INVESTIGATING A HIERARCHY OF STUDENTS' GRAPH INTERPRETATION

Kazuhiro Aoyama
University of Tsukuba, Japan
aoyama@criced.tsukuba.ac.jp

The ability to extract qualitative information from quantitative information, and/or to create new information from qualitative and quantitative information is the key task of statistical literacy in the 21st century. This paper presents a hierarchy of the graph interpretation aspect of statistical literacy that includes such ability. Participants from junior high to graduate students took part and some of them were interviewed. The SOLO Taxonomy is used for decoding the students' responses and the Rasch model is used for clarifying the construction of the hierarchy. Five different levels of graph interpretation are distinguished: Idiosyncratic, Basic graph reading, Rational/Literal, Critical, Hypothesising and Modelling. These results will supply guidelines for teaching statistical literacy.

INTRODUCTION

Statistics has achieved a new emphasis as one of the defining elements of scientific literacy in the 21st century. In the PISA 2000, one of the three dimensions of "Mathematical literacy" (OECD, 1999), namely, the content strand, is organised by "big ideas," namely "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 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 and he claimed that Data representation and interpretation are perhaps the most basic aspects of quantitative literacy.

The importance of statistics as one aspect of scientific literacy has been recognised in recent curriculum documents in many countries (NCTM, 2000; Department of Education and Employment, 2000). Gal and Garfield (1997, p. 2) 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.

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 using graphs to predict the result of a trend shown graphically (Curcio, 1987; Watson and Moritz, 1999; Ben-Zvi and Arcavi, 2001; Friel, Curcio, and Bright, 2001). Curcio (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 sizes. To judge between 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, 1998) 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. A Japanese science educator, Kimura (1999) 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. Theoretical considerations of his idea, and research to capture students' ability to extract qualitative information are described in my previous work (Aoyama and Stephens, 2003). It has become obvious that to achieve the tasks of extracting qualitative information and creating new information from qualitative and quantitative information is very difficult for students, and there are some differences in the performances of students. It is necessary to investigate more deeply students' performances when they try those tasks. The purpose of this paper is to clarify the hierarchy of graph interpretation. Watson and Callingham (2003) have already reported the hierarchical nature of statistical literacy. This paper is based on their argument and is focused on graph interpretation aspects of statistical literacy.

METHODOLOGY

The participants in the study were from junior high school students through to graduate students to get different performances (see Table 1). The total number was 175. Though these participants were not randomly chosen, the result of analysis is not compromised. It is important to involve from lower to higher performances in the Rasch analysis (Rasch, 1980) as were used in this study. Of course, it is the fact that these participants are not representative of their grades and we can not prove the ordinary state of them. But the purpose of this study at this stage is to clarify the hierarchy of graph interpretation. In other words, the following questions are the keys of this study: What is a higher performance when they tried to interpret the statistical data expressed in a graph? What is a lower performance? And what are the differences between higher and lower performances?

Table 1: Sample size for each grade

Grade	Number
Junior High (7^{th})	39
High School (10th)	80
Junior College	23
Graduate	33
Total	175

All participants filled in a questionnaire including three or four themes, and each theme has from three to five questions asking about interpretation of a graph and its context. After answering, some of them were interviewed to capture their thinking clearly.

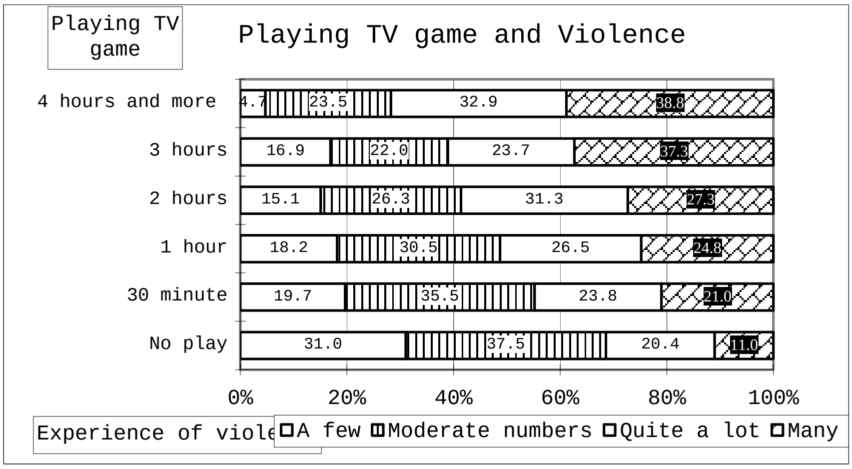
In test items, the first question of each theme is about the basic graph reading, namely to read a certain value or to compare some values. The second or third question is about the trend of the data. In the final part of each theme, participants are asked to evaluate some statements which are based on the presented graph and also asked to conjecture about the context or phenomenon (See Table 2).

Their responses are coded based on the SOLO Taxonomy (Biggs and Collis, 1991). Successful completion of tasks used in this study seemed to correspond to SOLO's "Extended Abstract Level." It was possible to apply "Uni-structural," "Multi-structural," and "Relational" levels to less sophisticated responses by students. Below are five criteria derived from SOLO those were used in this study to decode students' performances.

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 and Collis, 1991, p.65)

The following graph is the result of research to investigate how many hours elementary students play TV games at home in a day and how many experiences of violence they have. Please answer the following questions.



3-1 What is the percentage of "Quite a lot" in the group playing a TV game for one hour?

3-2 As the hours of playing a TV game increase, is the proportion of "a few" also increasing?
- a Increasing
- b Decreasing
- c Not so changed
- d I don't know

3-3 As the hours of playing a TV game increase, is the proportion of "many" also increasing?
a Increasing b Decreasing c Not so changed d I don't know

3-4 It is likely that a long time spent playing TV games will cause violence. Do you agree with this opinion? Please write your opinion.
- a Agree
- b Disagree
- c Neither
- d I don't know

Figure 1: Example of test items (translated from Japanese)

In this example, Question 3-1 is to ask participants to read a specific value of the graph. This shows whether a participant can understand and read axes of a graph or not. Questions 3-2, and 3-3 are about the trend of the graph. This shows whether a participant can compare values or proportions and understand the meaning of change. Question 3-4 is about the conclusion from these data. It is likely that it is true that to play TV games for a long time will affect one's character. But this data doesn't tell us such a thing. It is only sure that some correlation between playing TV games and violence exist. That is not cause and effect. To criticize some argument based on statistical data like this is very important for students who live in an information laden society. Question 3-4 is prepared for probing students' critical interpretation of statistical data.

After coding, all scores are analysed by the Rasch model (Rasch, 1980) to identify the hierarchy. Rasch analysis is based on correlations of task items and participants. It is hypothesised that participants who scored high points in task items have high ability in something related to this test. On the other hand, it is hypothesised that test items which were answered by very few participants are difficult and require high ability from participants. A high scored participant may be able to answer a difficult item. If there is an item which low scored participants could answer and high scored participants failed, this test item can be regarded as a noise or require other abilities from participants. Rasch analysis is based on judgements like the above. As a result, Rasch analysis shows the difficulties of each task and the total performance of each participant in numerical values.

By consideration of features of items and responses from participants who scored similar points, the character of a certain cluster is clarified. There are some clusters which partially overlap with others. So, qualitative differences of clusters are probed. Through this procedure, five different levels are found in this study.

RESULTS

Table 2 shows the reliabilities of the analysis in this study. The top two lines in Table 2 which start with “Item” show the reliability of test items. The next two lines which start with “Case” show the reliability of participants. “Separation Reliability” relates the correspondence between the differences of ability and the test result. The value is on a scale of 0 to 1, and to close to 1 means reliable. “Infit Mean Square” relates the differences between expected scores and actual scores. The value is not on a scale of 0 to 1. An acceptable range is usually from 0.77 to 1.3. “Cronbach Alpha” is calculated by variances of expected and actual values. More than 0.7 (or 0.8) is the criterion of which these data are reliable.

All indicators are high enough to suggest the existence of a hierarchical construct within these test items.

Table 2: Reliability and Fit Indices

Item Separation Reliability	0.82
Item Infit Mean Square	0.98(S.D. 0.13)
Case Separation Reliability	0.74
Case Infit Mean Square	0.97(S.D. 0.69)
Cronbach α	0.71

In the result of a Rasch analysis, all items and participants are plotted along the one numerical scale (Figure 2). The left side shows the distribution of participants. Each X represents one participant. The right side shows the distribution of items. Participants plotted at high scores have high ability. Items plotted at high scores are difficult tasks.

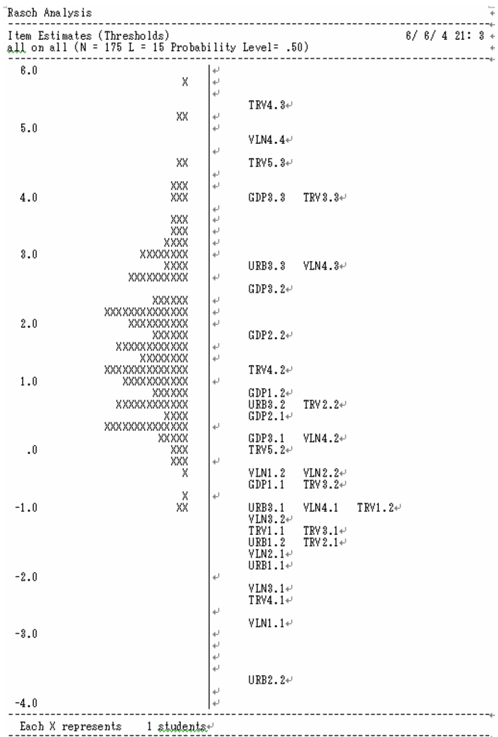


Figure 2: Distribution of participants and items

There are some borders in which the responses of participants and the nature of items change. Five different levels can be distinguished from such differences (Table 3).

Table 3: Hierarchy of Graph Interpretation

Level 1	Idiosyncratic
Level 2	Basic Graph Reading
Level 3	Rational/Literal
Level 4	Critical
Level 5	Hypothesising and Modelling

The features of each level and examples of responses are as follows.

- *Level 1: Idiosyncratic.* Students at this level cannot read values or trends in graphs. They fail to connect some features extracted from graphs with context. Usually, their responses in interview are based on their limited individual experience. For example, students in this level respond as follows to the Question of 3-4 shown in the Table 2: “Answer: a, Reason: Because if I play a TV game, I and my family sometimes fight against one another.”
- *Level 2: Basic Graph Reading.* Students at this level can read values and trends in graphs. But they cannot explain contextual meanings of trends or features, which they could see, and can’t contextualize events presented. An example of students’ responses in this level is the following: “Answer: b, Reason: Because I think that TV game and violence are not related.”
- *Level 3: Rational/Literal.* Students at this level can read values and trends. They explain contextual meanings literally in terms of features shown in a graph. They cannot suggest any alternative interpretations but use only presented meanings. They are generally unable to question the reliability of information. An example of students’ responses in this level is the following: “Answer: a, Reason: The group who play games for a long time have many experiences of violence.”
- *Level 4: Critical.* Students at this level can read graphs and understand presented contextual meanings. Still more, they can evaluate the reliability of presented contextual meaning. They can question information presented. An example of students’ responses in this level is the following: “Answer: b, Reason: Because I think that there are other causes. But I don’t know those.”
- *Level 5: Hypothesising and Modelling.* Students at this level can read graphs, and accept and evaluate some presented information. They can form their own explanatory hypotheses or models. At this level, students act as active statistics “researchers” not just as information receivers. An example of students’ responses in this level is the following: “Answer: b, Reason: Because I think that home environment seems to be a cause of both.”

DISCUSSION

The contents of statistics education in Japan are mainly focused on levels 2 and 3. Though some participants from junior high school could achieve level 4 tasks, those are not an effect of formal statistics education in school. They may learn in subjects of sociology, science and so on in school, and also in social contexts.

Level 4 was emphasised from many years ago (e.g., Huff, 1954). But it has recently become one of the priorities of statistics education in the context of statistical literacy (e.g., Watson, 1997; Gal, 2002). Because of advancing technology, children at present, and in the future, will be exposed to too much information. Information is often based on statistical surveys and is sometimes biased. It is very important for children to have an insight into such biased information. This is a task and role of statistics education in the 21st century.

Level 5 is different from level 4. Level 4 is important to an information consumer or receiver. On the other hand, level 5 is focused on information producers. When students try their own problems, they can tackle the problems in a statistical way. Statistics education must help them to learn how to implement that, and how to extract valuable knowledge from statistics data. This idea is due to Kimura (1999). To make his idea clear is one of the purposes of this study.

As mentioned above, the participants were not randomly chosen in this study, so the state of each grade is not clear. It is the next step to clarify the state and development of students.

Hierarchy of graph interpretation is clarified in this study. This result will supply the guidelines for teaching statistical literacy. But there are still more differences within the levels clarified in this study. With more probing interviews and more data, a more sophisticated construction of the hierarchy might be tried.

There are other possibilities for the construction of a hierarchy. In this study, reliabilities are high and suggest the existence of a one dimensional scale of graph interpretation ability. It means that there is no conflict to think of the ability of graph interpretation as a one dimensional construction. But to say it in a different way, the possibilities of a two or more complex dimensional construction still remains. It is the next challenge of this study to investigate these possibilities.

REFERENCES

Aoyama, K. and Stephens, M. (2003). Graph interpretation aspects of statistical literacy: A Japanese perspective. *Mathematics Education Research Journal* 15(3), 3-22.

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

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

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

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

Dossey, J. (1997). Defining and measuring quantitative literacy. In L. Steen (Ed.), *Why Numbers Count –Quantitative Literacy for Tomorrow's America*, (pp. 173-186). Boston: College Entrance Examination Board.

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

Gal, I. and Garfield, J.(1997). Curricular goals and assessment challenges in statistics education. In I. Gal and J. Garfield (Eds.), *The Assessment Challenge in Statistics Education*, (pp. 1-13). Amsterdam: IOS Press and The International Statistical Institute.

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

Huff, D. (1954). *How to Lie with Statistics*, New York: W. W. Norton and Company.

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

Ministry of Education.. (1998). *Tyuugakkou gakusyusidouyouryou (The National Course of Study of Elementary schools)*. Okurasyou Insatukyoku (in Japanese).

NCTM. (2000). *Principles and Standards for School Mathematics*. Reston, VA: Author.

OECD. (1999). *Measuring Student Knowledge and Skills -A New Framework for Assessment*. Paris: OECD Publications.

Rasch, G. (1980). *Probabilistic Models for Some Intelligence and Attainment Tests*, The University of Chicago Press.

Watson, J. (1997). Assessing statistical literacy using the media. In I. Gal and J. B. Garfield (Eds.), *The Assessment Challenge in Statistics Education*, (pp. 107-121). Amsterdam: IOS Press and The International Statistical Institute.

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

Watson, J. and Callingham R. (2003). Statistical literacy: A complex hierarchical construct. *Statistics Education Research Journal*, 2(2), 3-46.