organisation of the same data. In each case non-parametric tests were used because levels of thinking were assumed to be ordinal rather than interval or ratio data. Moreover, distributions of levels of thinking were expected to deviate from *normality*.

Results

The presentation of the results is linked to the four research questions posed in the introduction. Even though organising data presumably precedes representing data as a mental action, we have discussed representation prior to organisation because the basic source of data was children's representations, and organisational thinking was inferred from these representations.

Question 1: How is Children's Statistical Thinking in Representing and Organising Data Related to Grade Level?

Representing data. Table 1 gives the children's levels of thinking, for three children in each grade, for the construct representing data. The levels of thinking were based on their representations of both categorical and numerical data.

Table 1
Children's Levels of Thinking by Grade and Type of Data (categorical v numerical)

Grade 1		Grade 2		Grade 3		Grade 4		Grade 5	
C[a]	N[b]	C	N	C	N	C	N	C	N
1	1	1	2	4	3	3	2	3	2
1	1	3	3	3	2	3	2	2	2
1	1	3	1	1	1	3	2	3	3

Note. $n = 15$, with 3 children at each grade level. [a]C stands for categorical data; [b]N stands for numerical data.

Categorical data were associated with the protocol on children's modes of transport (see Figure 1) while numerical data were associated with the protocol on numbers of fish (see Figure 2).

With respect to statistical thinking in representing data, Table 1 reveals a sharper difference between Grade 1 thinking levels and thinking levels beyond Grade 1 than it does between thinking levels for Grades 2 to 5. Although there are differences between students' statistical thinking in relation to categorical and numerical data, the dichotomy between Grade 1 children and the rest of the children holds for both types of data. Moreover the similarities in levels of thinking across Grades 2 to 5 are consistent for both types of data.

As shown in Table 1, all 3 children in Grade 1 exhibited Level 1 thinking in representing both categorical and numerical data. By way of contrast, 11 out of the 12 children beyond Grade 1 exhibited at least Level 2 thinking in one or more protocols. For categorical data, 2 (out of 12) children exhibited Level 1 thinking, 1 child exhibited Level 2 thinking, 8 children exhibited Level 3, and 1 child exhibited Level 4 thinking. For numerical data, the corresponding numbers were 2 children at Level 1, 7 children at Level 2, and 3 children at Level 3.

All 3 Grade 1 children produced Level 1 displays, that is, their representations of the data were incomplete or idiosyncratic. For the categorical data, they could identify the four transport categories (bicycle, bus, car, and walk) and were able to represent all or some of these in their drawings. However, their representations showed no evidence that they made links between children and transport categories. Kylie's representation was typical. She simply drew a bus, a car, and a child walking. Jason went a little further. He attempted to represent a relationship between modes of transport and number of children. However, his attempt was essentially idiosyncratic, utilising "his own data set" rather than the given data set. For example, when asked "Why did you draw 3 people [in the bus]?" Jason responded, "One's a mum and 2 are kids." There was no link with the four children in the actual data who came by bus. When dealing with numerical data, the Grade 1 children produced largely idiosyncratic representations. They usually eliminated the link with the people who owned the fish and drew pictures of fish or groups of fish that did not relate to the given data. For example, Bruce drew 14 fish that bore no relation to the context of 10 children and 40 fish.

The students beyond Grade 1 who exhibited Level 1 responses produced incomplete representations rather than idiosyncratic ones for both categorical and numerical data. Their incomplete representations were similar to the incomplete representations of the Grade 1 students. Only one student beyond Grade 1, Joshua (Grade 5), exhibited Level 2 thinking for categorical data but 7 students exhibited Level 2 thinking for numerical data. Joshua did not reorganise the categorical data but maintained the essence of the data by showing the name of each child in the sample together with an icon that represented the mode of transport used by that child. Joshua's reluctance to create a display that reorganised the categorical data was not typical of students in Grades 2 to 5. However, this reluctance to reorganise was more typical for numerical data. For example, the 7 children exhibiting Level 2 thinking for numerical data generally produced valid graphs that maintained the original organisation of the data (child by number of fish). Four of the children drew pictographs and 3 drew bar graphs.

Of the 8 students beyond Grade 1 who exhibited Level 3 thinking in representing categorical data, 6 drew bar graphs, 1 drew a tally graph and 1 drew a pictograph. They showed a clear link between mode of transport and the number of children who used that mode. As such they reorganised the data and accommodated the data reduction that resulted from the children's names being omitted. With respect to numerical data, 1 student in each of Grades 2, 3 and 5 exhibited Level 3 thinking. Two of them produced a tally table (number of fish by frequency) while the third (Aldo) generated a line pictograph that was based on frequencies and used people icons to represent the number of people (frequency). Figure 4 shows Aldo's representation of the data.

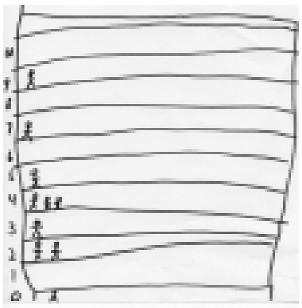

Figure 4. Aldo's representation.

Organising data. Table 2 gives the students' levels of thinking, by grade, for organising data. The levels of thinking for organising refer to both categorical and numerical data.

The changes in levels of thinking, by grade, for organisation of data, mirrored those for representation of data. Once again the organisational thinking of the three Grade 1 students was assessed at Level 1 and was noticeably lower than the children in the other grades for categorical data but less so for numerical data. The overall differences among Grades 2 to 5 children were relatively small, even though the levels of thinking for numerical data were substantially lower than those for categorical data.

Table 2

Students' Levels of Thinking, by Grade, with Respect to Organisation of Both Categorical and Numerical Data

Grade 1		Grade 2		Grade 3		Grade 4		Grade 5	
C[a]	N[b]	C	N	C	N	C	N	C	N
1	1	1	2	3	3	3	1	3	1
1	1	3	3	3	1	3	1	1	1
1	1	3	1	1	1	3	1	3	3

Note. $n = 15$, with 3 children at each grade level. [a]C stands for categorical data; [b]N stands for numerical data.

The organisations of data by the Grade 1 students reflected Level 1 thinking for both categorical and numerical data, and were either incomplete or idiosyncratic. Twenty-five percent of students beyond Grade 1 also exhibited Level 1 thinking on the categorical data. Like the Grade 1 students, they generally listed the transport categories without links to the number of children. Joshua was the exception in that he reproduced the raw data without reorganising it. With respect to *numerical* data, the situation for the 9 (out of 12) children in Grades 2 to 5 who exhibited Level 1 thinking was quite different. Whereas all Grade 1 students exhibited idiosyncratic thinking with respect to numerical data organisation, 7 out of these 9 children produced a valid representation. Their representations were typically pictographs or tables showing the students' names and the number of fish essentially as they were presented in the raw data. These students saw no need to reorganise the data. In their minds it could be represented perfectly in its raw form.

Nine students in Grades 2 to 5 exhibited Level 3 thinking with respect to the categorical data. They correctly organised the data by transport categories and were able to justify their organisation. Alex (Grade 4) was typical of this group when he remarked, "That just tells me how many walk, how many ride a bike ...." Unlike the Level 1 thinkers, children exhibiting Level 3 thinking did not seem to have any anxiety about losing the children's names through reorganisation of the categorical data. The situation was different for numerical data in that only 3 students beyond Grade 1 performed a reorganisation of the data. Tilli (Grade 2), Sally (Grade 3), and Aldo (Grade 5) reorganised the data by numerical groups (those who had 0 fish, those who had 1 fish, etc.) and used frequencies to represent the number in each numerical group. For example, Aldo drew a pictograph with the number of fish shown vertically and the frequencies (represented by stick-people) shown horizontally (see Figure 4). Moreover, all of these students were able to explain their reorganisation of the data. For example, Aldo explained his organisation as follows: "I've got numbers from 0 through 10: 1 person had zero fish, 0 people had one fish, ... and 1 person had nine fish." Sally and Tilli used a similar reorganisation to Aldo but represented it using tally graphs. However, Tilli did not show all of the numbers of fish on her tally graph. In fact, she omitted numbers like 1 and 5 (fish) where the frequencies (of students) were zero. Hence, although there were differences in their representations, the key feature that distinguishes these three students' thinking is their readiness to carry out and explain a frequency reorganisation of numerical data.

Question 2: What is the Effect of Data Type (Categorical Versus Numerical) on Children's Statistical Thinking in Representing and Organising Data?

A Wilcoxon Signed Ranks Test (Siegel & Castellan, 1988) was used to test for differences between students' *representations* of categorical and numerical data. In implementing this matched-pair test we compared each child's level of thinking in representing categorical data with their level of thinking in representing numerical data. The Wilcoxon Test showed that there was a significant difference between the two sets of representations in favour of the categorical data representations (*Mdn* [categorical] = 3, *Mdn* [numerical] = 2, $p = .05$).

Looking at this situation more closely (see Table 1) we note that the differences in representational thinking for the two data sets occurred for students beyond Grade 1. The Grade 1 students' were all at Level 1 for both categorical and numerical data. For the students beyond Grade 1, Table 1 reveals that 9 out of 12 were at Level 3 or above when representing categorical data compared with 3 out of 12 for representing numerical data. As discussed earlier, this difference in thinking level resulted from the fact that Level 3 thinking in representing data required students to show evidence of being able to reorganise the data while Level 2 did not. The fact that there were 50% more reorganisations of categorical data than numerical data by these children highlights the key difference in the children's representational thinking with categorical data versus that with numerical data.

A Wilcoxon Signed Ranks Test (Siegel & Castellan, 1988) was also used to test differences between children's *organisations* of categorical and numerical data. Children's levels of thinking for each type of data organisation were again used as the comparison measure. The Wilcoxon Test showed that there was a significant difference between the two types of organisational thinking in favour of the categorical data (*Mdn* [categorical] = 3, *Mdn* [numerical] = 1, $p = 0.03$).

The result is a slightly more pronounced version of the differences between categorical representations and numerical representations. Once again the key difference was that the majority of children (9 out of 12) beyond Grade 1 were able to reorganise categorical data (Level 3 thinking) but only 3 out of 12 were able to reorganise numerical data.

Question 3: Is Children's Statistical Thinking in Representing Data Related to their Thinking in Organising Data? Is this Relationship Influenced by Whether the Data is Categorical or Numerical?

For categorical data, there was a significant Spearman rank correlation between children's thinking levels on representation and their thinking levels on organising data ($r_s = 0.95$, $p < 0.005$). There was also a significant but smaller correlation between children's thinking levels on representation and organising numerical data ($r_s = 0.87$, $p < 0.005$). While we cannot assume a causal or even directional relationship between organising and representing data, the correlations do indicate a strong association between the cognitive activities that children undertake in organising and representing data.

In responding to the first research question we have already provided qualitative evidence that helps explain the difference between the strengths of the correlations for categorical data and numerical data. With *categorical* data, organisation appears to be a necessary and sufficient condition for representation of data. That is, the only students who could not represent data in a viable way were the ones who could not organise the data into transport categories. Moreover, all of the students who could organise the data by transport categories could also represent it in a valid way.

Within *numerical* data the situation was different because 7 out of the 15 students were able to provide a valid representation but did not reorganise the data; that is, did not build an organisation of the data that was structurally different from the raw data. In fact, our qualitative evidence suggests that these students did not see a need to reorganise numerical data. For example, note how easily Leslie (Grade 5) supports maintaining the organisation of the raw data, "Oh, I know a pictograph. I worked out how many fish they had, like 7, so I drew 7 fish. I just looked at how many fish each one had, so I drew that many fish." The fact that these students chose not to reorganise might indicate that they were reluctant to lose the data labels (children's names) or to engage more generally in data reduction. Although such a position may have limited their predisposition to generate multiple representations of data, it was not an impediment to representing numerical data in its raw form.

While the size of the sample in this study necessitates caution in drawing conclusions about the influence of data type on statistical thinking, the study provides some evidence that children's thinking in organisation and representation are in greater synchronisation when the data are categorical than when they are numerical. In an overall sense there is some evidence in this study to suggest that children's statistical thinking in both organising and representing data need to be in harmony for them to demonstrate real flexibility in representing data.

Question 4: Is Students' Performance in Recognising Organisations of Data Related to their Performance in Creating Data Organisations?

In order to assess students' performance in *recognising* organisations of data, Protocol 2 (see Figure 2) contained a series of questions on Mary's hypothetical representation of the Pet Fish Club data (see Figure 3). Mary's representation was a line plot which organised the data by number of fish and frequency. Each child's response to Mary's organisation was assigned a level of thinking by adapting the descriptors of the Framework that related to organising/grouping data. In essence, the descriptors were changed so as to reflect *recognition* of grouping rather than *creation* of grouping. For example, the Level 3 descriptor, "groups or orders data into classes and can explain the basis for grouping" was adapted to "recognises a grouping or ordering of data into classes and can explain the basis for the grouping." Subsequently the students' levels of thinking for recognition of Mary's organisation were correlated against their levels of thinking for actually organising the Pet Fish Club data. There was a significant correlation between recognition and creation with $r_s = 0.69$ and $p = 0.005$. This represents approximately 48% of shared variance between the two mental actions.

In analysing children's individual data, we observed that all 3 of the students who reorganised the Pet Fish data (Tilli, Sally, and Aldo) also recognised and explained the organisation in Mary's line plot. That is, these children exhibited Level 3 thinking whether creating or recognising data reorganisation. Moreover, they did this with consummate ease connecting Mary's representation to their own. For example, Tilli (Grade 2) said, "It's the same in a different way, it's got X's, I've got tally marks." Having noted the connection with their own representation they all went on to explain the meaning of the crosses as if they were tally marks. Tilli's response was typical, "The crosses mean that they have fish—that many people have fish. Two people have 2 fish, …."

A more interesting result is the fact that 3 children (Marcia, Lesley, and Joshua) gave valid and complete explanations (Level 3 thinking) of Mary's organisation of the Pet Fish data even though they had not previously reorganised the data in constructing their own representation. These students also handled the explanation with ease and spontaneity suggesting that they were able to understand this kind of data reorganisation even though they hadn't chosen to use it in their own representation. Marcia exemplified the kind of response given by these 3 children, "The number of fish. Oh, 1 person had zero fish, 2 people had two fish, 1 had three, 3 had four fish, …. Each of the crosses is a person."

Three other children (Anton, Adrian, and Meg) gave partial explanations to Mary's organisation of the Pet Fish data even though they had not shown any inclination to reorganise the data in their own representation. Their thinking was less spontaneous than the previously mentioned 6 children and exhibited Level 2 thinking rather than Level 3. However, as the following typical excerpt of Anton (A) shows, they were able to make accommodations in response to the researcher's (R) questions.

R: Can you read what it (Mary's graph) says?
A: Yes. Actually it's not a very good graph, because it doesn't tell *who* has 3 and who has 1. [Italics added.]
R: What do you think these crosses mean? [Points to 3 crosses above the 4 fish mark]
A: 3 fishes. Um, which one?
R: What if I picked that cross there? [R points to 1 of the crosses above the 4 fish mark.]
A: Ah! It means one 4. Which way is it going?
R: What does this say along here? [R points to the horizontal axis.]
A: Number of fish.
R: OK. So what does this cross mean? [R points to the cross above 9 fish.]
A: Oh! 9. Izra had 9 fish. [Anton looks back at the raw data.]
R: I wonder what that cross next to the zero means?
A: Gary.

At first, Anton was confused because he couldn't find the children's names on Mary's graph. He considered the crosses to be fish possibly because that was more compatible with the format of the raw data. Eventually, Anton was able to recognise that each cross represented a person but only after he had actually linked up the single crosses for 9 fish and 0 fish with the children in the raw data. Anton went on to explain the organisation in Mary's representation, "Two children have 2 fish, 1 has 3 fish, ..." but his thinking was very fragile and he constantly referred to the raw data.

Accordingly, while there is a significant association between children's ability to recognise and create data reorganisation, some of the unexplained variance (52%) may result from the fact that children's thinking in the recognition mode is stronger than their thinking in the creation mode. Our data reveal that 6 students exhibited higher-level thinking in the recognition tasks than the creation task, and no students performed at a lower level on the recognition task. Hence, there is evidence, albeit limited, that the mental actions required to create a reorganisation lag behind those needed to recognise one.

Discussion

This is an exploratory study that examined how young children organised data while they were in the process of representing data. More specifically, the study used the cognitive thinking levels of the Jones et al. (2000) Framework to gain a picture of how 15 children, 3 in each of Grades 1 to 5, represented and organised both categorical (school transport) and numerical (pet fish) data. The study also examined children's responses to a representation and organisation of the pet fish numerical data that was generated by an imaginary child.

The results of the study showed that children in Grade 1 were more idiosyncratic and incomplete in their thinking with respect to organising and representing data than their counterparts in Grades 2 to 5. These younger children did not make connections between category of transport and number of riders (nor number of fish and frequency), whereas children beyond Grade 1 made these connections, at least for the categorical data. The finding that Grade 1 children were unable to organise the categorical transport data surprised us, given that Jones et al. (2000) concluded that children in Grade 1 could reorganise *Beanie Baby* data into categories according to animal name. The fact that the children in Jones et al. could organise the Beanie Baby data may have been due to the transparent nature of the *Beanie Baby* categories and the children's familiarity with these toys. It may also have been due to the fact that the children could move concrete versions of the Beanie Babies but could not manipulate the transport categories or the fish. Whatever the reason, the result again points to the importance of mode of presentation and context in data exploration especially with young children (Curcio & Folkson, 1996; Friel, Curcio & Bright, 2001; Greer, 2000, Konold & Higgins, in press). Clearly further research is needed on *mode of presentation* and *context* in data exploration with a view to identifying and classifying tasks and contexts that are suitable for young children. Young children in kindergarten and in the primary grades regularly engage in sorting activities with blocks and other manipulatives. However, there is obviously a need for research to build a more pervasive understanding of children's sorting schemata not only in the context of manipulative blocks but also in the context of data exploration.

The ability to make connections between different aspects of the data (e.g., between transport category and number of riders) enabled students beyond Grade 1 to produce more normative (in a mathematical sense) organisations and representations of the data. Interestingly, and in sharp contrast to earlier findings (Jones et al., 2000; Lehrer & Schauble, 2000), there were minimal *grade-level* differences for students in Grades 2 through 5 with respect to the processes they used in organising and representing data. The lack of grade-level differences beyond Grade 1 may well be a manifestation of the limited sample size in this study; however, there are recent studies in the field of probability and statistics where neither age nor grade have not been significant indicators (e.g. Batanero & Serrano, 1999; Fischbein & Schnarch, 1997). What may be of more importance in this study is the fact that, as a group, the students in Grades 2 to 5 showed a preference for categorical organisations of data and the use of pictographs, bar graphs, and tally graphs when representing data. Tally graphs were also helpful in reorganising numerical data by frequencies and 2 of the 3 students who were able to produce such a reorganisation used a tally graph. The third student, Aldo, used a line pictograph (see Figure 4), which was conceptually equivalent to a tally graph.

Another potentially important finding of this study is that numerical data appears to be more difficult for students to organise and represent than categorical data. The difference between the two kinds of data did not arise with the Grade 1 children because none of them produced connected organisations of either the categorical or the numerical data. However, the difference in complexity between numerical data and categorical data was apparent with respect to the students beyond Grade 1. Only 3 out of 12 made the reorganising link between number of fish and frequency for the numerical data, whereas 9 out of 12 made the link between modes of transport and numbers of students for the categorical data. Our correlation analysis also suggests that students' thinking in organisation and representation is in greater synchronisation for categorical data than for numerical data. The difficulty these elementary children experienced in reorganising the numerical data according to frequencies is consistent with Bright and Friel's (1998) finding for middle school students.

Bright and Friel (1998) found that reorganising *ungrouped* data into *grouped* data was not an intuitive sorting process for middle-school students. The students beyond Grade 1 in our study reflected this same difficulty. Maybe it is too much of a mental leap to expect elementary and middle school students to create an organisation incorporating frequencies. Alternatively these students may simply see no need to reorganise the data by frequencies when the raw organisation and its representation by children's names and number of fish is perfectly meaningful and comfortable. The latter alternative has even more credence when it is noted that students in this study expressed reluctance about breaking the link between the children's names and the number of fish they owned (see Figure 2). Interestingly breaking the link between children's names and their mode of transport did not seem to produce the same resistance for the categorical data. Is it easier to keep track of data reduction (loss of students' names) in dealing with categories than in dealing with numbers? Do frequency representations of numerical data that have numbers on both axes, that is, "double numeric" labels, produce accommodations that are beyond elementary students' conceptions? We simply do not know the answer to this dilemma. Further research on organizing and representing data should incorporate teaching experiments or microworlds that produce perturbations capable of stimulating children to create frequency distributions. Such perturbations would need to challenge their reluctance to reduce data. In addition such research might try to explain the dispositions, meta-cognitive processes, and conceptions that lead some elementary students but not others to reorganise numerical data into frequencies.

Although the *creation* of frequency representations was rare and seemingly non-intuitive, students beyond Grade 1 revealed stronger sense making when they were asked to analyse and interpret Mary's hypothetical line plot (Figure 3) that was organised by number of fish and frequency. Fifty percent of the students beyond Grade 1, including all 3 who were able to create their own frequency representation, interpreted Mary's line plot fluently and spontaneously. The ones who had created a frequency representation simply linked Mary's representation to their own. The others discerned what was meant by the crosses (people) in Mary's line plot and then built an explanation that showed they understood reorganisation by frequencies even though they had not initiated such a reorganisation when asked to represent the original data. In addition to these 6 students, there were 3 other students who interpreted Mary's line plot after some probing. Interestingly, all 3 could only make the connection between the crosses (people) and the number of fish after they had assigned names to the crosses. In other words they appeared to have to mentally rebuild the original raw data before they could make sense of the frequencies and their connection to the number of fish. The evidence associated with Mary's line plot suggests an implication for instruction. Given the students' ability to interpret a frequency representation constructed from raw data, teachers might use recognition tasks like Mary's hypothetical line plot to build an understanding of frequency representations and to forge in students a stronger sense of the usefulness of frequency in reorganising data. The idea of presenting both the raw data and its frequency representation and having the students make links between them seems especially powerful. Such an approach might well resonate with students like those in this study who could only make sense of the frequency representation when they assigned original data names to the crosses in Mary's line plot. Consistent with this pedagogical approach, Perry et al. (1999) argued that, before learning how to construct an unfamiliar visual display, students should begin by examining and analysing the characteristics of the visual display and its connections to the original data.

The size of the sample limits the generalisability of the conclusions evidenced in this study. Further research might explore organisation and representation of data with larger groups of elementary students and in instructional settings that trace students' thinking over an extended period. Notwithstanding the limitation associated with sample size, this study has built on the statistical thinking framework of Jones et al. (2000) and has generated new insights into our understanding of elementary students' cognitive functioning when they are faced with tasks that provided them with the opportunity to organise categorical and numerical data while they were in the process of representing such data.

References

Australian Education Council (1994). *Mathematics—A curriculum profile for Australian schools.* Melbourne: Curriculum Corporation.

Batanero, C., & Serrano, L. (1999). The meaning of randomness for secondary students. *Journal for Research in Mathematics Education, 30*, 558-567.

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: Erlbaum.
Curcio, F. R. (1987). Comprehension of mathematical relationships expressed in graphs. *Journal for Research in Mathematics Education, 18*, 382-393.
Curcio, F. R., & Folkson, S. (1996). Exploring data: Kindergarten children do it their way. *Teaching Children Mathematics, 6*, 382-385.
Fischbein, E., & Schnarch, D. (1997). The evolution with age of probabilistic, intuitively based misconceptions. *Journal for Research In Mathematics Education, 28*, 96-105.
Friel, S. N., Curcio, F. R., & Bright, G. W. (2001). Making sense of graphs: Critical factors influencing comprehension and instructional implications. *Journal for Research in Mathematics Education, 32*, 124-158
Greer, B. (Ed.). (2000). Special Issue: Statistical thinking and learning. *Mathematical Thinking and Learning, 2*, 1-155.
Jones, G. A., Thornton, C. A., Langrall, C. W., Mooney, E. S., Perry, B., & Putt, I. J. (2000). A framework for characterizing students' statistical thinking. *Mathematical Thinking and Learning, 2*, 269-307.
Konold, C., & Higgins, T. L. (in press). Reasoning about data. In J. Kilpatrick, W. G. Martin, & D. Schifter (Eds.), *A research companion to NCTM's Standards.* Reston, VA: National Council of Teachers of Mathematics.
Lehrer, R., & Romberg, T. (1996). Exploring children's data modeling. *Cognition and Instruction, 14* (1), 69-108.
Lehrer, R., & Schauble, L. (2000). Inventing data structures for representational purposes: Elementary grade students' classification models. *Mathematical Thinking and Learning, 2*(1 & 2), 51-74.
Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis* (2nd ed.). 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.
Moritz, J. (2002). *Reasoning about covariation*. Manuscript submitted for publication.
National Council of Teachers of Mathematics (2000). *Principles and standards for school mathematics.* Reston, VA: Author.
Nisbet, S. (1998, September). *Displaying statistical data*. Paper presented at the annual conference of the Queensland Association of Mathematics Teachers, Gold Coast, Australia.
Perry, B., Jones, G., Thornton, C., Langrall, C., Putt, I., & Kraft, C. (1999). Exploring visual displays involving Beanie Baby data. *Teaching Statistics, 21*, 11-13.
Shaugnessy, 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* (pp. 205-238). Dordrecht: Kluwer Academic Publishers.
Siegel, S., & Castellan, N.J. (1988). *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill.
Watson, J. M., & Moritz, J. B. (2000). The longitudinal development of the understanding of average. *Mathematical Thinking and Learning, 2*, 11-50.

Authors

Steven Nisbet, Faculty of Education, Griffith University, Brisbane, QLD 4111. Email: <s.nisbet@mailbox.gu.edu.au>.

Graham A. Jones, Department of Mathematics, Illinois State University, Normal, IL 61790-4520, USA.

Carol A. Thornton, Department of Mathematics, Illinois State University, Normal, IL 61790-4520, USA.

Cynthia W. Langrall, Department of Mathematics, Illinois State University, Normal, IL 61790-4520, USA.

Edward S. Mooney, Department of Mathematics, Illinois State University, Normal, IL 61790-4520, USA.