Mathematics Education Research Journal 2003, Vol. 15, No. 3, 207-225

Graph Interpretation Aspects of Statistical Literacy: A Japanese Perspective

Kazuhiro Aoyama
University of Tsukuba

Max Stephens
University of Melbourne

Many educators and researchers are trying to define statistical literacy for the 21st century. Kimura, a Japanese science educator, has suggested that a key task of statistical literacy is the ability to extract qualitative information from quantitative information, and/or to create new information from qualitative and quantitative information. This article presents research that offers a theoretical basis using the SOLO Taxonomy to capture students' ability to create new information from qualitative and quantitative information. This research shows that the "creation of dimensionally new information" is a complex construct requiring further research and a deeper analysis than Kimura appears to have used.

Statistics has achieved a new emphasis as one of the defining elements of scientific literacy in the 21st century (American Association for the Advancement of Science, 1990, 1993; Organisation for Economic Cooperation and Development (OECD), 1999, 2000; Steen, 1990, 1997, 1999). In the 2000 Programme for International Student Assessment (PISA), one of the three dimensions of Mathematical Literacy (OECD, 1999, 2000), namely, the content strand, is organised by big ideas—*change and growth* and *space and shape*—rather than being based on traditional mathematical components such as algebra, geometry, and calculus. Data analysis and statistics are both central to change and growth. Dossey (1997) also tried to capture quantitative literacy, as defined by Steen (1997), from the perspective of categorisation of mathematical behaviours and constructed six major aspects of quantitative literacy. *Data representation and interpretation* is the first item of these six. According to Dossey (1997):

> Data representation and interpretation are perhaps the most basic aspects of quantitative literacy because they are the aspects through which people perceive data, collect information, and construct models for decision making in quantitative settings. (p. 176)

The importance of statistics as one aspect of scientific literacy has been recognised in recent curriculum documents in many countries (Australian Educational Council, 1991; Department of Education and Employment, 2000; Ministry of Education, 1992; National Council of Teachers of Mathematics, 1989, 2000). Gal and Garfield (1997) proposed a goal of statistics education that

> by the time students finish their encounters with statistics, they become informed citizens who are able to:
>
> - Comprehend and deal with uncertainty, variability, and statistical information in the world around them, and participate effectively in an information-laden society.
> - Contribute to or take part in the production, interpretation, and communication of data pertaining to problems they encounter in their professional life. (p. 2)

To achieve these goals, students' understanding and interpretation of data need to be developed. That development is now a key aspect of research. Many researchers have focused on students' ability to extract statistical information from graphs and to predict the result of a trend shown graphically (Ben-Zvi & Arcavi, 2001; Curcio, 1981, 1987; Friel, Curcio, & Bright, 2001; Watson & Moritz, 1999). Curcio (1981, 1987) defined three levels of graph reading and studied the effect of prior knowledge, reading achievement, mathematical content, and gender on graph reading ability. Watson and Moritz (1999) focused on students' statistical thinking under the setting of different sample size. To judge among groups of different sample size, appropriate use of the arithmetic mean and proportional reasoning are needed. Watson and Moritz analysed the changes in students' statistical thinking from a cognitive development perspective. Ben-Zvi and Arcavi (2001) researched the process of students' acquisition of global views of data from the perspective of enculturation. Friel, Curcio, and Bright (2001) reviewed prior research on graph reading and identified factors influencing graph comprehension. They also defined *Graph Sense*, which covers all tasks related to graphs, including graph making and reading graphs. In Japan, textbooks and the National Course of Study (Ministry of Education, 1998a, 1998b, 1999) focus more narrowly on processing and converting statistical data into graphs and reading simple quantitative information from graphs. Graph interpretation, in these school documents, hardly ever extends to making qualitative interpretations of statistical information.

Kimura's Work

Kimura (1999), a distinguished Japanese science educator, has suggested that a key component of statistical literacy is the ability to extract qualitative information from quantitative information, and/or to create new information from qualitative and quantitative information. Kimura argues that to be an intelligent citizen in this information age, the abilities to think critically using statistics and to create dimensionally new information are important. His notion of statistical literacy is rich and complex, and is intended to challenge a weaker expression of statistical literacy that is presented in Japanese textbooks and in the National Course of Study (Ministry of Education, 1998a, 1998b, 1999).

The focus of Kimura's work is on data interpretation. Kimura's notion of statistical literacy is consistent with that of Gal (2002) "as the ability to interpret, critically evaluate, and communicate about statistical information and messages" (p. 1). Those who assemble and manage information clearly need to create appropriate forms for its presentation, but everyone needs to be able to understand and interpret that information. Statistical literacy becomes all the more important for students and citizens as technology provides increasing access to information for domestic, leisure, and work purposes. This abundance of information in various forms requires the increased capacity to evaluate critically what is being presented.

Significance of Kimura's Ideas

Although not well known outside Japan, Kimura's work is significant for four reasons. First, Kimura is one of a number of contemporary researchers and writers who strongly advocate that statistical literacy be seen as a requirement for informed citizenship, and hence an important area for the school curriculum. Shaughnessy, Garfield, and Greer (1996) refer to Paulos (1988, 1995), for example, to illustrate a growing conviction of the importance of having an informed public with a better "feel" for numbers and data, and capable of contributing to the quality of public discourse by using data in newspaper and other media (p. 226). In endorsing this view, Gal (2002) claims that "statistical literacy is a key ability expected of citizens in information-laden societies" (p. 1). Definitions of statistical literacy, or, to use Kimura's expression, of statistical thinking, will be discussed later, but common elements include both knowledge and dispositions. Gal's (2002) knowledge elements include literacy skills, statistical and mathematical knowledge, knowledge of contexts, and critical questions. These have to be supported by beliefs and attitudes conducive to investigation and critical analysis (p. 4). Kimura also sees statistical thinking as a complex and interrelated set of knowledge and dispositions. Especially relevant to Kimura's position and to this paper is what Gal (2002) refers to as people's disposition "to discuss or communicate their reactions to such statistical information" (p. 3).

Second, Kimura's emphasis on graphical representations of statistical phenomena is consistent with the views of Tufte (1983) who drew attention to the importance of and challenges posed by graphical displays as a means of communicating complex ideas, and also with those of Tukey (1977) who noted the strong impact of visual phenomena in presenting quantitative information. Kimura's focus on statistical thinking in relation to graphical representations is supported by Gal (2002) who notes that "the many examples of contexts where statistical literacy may be activated indicate that most adults are consumers (rather than producers) of statistical information" (p. 3). Both authors insist on a high level of skills required to be a discerning consumer.

Third, Kimura's proposed hierarchy of statistical understanding reflects a growing trend among researchers to use frameworks denoting increasing sophistication in statistical thinking that have emerged from cognitive and developmental psychology (Shaughnessy, Garfield, & Greer, 1996, p. 206). This trend is exemplified by a three-tiered approach used by Watson (1997) to describe the skills needed to interpret statistical information presented in society. Watson also refers to related frameworks presented by Biggs and Collis (1982, 1991), and Case (1985). The SOLO Taxonomy developed by Biggs and Collis (1982), with its focus on the structure of observed learning outcomes, is especially relevant to the latter part of this paper.

Finally, any proposed hierarchy implies an invitation for researchers to test and probe what is being claimed. Gal (2002) lists a major challenge for research "to identify levels of statistical literacy in a similar fashion to the continuum proposed to describe levels of scientific literacy" (p. 21). According to Gal (2002), "clarity on the characteristics of the building blocks of statistical literacy is needed before other questions can be addressed in earnest regarding assessment and instruction focused on statistical literacy" (p. 3). The research reported in this paper is intended to contribute to this important agenda.

This article outlines Kimura's levels of statistical ability and discusses some of the tasks that he used to investigate primary students' ability. The main focus of this article is on Kimura's Level F, the highest level, which refers to the ability to draw dimensionally new information from quantitative data representations. The research reported in this article is based on research carried out by one of the authors (Aoyama, 2002). This research points to the importance of analysing students' performances in Level F from a theoretical perspective. It highlights attractive features of Kimura's position and also draws attention to other features that are more problematic for researchers. Some of these reservations indicate a need for greater conceptual clarity if Kimura's work is to inform research. It is also argued that a stronger empirical basis is needed to support Kimura's levels and to explain the relationships among them.

Kimura's Categories of Statistical Ability

Kimura (1999) argues that in order to make maximum use of the power of statistics, students have to learn how to derive new information from statistical information. Kimura proposed six levels of statistical ability, shown in Table 1, each of which involves dealing with statistical information.

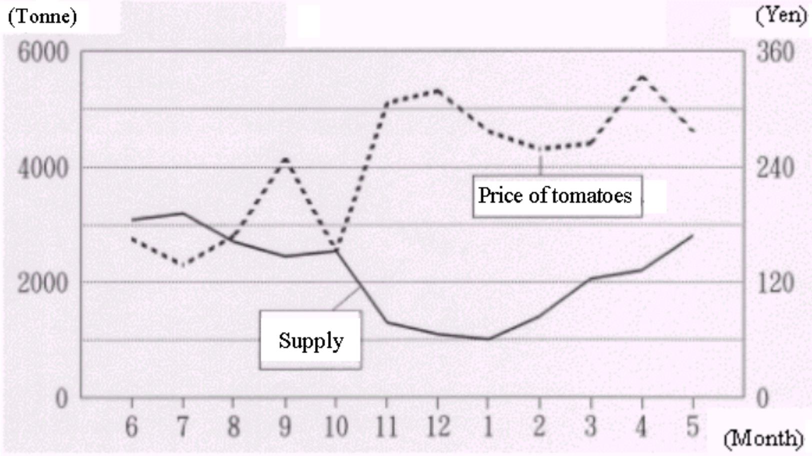
Table 1
Categories of Statistical Ability (Kimura, 1999)

Level	Features
Level A1	Basic reading of tables and graphs
Level A2	Reading key features from graphs
Level A3	Comparing information from two graphs
Level A4	Reading a simple trend in graphs
Level B	Knowing what constitutes an appropriate source of data for a given question
Level C	Statistical computation skills
Level D	Reading global trends in graphs
Level E	Extracting qualitative information from quantitative information
Level F	Creating new dimensional information

The order from A to F does not imply a strict sequence of development. Levels A, B, and C, for example, can be seen as different facets of the same ability. Levels A, D, E, and F, however, do imply a developmental sequence according to Kimura, in relation to the interpretation of statistical information. Levels A, D, and E will be discussed briefly in order to show the context in which Kimura understands Level F to operate.

Level A captures to a large extent the basic skills that students are taught in school mathematics in Japan and in other countries. Level A1 is about reading the title or theme of a graph, or naming its units, or locating particular values. Curcio's (1987) *Read the Data* corresponds to Level A1. Level A2 is about reading maximum and minimum values in graphs, or differences between values, or finding ratios between several values. Level A3 refers to the ability to compare two features in a graph or graphs. Level A2 and Level A3 seem to correspond to what Curcio (1987) refers to as the ability to *Read between the Data*. Level A4 involves reading a clear trend from data presented in a simple non-fluctuating graph.

Reading a trend from fluctuating phenomena, like stock prices, corresponds to Level D. Curcio's (1987) reference to *Reading beyond the Data* appears to correspond most closely to this level. Level D is especially important in dealing with information presented as a time series. Basic questions need to be asked, such as, "Is there a trend?" and "What predictions can be made on the basis of the trend?" In Figure 1, for example, the first part of the task asks students to identify an overall trend in the price of tomatoes.



(The horizontal axis represents months of the year. 6=June, 7=July, 12=December, etc.)

D Look at the graph and think about the price of tomatoes, then choose the correct statement.

1. Tomatoes are expensive in winter and cheap in summer. But overall, the price of tomatoes is becoming gradually higher.*
2. Although the price of tomatoes has changed up and down, overall it is about the same.
3. Because the price of tomatoes continues to change, I can't say anything overall about the price.

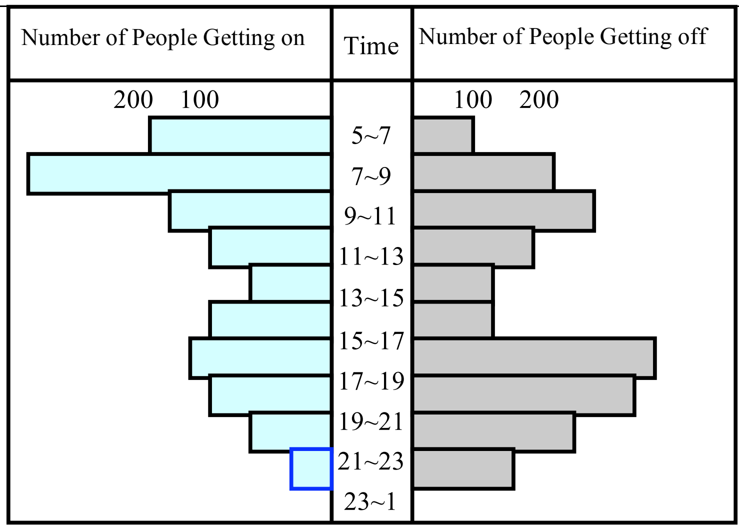
E Based on the information shown in the graph, please choose the most suitable statement below.

1. When the weather is warm, the price of tomatoes is high, and becomes cheaper when the weather is cold.
2. When the weather is warm, the supply of tomatoes increases, and when cold, the supply decreases.
3. Since many people buy tomatoes when they are plentiful, the price of tomatoes increases.
4. When the supply of tomatoes is low, their price is generally high, and it is generally cheap when the supply of tomatoes is high.*

Figure 1. Sample question items used by Kimura for Level D and E. (Correct responses are indicated by *.)

Level E involves extracting qualitative information from quantitative information. For example, the second part of the task in Figure 1 requires students to relate the price of tomatoes and their supply. Students are required to extract qualitative information, such as "the supply of tomatoes influences their price" or "the price of tomatoes becomes cheaper when they are in more plentiful supply," from the quantitative information. This information is not new, however, because it is directly implied by the information embedded in the graphical representation.

By contrast, Kimura's Level F involves creating dimensionally new information from, or imposing new information on, pre-existing qualitative and quantitative information. Two interpretations can be offered to elucidate Kimura's meaning. In the sense of *creating new information*, it may take the form of positing a likely explanation of a trend shown in a graph. In this sense, it may refer to a capacity to provide an explanation that goes beyond the data. In the other sense, that of *imposing new information*, it may involve asking questions about the quality or reliability of the data or the conclusions drawn from data, such as represented in opinion polls. Level F can be viewed as the signature feature of Kimura's levels. It is also the most difficult to interpret and analyse. The task in Figure 2 was used by Kimura to investigate this level.



Train Station X

F Please choose the most suitable explanation.

Station X could be located:
1. near a big city
2. far from the city*
3. in the center of a big city

Station X is also likely to be close to:
1. a shopping area
2. a factory area
3. a residential area*

Figure 2. Sample question item used by Kimura for Level F.
(Correct responses are indicated by *.)

The graph in Figure 2 shows the number of people getting on and off the train at Station X according to the time of day. In the morning, there are more people getting on the train, and in the evening there are more people getting off the train. One explanation of this pattern could be: "Station X is in a dormitory suburb that has many office workers who work for companies in a big city." This is only one of a range of acceptable explanations. An alternative explanation could be: "Station X is next to a ferry terminal connecting a small town to a larger town." The important feature of these explanations is that they do not involve simply an extraction of qualitative information from pre-existing quantitative data as in Level E. These explanations essentially involve the creation of a scenario to account for the data. The scenario contains new information—Kimura's "new dimension"—in order to supply an explanatory context or set of conditions to the data. For the task in Figure 2, students are required to superimpose some characteristics on the area served by Station X in order to explain the pattern of movements of passengers throughout the day.

The Challenge of Kimura's Level F

According to Kimura, Level F represents the highest level of statistical thinking that a school system should have as its goal, and that he believes is required for intelligent and informed citizens. Other statistics educators such as Watson (1997) and Gal (2002) also propose end points of a statistical thinking hierarchy. For Watson, this refers to students having both the ability and confidence "to challenge what they read in the media" (p. 110). Gal (2002) also points to the importance for adults "to be concerned about the validity of messages, the nature and credibility of the evidence underlying the information or conclusions presented, and to reflect upon possible alternative interpretations of conclusions conveyed to them" (p. 17). There is a distinction between Kimura's Level F and what these authors refer to as the ability and confidence to ask critical questions regarding statistical information. Similar to the ability to take a critical stance in evaluating statistical information, however, Kimura's Level F presupposes a capacity to step back from what is directly implied by the data, and to bring to that stance additional knowledge of other factors and of likely interpretative frameworks. For researchers and educators, the challenge is, as Watson (1997) suggests, to understand and "be sensitive to the processes students use to learn critical statistical thinking, the stages of development, and the processes available to facilitate this contextual statistical learning" (p. 121). Kimura's Level F is a prerequisite for critical statistical thinking. The remainder of this paper aims to probe more deeply the types of performances implied by Kimura's Level F. Without a more careful and clearer analysis of what is implied by Kimura's Level F, it is not possible to address the critical tasks that Watson has identified.

Results and Limitations of Kimura's Research

Using a range of questions, including those shown above (Figures 1 and 2), Kimura (1999) carried out a pilot study among Japanese students, across Grades 4, 5, and 6, in order to show relationships among his categories. Tasks used in the pilot study were intended to test performance on Levels A, D, E, and F only. Approximately 120 students were involved at each grade level. Kimura's findings indicated high levels of performance overall on tasks representing Level A. Performances diminished for Level D, with a range of 49% considered to be operating at this Level in Grade 4 to 69% in Grade 6. For Level E, the range was 31% to 60%, and further diminished for Level F where only 35% in Grade 4 were considered at Level F and 41% in Grade 6.

It may not be unexpected that performance at Level A should be quite high among students in Grades 4 to 6, since tasks corresponding to these levels are included in the course of study for elementary schools in Japan. By contrast, tasks corresponding to Levels D to F are not included in the content of elementary schools, and so it is understandable that the performances of these levels in Kimura's pilot study would be lower than Level A. Although almost all performances improved from Grade 4 to Grade 6, these improvements may not be the result of statistics education in elementary schools. They may be the product of cognitive development, learning in other subjects, and everyday experiences. Kimura's pilot study indicated that around 40% of Grade 6 students were able to operate at Level F. This proportion would seem, at first sight, to be rather high, given that Level F represents the pinnacle of performance for a school system to achieve, that is, after twelve or thirteen years of schooling.

Kimura's results for Level F were based on students' responses to multiple-choice items, of which Figure 2 is an example. These items did not require students to give reasons for their choices. This deficiency is specifically addressed in the research to be discussed in the following section. Analysis of students' responses also needs to include comments on how well students understood the particular context presented to them. Some contexts are likely to be more difficult for students in general or for some groups of students. Kimura's task relating to Railway Station X, for example, may be difficult for students who live in a country town or who have little experience of city life. A further limitation of Kimura's study was his use of a very small number of items to test for performance at each level.

Research into Kimura's Level F

The current study (arising from the first author's Master's course paper, Aoyama, 2002) builds upon Kimura's pilot study and looks especially at Level F tasks, which are the most complex and difficult tasks for students. As noted, Kimura's questionnaire items did not require students to give reasons for their choices. The present study required students to articulate and to justify their choices. In this section, the tasks used to probe students' graph interpretation abilities are described, and a theoretical framework is provided to analyse students' performances more deeply and to verify achievement of Level F tasks. To achieve Level F tasks, students need to read and understand all information presented graphically, and then to draw on their own knowledge related to the topic, and finally to supply a suitable scenario to explain all information from the graph.

Development of a Revised Questionnaire

To achieve Level F on a task, students need to understand and attend to relevant features of the information shown in a graph. The following research design set about analysing students' basic ability to read graphical information before analysing their performances at Level F. If we relied solely on a multiple-choice format (as was done in Kimura's (1999) study), students' concerns about a given task and differences in their thinking would not necessarily appear. To avoid this risk, a written questionnaire was designed that allowed students to state their reasons for supporting the option that they chose. In addition, all students in the study were interviewed about their responses. The questionnaire consisted of three tasks, the first two concerned with basic graph reading ability, based on Curcio's descriptors (1981) and Kimura's earlier levels, and a third task which focused on probing performance at Level F.

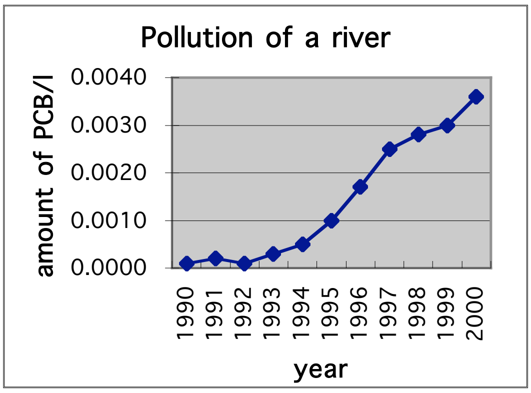
Task 1

The graph below shows the result of research to investigate the state of a river's pollution flowing through town X over the last decade.

1-1 Please choose the most suitable explanation.

1. The amount of PCB/l in 1996 is 0.0020 mg/l.

2. The amount of PCB/l in 1997 is about 0.0025 mg/l.*
3. In 1998 the amount of PCB/l is about 0.0035 mg/l.
4. In 1999 the amount of PCB/l is about 0.0015 mg/l.



1-2 Please choose the most suitable explanation.

1. The amount of PCB/l doubles from 1994 to 1995.*
2. The amount of PCB/l increases three fold from 1995 to 1996.
3. The amount of PCB/l does not change from 1996 to 1997.
4. The amount of PCB/l doubles from 1997 to 1999.

1-3 Please choose the most suitable explanation.

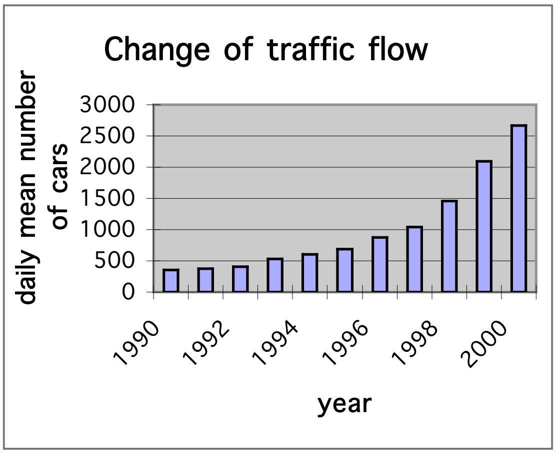
1. Since the amount of PCB/l was increasing until 2000, after that it is likely to decrease.
2. The river has not become polluted over the past decade.
3. Since the amount of PCB/l was increasing until 2000, after that it is likely to get higher unless action is taken.*
4. The river has become cleaner over the past decade.

Figure 3. Task 1. (Correct responses are indicated by *.)

Task 1 (Figure 3) and Task 2 (Figure 4) each dealt with a separate graph. The first graph shows increasing pollution of a river flowing through a town over a period of ten years. The second graph shows increasing traffic flow through the same town during the same period. Each task had three questions that correspond to Curcio's descriptors: (i) *Read the data,* which corresponds to Kimura's Level A1; (ii) *Read between the data,* which corresponds to Levels A2 and A3; and (iii) *Read beyond the data,* which corresponds to Levels A4 and D.

Task 2

The graph below shows changes in the traffic flow through town X over the past ten years. The data shows the average (mean) number of cars each day that run through one point in town X.



2-1 Please choose the most suitable explanation.

1. The daily mean number of cars in 1993 is about 500.*
2. The daily mean number of cars in 1995 is about 900.
3. In1997 the daily mean number of cars is about 1200.
4. In 1999 the daily mean number of cars is about 1500.

2-2 Please choose the most suitable explanation.

1. The daily mean number of cars is greatest in 2000, and this is about 20 times as much as traffic of 1990.
2. The daily mean number of cars is least in 1991, and it this about half as much as the daily mean number of cars in 1993.
3. The second highest daily mean number of cars is in 1999, and this is about four times as much as the daily mean number of cars in 1993.*
4. The third highest daily mean number of cars is in 1998, and this is about twice as much as the daily mean number of cars in 1997.

2-3 Please choose the most suitable explanation.

1. The daily mean number of cars has increased over the past decade, and is likely to be over 3000 in 2001.*
2. The daily mean number of cars has decreased over the past decade, and is likely to be lower in 2001.
3. The daily mean number of cars has increased over the past decade, and is likely to be about 3500 in 2010.
4. The daily mean number of cars has increased over the past decade, and is likely to be about 4000 in 2020.

Figure 4. Task 2. (Correct responses are indicated by *.)

Task 3 of the questionnaire (Figure 5) combined both graphs of Tasks 1 and 2, and was intended to examine students' Level F responses. A specific conclusion drawn by a fictitious student (Takashi) from these graphs was given for students to evaluate. Takashi concludes: "I think this town has become more urbanised in the last decade." Students were asked to evaluate his conclusion. Did they agree with Takashi, or did they disagree? Alternatively, did they believe that, based on the information presented in two graphs, Takashi's conclusion may not be strictly warranted? Students were asked in Task 3 to give reasons for their choice. In addition, all students were subsequently interviewed to elaborate further on their written response to Task 3. The interview provided an opportunity to probe students' understanding of Tasks 1 and 2 where, for example, their responses to these two tasks were both incorrect, or otherwise suggested that they had not understood what the graphs were showing. With respect to Task 3 the interview consisted of three questions:

1. Why did you choose this answer?
2. Can you explain these phenomena by other reasons?
3. Can you design any research to make clear your explanation?

Question 1 was used for all students. Question 2 was used to promote further thinking for those students who had chosen the answer "I can't judge from the information in the graphs" with appropriate evaluation. Question 3 was used for students who gave other possibilities of phenomena in graphs at Question 2.

Task 3

(In this task, students are shown the two graphs from Tasks 1 and 2 relating to town X alongside one another: Pollution of a river and Change of traffic. They are then given a statement by a hypothetical student, Takashi:)

"I think this town has become more urbanised in the last decade."

What do you think about Takashi's opinion? Please give your explanation. Put a circle around one of the three choices below and then write briefly about your reasons.

Agree Disagree I can't judge from the information in the graphs

My reasons are:……………………………………………………………………………

………………………………………………………………………………………………

Figure 5. Task 3, focusing on Level F.

An urbanization scenario is, of course, one possible explanation. In order for students' responses to be judged at Level F, it was essential that they express some clear reservations about Takashi's conclusion, even if they were inclined to agree with him. The reservations would recognise a need to draw on information beyond that given in the two graphs, and introduce what Kimura would call "new knowledge." For example, a Level F response might say: "Takashi may be right, but there could be other reasons, apart from urbanisation, which would explain the information given. For example, the pollution in the river may be caused by a factory upstream some distance from the town. The increased traffic flow may be the result of a new highway being built near the town. We can't tell for sure from the information presented in these two graphs."

Theoretical Framework for the Analysis of Results

Students' explanations were analysed using the levels of the SOLO Taxonomy (Biggs & Collis, 1982, 1991), as summarised below. Successful completion of tasks used in this study seemed to correspond to SOLO's Extended Abstract Level. It was possible to apply Unistructural, Multistructural, and Relational Levels to less sophisticated responses by students. Below are the five criteria derived from the SOLO Taxonomy that were used to analyse students' performances in this study.

1. *Prestructural.* Student engages in the task, but is distracted or misled by an irrelevant aspect.
2. *Unistructural.* Student focuses on the relevant domain, but picks up only one aspect to work with.
3. *Multistructural.* Student picks up more and more relevant or correct features, but does not or is unable to integrate them.
4. *Relational.* Student now integrates several relevant or correct features with each other, so that the whole has a coherent structure and meaning.
5. *Extended Abstract.* Student now generalises the structure to take in new and abstract features, representing a new and higher mode of operation. (Biggs & Collis, 1991, p. 65)

It is not easy to make inferences about students' underlying ability on the basis of one or more responses. First, students may not be disposed to think deeply, or even see a need to do so. They may need to be prompted to engage with the task. Second, students bring different knowledge, concerns, and viewpoints about data. For this reason, it is important to provide a format in an interview setting that allows students the maximum opportunity to show distinctive features of their own knowledge. The SOLO Taxonomy is neutral about students' motivations and backgrounds. The purpose of applying the SOLO Taxonomy is simply to categorise those responses that are observed. For that reason, the structure and design of tasks needs to be such that students are encouraged to perform at their highest potential. This feature of the task design is intended to overcome a weakness of a multiple-choice format, namely its inability to probe more deeply into students' thinking.

Employing the SOLO taxonomy, the students' responses were categorised as follows. Students who gave an incomplete or inadequate response, or who when interviewed said they did not understand the graphs, were counted in the Prestructural Level. Unistructural responses were indicated where students referred to one of the graphs only and Takashi's conclusion. When both graphs and Takashi's conclusion were connected, or simply stated to co-exist, the response was counted as a Multistructural Level response. When both graphs and Takashi's conclusion were described in an interconnected fashion, this was considered evidence of a Relational response. Takashi's response itself can be considered an example of Relational Level thinking, because he sees a possible causal relationship between urbanisation and increased pollution in the river, together with increased traffic flow through the town. Takashi does not, however, offer other possible causal explanations. Extended Abstract or Level F responses were counted only if students showed that they fully understood both graphs and at the same time they were able to evaluate critically Takashi's opinion, by suggesting an alternative explanation, or by otherwise indicating why they were not convinced that Takashi's explanation was the only one possible. This is in contrast to Kimura's study where students' responses were classified at Level F only because they chose the appropriate multiple-choice option.

Participants

The sample consisted of 55 students from Grades 5 and 8 in two government schools from Ibaraki prefecture, Japan. The schools and students were not randomly selected, but were considered to be typical Grade 5 and 8 classes by Japanese standards. By Grade 5, students have studied bar graphs, line graphs, and pie graphs in mathematics. Grade 8 students would also have studied averages in Grade 6, but would have had no further teaching in statistics in lower secondary school. The numbers of students involved in the study were 17 in Grade 5 and 38 in Grade 8.

Results

Tasks 1 and 2: Basic Graph Reading Ability

The results of basic graph reading ability are given in Table 2. Among Grade 5 students, 16 students out of 17 were able to Read beyond the Data in at least one graph. Twelve students were able to Read beyond the Data in both graphs. Students' performances may have been influenced by the fact that two different graph types were used in the questionnaire (a bar graph and a line graph) and by the technical nature of the two themes portrayed in the graphs. One student was unable to Read beyond the Data in both graphs because of a lack of understanding of each theme. For example, in Task 1 about the pollution of river, he chose the answer "Since the amount of PCB/l was increasing until 2000, after that it is likely to decrease." In the interview, he was asked why he had chosen this answer. His reason was subjective, paying no attention to the trend shown in the graph, and expressing a wish that the level of pollution would decrease. He responded that "it is better for the pollution of river to be reduced, I think." This student's response may be an instance of what Friel, Curcio, and Bright (2001) refer to as interference of prior knowledge in graph reading.

Among Grade 8 students, 36 out of 38 students were able to Read beyond the Data in one graph at least. The number of students who were able to Read beyond the Data in both graphs was 32.

Table 2
Basic Graph Reading (Tasks 1 and 2): Maximum Levels of Performance Attained

Level	Task 1 Pollution of a river		Task 2 Change of traffic	
	Grade 5*	Grade 8*	Grade 5*	Grade 8*
Read the Data	0	0	3	0
Read between the Data	3	2	0	6
Read beyond the Data	14	36	14	32

* Grade 5 (N = 17), Grade 8 (N = 38).

Task 3: Level F Task

The levels of response given by students for the third task are shown in Table 3. Examples of actual responses are discussed in the following subsections.

Table 3
Results of Level F Task (Task 3)

Level	Grade 5*	Grade 8*
Prestructural	8	0
Unistructural	6	13
Multistructural	2	10
Relational	1	7
Extended Abstract	0	8

* Grade 5 (N = 17), Grade 8 (N = 38).

Prestructural Level. When interviewed, eight of the 17 Grade 5 students could not give an appropriate answer or reason for their choice, and their responses were classified at a Prestructural Level. Some students at this level chose Agree, but they could not explain why they so chose. An example of a Prestructural Level response was "I chose Agree but I don't have any reasons."

A few Grade 5 students chose the answer "I can't judge from the information in the graphs," but their reason was that urbanisation was different from the theme of graphs. Although they could achieve Read beyond the Data in both graphs, they appeared not to know the meaning of *urbanised*. No Grade 8 students gave Prestructural responses.

Unistructural Level. Six Grade 5 students and 13 of the 38 Grade 8 students chose Agree in reference to Takashi's opinion, referring only to one graph. The following is an example of a Unistructural response: "I chose Agree because the pollution in the river increased." Among this group, there was a tendency to refer to the graph where they had previously shown that they could Read beyond the Data.

Multistructural Level. Two Grade 5 students and 10 Grade 8 students chose Agree, but referred only to both graphs without connecting either to the idea of urbanisation. An example of a response at this level was "I chose agree because the pollution in river and traffic flow both increased." These responses referred to information contained in the two graphs, but without expressing any causality.

Relational Level. One Grade 5 student and seven Grade 8 students chose Agree, referring to a possible causal relationship between the data. As an example of a response at this level one student responded "I chose Agree. Because if a town becomes urbanised, traffic flow would increase, then gases of cars would pollute the air and river." Students at this level tended to mirror Takashi in using all information and their knowledge, but they did not take the additional step to evaluate Takashi's conclusion.

Extended Abstract Level. No student in Grade 5 gave a response that was classified as Extended Abstract. Students who showed the highest response in Grade 5 could only connect the information from the graph with urbanisation, and were not able to evaluate Takashi's conclusion.

Eight of the 38 Grade 8 students chose "I can't judge from the information in the graphs" with appropriate reasoning. An example of this kind of response was as follows.

> If a town became urbanized, similar phenomena are likely to occur. But these phenomena can occur because of other possibilities. So, I can't judge whether Takashi's opinion is right or not.

Nearly all of the students at this level could not, however, actually propose other possible scenarios. They could only hint at the possibility of other scenarios. Only one student elaborated a plausible alternative, suggesting that "The pollution in the river may be caused by the pollution upstream that is not in town X." Furthermore, this same student was able to suggest some research to make clear whether town X really has become urbanised or not.

> I would investigate firstly some towns that are properly called "urban cities." I would look at some items (such as number of houses and buildings, change of population, and so on). Secondly, I would investigate town X about the same items. Then I could judge the urbanisation of town X by the similarity in items.

The responses from these two students illustrate that even among Grade 8 students there are big differences within what has been classified as Extended Abstract Level. Some students appear to recognise the possibility of other scenarios and can go no further. Although students knew the word and meaning of *urbanisation*, they were not able to articulate particular features such as growth in shops, factories, transportation, and services, to support increased population. In other words, implications of the definition were not clear for them. In contrast, the responses of the second student just quoted are more sophisticated than the other Extended Abstract responses. The student tried to assess the alleged urbanisation of town X by comparing it with some other city or cities that are recognised as urban cities.

A Second Cycle of the SOLO Model

The differences described above suggest the desirability of creating a second SOLO cycle to capture the transition from least to most sophisticated responses. At the lowest level of Extended Abstract responses, students can be seen only to recognise the possibility of other scenarios, a single aspect of the higher order functioning. Seven of the Grade 8 students appear to be at this sub-level. They cannot see, however, or elaborate on other appropriate aspects of the problem. These seven responses could be classified at a second-cycle Unistructural Level. Similarly, a second-cycle Multistructural Level could be used to explain responses that recognise the need for other possibilities, state one or more possible scenarios, but these are not integrated. Each scenario is related to one graph. Only one Grade 8 student in the study reported here elaborated a possible scenario as being pollution upstream. This scenario was drawn from attending to the graph of pollution of the river, but did not explain the change of traffic. It is also possible to envisage a second-cycle Relational Level, where students can elaborate some possible scenarios, but, in addition, they can integrate their scenarios with both graphs. Using these distinctions, a second cycle of SOLO could be used for those responses initially classified as being at the Extended Abstract Level. Using this second cycle of the SOLO model, it could be argued that no Grade 8 student performed beyond a second-cycle Multistructural Level.

Discussion and Conclusions

Implications of the Outcomes

Using a combination of written responses and interviews from students, this study showed that the proportion of students giving Extended Abstract Level responses was considerably smaller than Kimura's data had suggested. In Kimura's pilot study, about 35% of Grade 4 and 5 students and 41% of Grade 6 students were found to achieve Level F. In this study, no response by an elementary grade student was classified at Level F. It appeared that the best one could expect from elementary grade students was the ability to understand and connect the information presented.

If Level F responses are best characterised as Extended Abstract, using the SOLO Taxonomy, then it appears that few, if any, elementary grade students operate in this fashion. One of the remaining issues for further research is to extend Kimura's analysis to students beyond Grade 8. Using a methodology that was designed to interpret performance in a way that did justice to Kimura's Level F, the research reported here showed that a few lower secondary students could approach Level F. Using a SOLO classification, the study showed a clear improvement in students' ability to evaluate statistical information from Grade 5 to Grade 8. Nevertheless, a reason why many Grade 8 students could not criticise Takashi's opinion appears to be their limited experience in evaluating statistical and graphical information in both educational settings and their everyday life. Most participants said that they had no experience in evaluating an opinion based on graphical information.

Levels of Sophistication in Extended Abstract Responses

The proposed second cycle of SOLO responses has permitted a fuller elaboration of Kimura's Level F. This second cycle would indicate that performances of Grade 8 students were quite limited. This finding points to the importance of administering the questionnaire to older secondary students and to tertiary students with the goal of seeing how performances within the second cycle of the SOLO model improve with age and educational background.

The improvement observed in this study across Grades 5 and 8 is unlikely to be attributed wholly or even largely to formal statistical education given the relatively sparse treatment of statistical topics in the Japanese school curriculum between these two grades. Causes for improvement might be attributed to general cognitive development; experiences with statistical information in other courses or in everyday life; training in critical thinking; and wider knowledge of topics in biology, chemistry, and the environment that may well elucidate the contexts of tasks given to students. These are important issues for future research.

Future Research

The study reported here points to the need to gather more data on the challenge of Kimura's Level F for students in the elementary school. Clearly, more research is required to confirm the existence of further levels associated with Extended Abstract responses as proposed in the previous section. Further analysis in this area is important for curriculum planning in statistics education in Japan and in other countries.

Other important questions for further research arise from this study. The first concerns the conceptual structure of Kimura's categories. As has been noted, there is a breadth of students' performance at Level F. Does Level F include several quite distinct interpretive performances? Are all of these performances strictly statistical? Or do they relate to more general abilities for critical thinking? What elements of statistical literacy appear to be fundamental for success at Level F? Are these elements sufficiently evident and represented in Kimura's tasks? If Level F is to form a coherent structure, its defining features and its upper and lower bounds need to be defined.

Further work is also needed to link Kimura's focus on creating "new dimensional information" to what Gal (2002) and Watson (1997, 2000) refer to as critical statistical literacy. Following Watson's (1997) suggestions, it is important for researchers and educators to understand better and be sensitive to factors promoting the development of statistical thinking along lines advocated by Kimura. In order to investigate this question, it is important to extend research in older age groups of secondary school students and include some university students. Among senior high school students, are there differences in the quality of statistical thinking between those students who are taking courses in descriptive statistics and those who are pursuing more general courses?

These conceptual, theoretical, and methodological issues are shaped by one's definition of statistical literacy for the 21st century. Graph interpretation is a very important element of statistical literacy. This paper has shown that Kimura is not alone in advocating a strong definition of statistical literacy as an essential goal for schools today. Schools, however, have not paid strong attention to fostering qualitative interpretation of information presented graphically. Research is certainly needed, but the task, as this paper shows, is far from simple. Unless these issues can be framed into concrete proposals and tested by careful research, however, little is likely to change in how statistical literacy is promoted and taught in our schools and universities.

References

American Association for the Advancement of Science. (1990). *Science for all Americans. Project 2061*. New York: Oxford University Press.

American Association for the Advancement of Science. (1993). *Benchmarks for science literacy. Project 2061.* New York: Oxford University Press.

Aoyama, K. (2002). *A study about the statistical education in school mathematics: Focused on interpreting of phenomena by graphs.* Unpublished Masters Course Paper of Doctoral Program in Education, University of Tsukuba, Japan.

Australian Educational Council. (1991). *A national statement on mathematics for Australian schools*. Melbourne: Australian Educational Council and Curriculum Corporation.

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

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

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

Case, R. (1985). *Intellectual development: Birth to adulthood*. New York: Academic Press.

Curcio, F. R. (1981). *The effect of prior knowledge, reading and mathematics achievement, and sex on comprehending mathematical relationships expressed in graphs* (Final Report). Brooklyn, NY: St. Francis College. (ERIC Document Reproduction Service No. ED 210 185).

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

Department for Education and Employment (2000). *Mathematics: The national curriculum for England. Key stages 1-4*. London: Bernan Associates.

Dossey, J. A. (1997). Defining and measuring quantitative literacy. In L. A. Steen (Ed.), *Why numbers count: Quantitative literacy for tomorrow's America* (pp. 173-186). Boston: College Entrance Examination Board.

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.

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

Gal, I. (2002). Adults' statistical literacy: Meanings, components, responsibilities. *International Statistical Review, 70*(1), 1-24.

Kimura, S. (1999). Toukeizyouhoukyouikuno Karikyuramuto 5-dankaino Toukeitekitankyu Purosesu [Curriculum of statistics education and five phases of statistical inquiry process] (in Japanese). In Zentouken (Ed.), *Toukeizyouhoukyouikuno Rironto Zyugyouzissenno Tenkai*. Tsukuba Syuppankai (pp. 33-46).

Ministry of Education. (1992). *Mathematics in the New Zealand curriculum*. Wellington, NZ: Author.

Ministry of Education. (1998a). *Syougakkou gakusyusidouyouryou* [The National Course of Study of Elementary schools]. Okurasyou Insatukyoku (in Japanese). Tokyo: Author.

Ministry of Education. (1998b). *Tyugakkou Gakusyusidouyouryou* [The National Course of Study of Lower-Secondary schools]. Okurasyou Insatukyoku (in Japanese). Tokyo: Author.

Ministry of Education. (1999). *Koutougakkou Gakusyusidouyouryou* [The National Course of Study of Upper-Secondary schools]. Okurasyou Insatukyoku (in Japanese). Tokyo: Author.

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.

Organisation for Economic Cooperation and Development. (1999). *Measuring student knowledge and skills: A new framework for assessment.* Paris: OECD Publications.

Organisation for Economic Co-operation and Development. (2000). *Measuring student knowledge and skills: The PISA 2000 assessment of reading, mathematical and scientific literacy*. Paris: OECD Publications.

Paulos, J. A. (1988). *Innumeracy*. London: Penguin Books.

Paulos, J. A. (1995). *A mathematician reads the newspaper*. New York: Basic Books.

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* (Part 1, pp. 205-237). Dordrecht, Netherlands: Kluwer Academic Publishers.

Steen, L. A. (1990). *On the shoulders of giants: New approaches to numeracy.* New York: National Academy Press.

Steen, L. A. (1997). *Why numbers count: Quantitative literacy for tomorrow's America*. Boston: College Entrance Examination Board.

Steen, L. A. (1999). Numeracy: The new literacy for a data-drenched society. *Educational Leadership, 57*(2), 8-13.

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

Tukey, J. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.

Watson, J. M. (1997). Assessing statistical literacy using the media. In I. Gal & J. B. Garfield (Eds.), *The assessment challenge in statistics education* (pp. 107-121). Amsterdam: IOS Press and The International Statistical Institute.

Watson, J. M., & Moritz, J. B. (1999). The beginning of statistical inference: Comparing two data sets. *Educational Studies in Mathematics, 37*, 145-168.

Watson, J. M., & Moritz, J. B. (2000). Developing concepts of sampling. *Journal for Research in Mathematics Education, 31*, 44-70.

Authors

Kazuhiro Aoyama, Graduate School of Education, University of Tsukuba, Tsukuba City, Ibaraki Prefecture, Japan 305-8572. Email: <kazao@human.tsukuba.ac.jp>

Max Stephens, Department of Science and Mathematics Education, University of Melbourne, VIC 3010. Email: <m.stephens@unimelb.edu.au>