



International Electronic Journal of
Mathematics Education

Volume 2, Number 3, October 2007

www.iejme.com

**INVESTIGATING A HIERARCHY OF STUDENTS' INTERPRETATIONS OF
GRAPHS**

Kazuhiro Aoyama

ABSTRACT. The ability to analyse 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. Although several studies have focussed on critical evaluation of statistical information, this aspect of research has not been clearly conceptualised as yet. This paper presents a hierarchy of the graphical interpretation component of statistical literacy. 175 participants from different educational levels (junior high school to graduate students) responded to a questionnaire and some of them were also interviewed. The SOLO Taxonomy was used for coding the students' responses and the Rasch model was used to clarify the construction of the hierarchy. Five different levels of interpretations of graphs were identified: Idiosyncratic, Basic graph reading, Rational/Literal, Critical, and Hypothesising and Modelling. These results will provide guidelines for teaching statistical literacy.

KEYWORDS. Statistical Literacy, Students' Interpretations of Graphs, Solo Taxonomy, Rasch Model.

INTRODUCTION

“Statistical literacy” as an important area of mathematical and scientific education has become recognized in recent years. Research into aspects of statistical literacy has also become more prominent. Gal (2002) has suggested two interrelated components of statistical literacy required by society: (a) people's ability to interpret and critically evaluate statistical information, data-related arguments, or stochastic phenomena, which they may encounter in diverse contexts, and when relevant; (b) their ability to discuss or communicate their reactions to such statistical information, such as their understanding of the meaning of the information, their opinions about the implications of this information, or their concerns regarding the acceptability of given conclusions (pp. 2-3). Gal has also proposed a component model based on his literature study. Watson (1997) discussed three levels of statistical literacy: 1) Basic understanding of terminology; 2) Embedding of language and concepts in a wider context; 3) Questioning of claims. Watson and Callingham (2003) suggested a hierarchical nature of statistical literacy, and identified six levels of this hierarchy using the results of a large survey.

In recent discussions of literacy for informed citizens of 21st century, statistical literacy has been seen by several authors as an essential component. For example, Steen (1990) argued that statistical literacy is a key component of quantitative literacy. Assessment tasks which exemplify the type of statistical literacy needed by informed citizens have been included in the Programme for International Student Assessment (PISA) conducted by the Organisation for Economic Cooperation and Development (OECD, 1999, 2003). One of tasks included in the 2003 PISA assessment is set in the context of a media report, in which a biased graph and false statement are presented. Students are expected to recognize how biased the given information is and to criticize how the information is presented (OECD, 2004).

As suggested by the definition of statistical literacy by Gal and the above PISA task, critical evaluation of statistical information is one of the most essential aspects of statistical literacy. In the hierarchy of statistical literacy, proposed by Watson and Calingham (2003), critical evaluation of information presented in a statistical form is posed as the most advanced component. Statistics is needed and used in various fields in our society such as medicine, politics or marketing. And every day the media reports results of statistical surveys or make claims based on statistics. Educating students who are able to evaluate statistical information appropriately is obviously a priority of statistics education.

Moreover, teaching students how to analyze or approach phenomena with statistics is also a key purpose of statistics education. Students need to have the experience of doing their own research including posing the questions to be investigated, designing surveys, summarising data, analyzing data, and drawing conclusions. In the phases of analyzing and concluding, it is important for students to generate models or new hypotheses. Research in statistics education needs to take into account the students' insights into statistical data and its development.

RESEARCH ON STUDENTS' UNDERSTANDING OF GRAPHS

Many researchers have focused on students' ability to extract statistical information from graphs and their using of graphs to make predictions or discover a trend (Curcio, 1987; Watson & Moritz, 1999; Ben-Zvi & Arcavi, 2001; Friel, Curcio, & Bright, 2001; Monteiro, & Ainley, 2006). Curcio (1987) defined three levels of graph reading: read with the data, read between the data, read beyond the data. She also studied the effects 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. Monteiro and Ainley (2006) investigated whether individual differences between participants who have different academic background correspond to differences in the emphasis placed on different kinds of knowledge in their interpretations of media graphs.

Despite the fact that this previous research has set the basis for our understanding of students work with graphs, more research is needed to focus on deeper aspects of students' insight into statistical graphs. Students are not only able to read and summarize statistical graphs but can also find new facts and/or make their own conjectures and hypotheses from these graphs. Such performances of students are reported in Aoyama and Max (2003). The purpose of this paper is to identify a hierarchy in students' interpretations of graphs which focuses on several of these deeper aspects. Watson & Callingham (2003) have already reported the hierarchical nature of statistical literacy (Table 1). This paper utilises their hierarchy and is focused on one aspect of statistical literacy, which is students' interpretations of graphs.

Table 1. Statistical literacy construct by Watson and Callingham (Watson & Callingham, 2003, p.14)

Level	Brief characterization of step of tasks
6. Critical, Mathematical	Task-steps at this level demand critical, questioning engagement with context, using proportional reasoning particularly in media or chance contexts, showing appreciation of the need for uncertainty in making predictions, and interpreting subtle aspects of language.
5. Critical	Task-steps require critical, questioning engagement in familiar and unfamiliar contexts that do not involve proportional reasoning, but which do involve appropriate use of terminology, qualitative interpretation of chance, and appreciation of variation.
4. Consistent, Non-critical	Task-steps require appropriate but non-critical engagement with context, multiple aspects of terminology usage, appreciation of variation in chance settings only, and statistical skills associated with the mean, simple probabilities, and graph characteristics.
3. Inconsistent	Task-steps at this level, often in supportive formats, expect selective engagement with context, appropriate recognition of conclusion but without justification, and qualitative rather than quantitative use of statistical ideas.
2. Informal	Task-steps require only colloquial or informal engagement with context often reflecting intuitive non-statistical beliefs, single elements of complex terminology and settings, and basic one-step straightforward table, and chance calculations.
1. Idiosyncratic	Task-steps at this level suggest idiosyncratic engagement with context, tautological use of terminology, and basic mathematical skills associated with one-to-one counting and reading cell values in tables.

THEORETICAL FRAMEWORK

SOLO Taxonomy

The Structure of Observed Learning Outcomes (SOLO) taxonomy (Biggs & Collis, 1982, 1991) is used for evaluating students' performances. SOLO is a framework based on cognitive development by Piagetian theory. But Biggs and Collis suggested that there are clear distinctions between development and learning. The SOLO framework does not identify the developmental stage of a certain student but identifies the level of a response to a task by a certain student. The five levels of the SOLO Taxonomy (Table 2) correspond to the various factors which students are able to take account of and imply how consistently they can operate at a particular level.

Table 2. Five levels of SOLO taxonomy (Biggs & Collis, 1991, p.65)

Level	Features
Prestructural	The task is engaged, but the learner is distracted or misled by an irrelevant aspect belonging to a previous stage or mode.
Unistructural	The learner focuses on the relevant domain, and picks up one aspect to work with.
Multistructural	The learner picks up more and more relevant or correct features, but does not integrate them.
Relational	The learner now integrates the parts with each other, so that the whole has a coherent structure and meaning.
Extended Abstract	The learner now generalizes the structure to take in new and more abstract features, representing a new and higher mode of operation.

The middle three levels of SOLO (unistructural, multistructural, relational) belong to "Concrete operational" mode, while the prestructural level belongs to a lower (Preoperational) mode, and the "Extended Abstract" level belongs to a higher (Formal) mode. Unistructural, multistructural, and relational levels are referred to as a UMR learning cycle in a target mode.

Understanding statistical claims, including terminologies, concepts, and contexts, is a performance identified within the concrete operational mode. But to criticize given statistical claims, students should be aware of some other aspects which are not presented and such performances belong to the formal mode. Furthermore, to make their own conjectures students should take into account additional aspects or resources. This ability is also included in the formal mode, but at a little higher level than critical evaluation. These differences can be seen as a second UMR cycle in the higher mode.

The students' responses to each task in this study were coded based on the five levels of SOLO. This does not mean that there are always five codes for each task. For example, in some simple tasks, it was difficult to distinguish a multistructural response from a relational one, so that in this case, only four codes were used (prestructural, unistructural, multistructural-

relational, extended abstract). In other tasks students were required not only to criticize statistical claims, but also to make their own conjectures or hypotheses. Because both responses were identified as belonging to the extended abstract level, two sub-levels within the extended abstract level were distinguished and used as codes in this case.

Rasch Model

A Rasch (Rasch, 1980) model is generally used for linking or equating tests through the use of common items or common participants (Bond & Fox, 2001). But in this study, a Rasch model was used to bring together various items in a questionnaire that was designed to measure similar aspects related to students' graph interpretations. If the different items in the various tests work together in a consistent and predictable fashion, it can be seen as evidence of a single underlying dominant variable and it can be argued that they are likely to be measuring the same construct (Bond & Fox, 2001). Placing all items on the same scale then provides an opportunity to examine the nature and validity of the underlying theorised construct.

In the Rasch model, the probability of a correct response is modelled as a function of both the person and the item parameters. Construct validity can be examined by considering the fit to the model of both items and cases. The most commonly used measure of fit is referred to as the infit statistic. The mean infit statistic has a theoretical value of 1 and the fit may be considered acceptable if its values lie between 0.77 and 1.3 (Adams & Khoo, 1996). If the items can be shown to be systematically and predictably related to each other along the variable, this is taken as confirmation that a single construct is being measured, and provides evidence of construct validity.

METHODOLOGY

The participants in the study were 175 students from different education levels in Japan: Students from junior high schools to graduate level were included in order to get different levels of reasoning (see Table 3). These participants were not randomly chosen. It was important to select groups of students with different abilities to maximise the spread of scores in the test, from low to high performances, so that a Rasch analysis could show the different levels we expected in this study. Of course, it is assumed that these participants are representative of their grades, although we are conscious that the study has an exploratory character, due to the use of an intentional sample.

The purpose of this study at this stage is to clarify a hierarchy of interpretations of graphs. In other words, we tried to answer the following questions: How can we define a high performance when students interpret the statistical data represented in a graph? What is a low performance? And what are the differences between high and low performances?

Table 3. Sample size for each grade

Grade	Age	Number
Junior High (7th)	12-13	39
High School (10th)	15-16	80
Junior College	18-19	23
Graduate	22-	33
Total		175

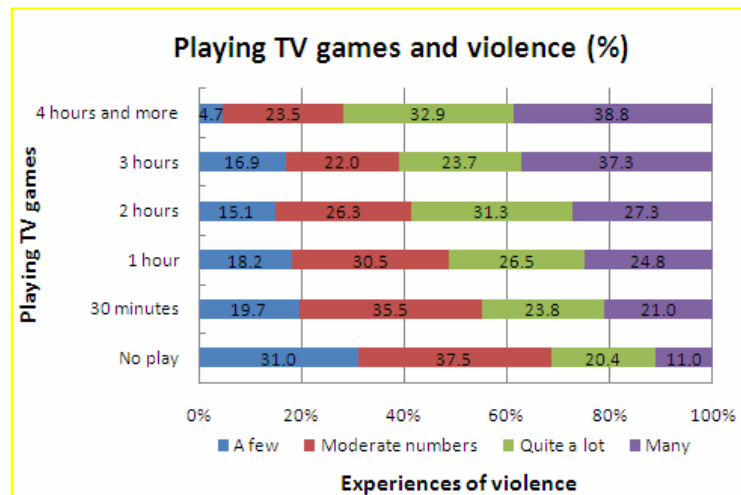
All participants completed a questionnaire including three or four items each based on a different theme, and having three to five questions asking about the interpretation of a graph and its context. After answering, participants who showed higher performances were selected to be interviewed to collect complementary data to give different evidence about the students' thinking.

In the construction of test items, the first question of each item deals with basic reading of graphs, namely students are asked to read a certain value or to compare some values. The second and third questions are related to the trend of the data. In the final question of each item, the participants are asked to evaluate some statements which are based on the presented graph and also asked to conjecture about the context or phenomenon implied by the particular graphical representation. One such item and its questions are given in Figure 1.

In this example, Question 3-1 (VLN1) asks participants to read a specific value from the graph to assess whether a participant can understand and read given values from a graph or not. Questions 3-2 (VLN2), and 3-3 (VLN3) are related to the trend of the graph and evaluate whether a participant can compare values or proportions and understand the meaning of change. Question 3-4 (VLN4) asks for a conclusion to be drawn from these data. Students are asked to judge the conclusion that playing TV games for a long time will affect a person's character. But the data in the graph does not give such information; it only suggests that some correlation might exist between playing TV games and violence, which does not prove cause and effect.

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

Item 3 (VLN). The following graph is the result of research that investigates how many hours per day elementary students play TV games at home and how many experiences of violence (e.g. hitting or pushing a classmate, or pulling someone's hair) they have.



- 3.1. What is the percentage of students experiencing “Quite a lot” of violence 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 people experimenting “a few” violence also increasing?
a:Increasing; b:Decreasing; c:Do not change; d:I don't know
- 3.3. As the hours of playing a TV game increase, is the proportion of people experimenting “many” violence also increasing?
a:Increasing; b:Decreasing; c:Do not change; d:I don't know
- 3.4. It is suggested that a long time spent playing TV games will cause violence. Do you agree with this opinion?
Please choose and write your idea.
a:Agree; b:Disagree; c:Neither; d:I don't know

The capacity to criticize this type of simplistic cause-and-effect argument is very important for students who live in an information based society. Question 3-4 is therefore intended to assess students' critical interpretation of statistical data. Those who showed higher performances in this part were selected to be interviewed in order to explore their reasoning. Higher codes in this question were given to students through the interview. A similar format was used with the other items and their sub-questions.

RESULTS

After collecting the open responses and coding them, all the data were analysed by the Rasch model in order to identify a hierarchy in the students' responses. The data were then analyzed using the Quest software (Adams & Khoo, 1996).

Table 4 shows the reliability indexes of the analysis in this study. Item and Case infit mean square are 0.98 and 0.97 respectively, lying between 0.77 and 1.3, which suggests that the tasks form a hierarchical one-dimensional scale. The value of separation reliability ranges on a scale 0 to 1, and being close to 1 implies that the items or cases are reliable. Item separation reliability 0.82 is high, suggesting that the tasks indeed describe a spread of difficulty along the variable. Case separation reliability is an estimate of how well one can differentiate persons on the measured variable (Bond & Fox, 2001) and its value 0.74 is moderate. "Cronbach Alpha" is the test reliability coefficient, that estimates the quotient between the variances of expected and observed values and usually a value higher than 0.7 is the criterion to consider the test to be reliable. All indicators are high enough to suggest the existence of a hierarchical construct being measured by these test items.

Table 4. 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 Alpha	0.71

In a variable map of a Rasch analysis, all items and participants are plotted along a single numerical scale (Figure 2). The left side shows the distribution of participants' scores. Each X represents one participant. The right side shows the distribution of item difficulty. Each item is represented by a combination of three letters and two numbers separated by a period. Each of the four items in this study corresponds to three alphabets in the variable map. The first number represents the sub-questions in each item. The second number represents the level of response (code) in each sub-question. For example, VLN4.3 means to get code 3 in the sub-question 4 of item 3 (VLN). Participants plotted at high scores are presumed to have high ability. Items plotted at high scores are more difficult tasks than those plotted lower on the scale.

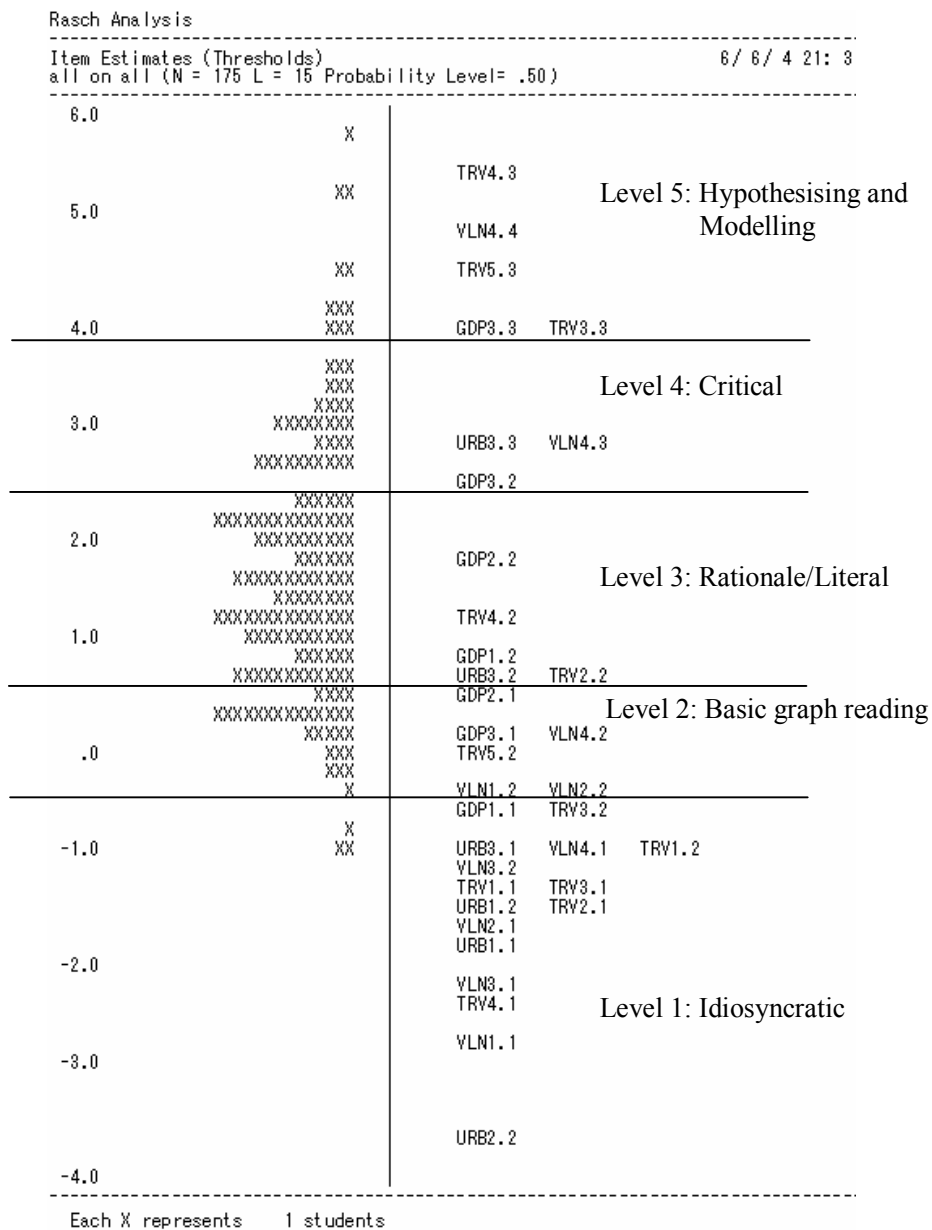


Figure 2. Variable map of a Rasch analysis: Distribution of participants and items

Questions 1 and 2 in each item are generally placed on a lower position and this result agrees with the design of the questionnaire. There is also a clear tendency for questions which ask students to criticize or make a model/hypothesis to be placed in a much higher position on the scale. For example, VLN4.3 represents critical responses (responses where students recognise correlation but do not conclude causality) for sub-question 4 in item 3 (VLN) concerning Violence and TV games (Figure 1). VLN4.4 represents responses that include a model/hypothesis in the same sub-question (students recognise correlation, do not conclude causality and refer specifically to other possible variables that might explain correlation). While

responses coded as VLN4.3 reject the presented opinion but do not propose any other explanations for the data features, VLV4.4 responses do not only deny causality but also suggest a scenario (for example, the home environment of those children) which explains the features of the data. TRV4.3, TRV5.3, and GDP3.3 that appear in the upper part of the map are also related to making a model/hypothesis. These results support the premise that the ability to interpret graphical information beyond the context of each item is a far more difficult task.

By analyzing clusters of items and participants as shown in the previous paragraph, five different levels could be distinguished. These levels are related to the hierarchy of Watson and Callingham, although the hierarchy of this study is focused on students' interpretations of graphs, that is one aspect of statistical literacy. Furthermore, to identify higher levels of students' interpretation, participants were taken from different educational levels, including graduate students. The features of each level and examples of responses are described below.

Level 1: Idiosyncratic. Students at this level cannot read values or trends in graphs. Students at this level provided wrong values when reading the graph or gave up answering the question 3-1. They failed to connect some features extracted from graphs with the context. Usually, their responses in the interview were based on their limited individual experience or on purely personal perspectives. For example, a student in this level, who agreed to the statement presented in question VLN4 (figure 1), gave the following justification: "Because if I play a TV game, I sometimes argue with my family."

Level 2: Basic Graph Reading. Students at this level can read values and trends in graphs, but they can neither explain the contextual meanings of trends or features they see, nor contextualize the events presented. An example of a response to question VLN4 in this level is the following: "Answer: Disagree, Reason: Because I think that playing TV games is different from doing some violence. So, we can't compare these things, or put them together."

Level 3: Rational/Literal. Students at this level can read particular values and trends. They explain contextual meanings literally in terms of features shown in a graph, but they cannot suggest any alternative interpretations; they only use the presented meanings. They are generally unable to question the reliability of information. An example of a response to question VLN4 in this level is the following: "Answer: Agree, Reason: The group who played games for a long time had many experiences of violence."

Level 4: Critical. Students at this level can read graphs and understand the contextual variables presented. Still more, they can evaluate the reliability of the contextual meaning described in the graph and question the information presented. For example, a student in this level, who disagreed in question VLN4 gave the following response: "Because I think that there are other causes. But I don't know which."

Level 5: Hypothesizing 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 a response to question VLN4 in this level is the following: “Answer: Disagree, Reason: Because I think the home environment seems to be a cause of both playing TV games and violence.”

Figure 3 shows the distribution of students in each of the above Levels. Very few students performed at the Idiosyncratic Level. The proportion of students performing only at Level 2 – Basic Graph Reading – was quite significant among the junior high school cohort and steadily reduced in the high school and junior college cohorts; with no student in the graduate group performing at this level. Figure 3 also shows that the majority of respondents from the high school and junior college groups performed at Level 3 – Rational/Literal. The proportion of students performing at Level 4 – Critical – remains small, with some slight variations, across the junior high school, high school and junior college cohorts. The proportion of respondents operating at Level 4 increases substantially in the graduate cohort with nearly 50% operating at this level. In this cohort, there is also clearer evidence of the emergence of performances at the highest level, Level 5 – Hypothesizing and Modelling. In the preceding cohorts, there were individual instances of students performing at Level 5 but these were not strong enough to indicate any trend. There were no instances of Level 5 performances in the junior college cohort.

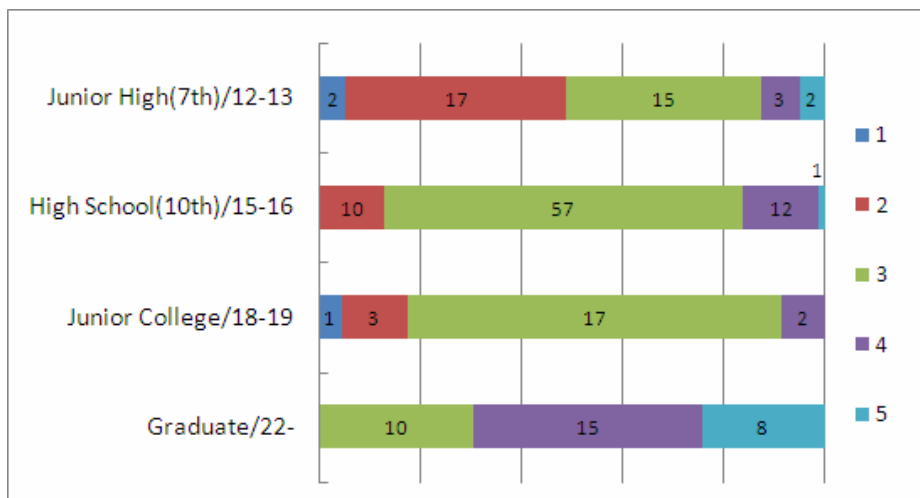


Figure 3. Number of Students in Each Level

The students’ responses to tasks were strongly affected by their familiarity with the context and their knowledge about the related phenomena. There were some participants who presented high performances in difficult tasks though failed in easier tasks. One possible reason is that they were influenced by the context of tasks because of lack of familiarity with it. For example, the context of GDP and literacy, as a social topic, appears to be less familiar than the context connecting TV games and violence for students. Although, some questions arise as a result of Rasch analysis in respect to individual participants and items, the overall distribution of difficulty shown in the variable map is valid for the global performances of students.

In summary, performances progressed with age generally, but performances of junior college students were slightly lower than those of high school students. This result possibly suggests that there are some constraints for junior college students to think more narrowly about open questions caused by their learning experiences, although this hypothesis needs to be investigated more thoroughly in view of the fact that participants of this study were not randomly chosen.

DISCUSSION

In Japan, textbooks and the National Curriculum (Ministry of Education, 1998) focus fairly narrowly on processing and converting statistical data into graphs and reading simple quantitative information from graphs. In these school documents, interpretation of graphs hardly ever extends to making qualitative interpretations of statistical information. The contents of statistics education in Japan appear to be mainly focused on Levels 2 and 3 of reading graphs. Though some participants from junior high school could perform at Level 4, this is not likely to be an effect of formal statistics education in school, and may be attributed to experiences in reading and interpreting graphs in social studies, or science in school, or in non-school contexts.

Reports from the Ministry of Education regarding the next curriculum revision for Japanese schools suggest that statistical contents will probably be enriched. Kodera and Shimizu (2007) argue that the definition of mathematical literacy posed in the framework of PISA is not just a framework for evaluating students' capacity, but represents a clear shift in the goals and contents of education for the 21st century. It implies that future curricula of Japan will reflect more clearly the PISA framework in which Japan has participated and which Japan is already committed to use for its 2007 round of national assessments in Years 5, 7, and 9.

The importance of Level 4 or reasoning was emphasised many years ago (e.g. Huff, 1954) and it has recently become recognised as one of the priorities of statistics education in the context of statistical literacy (e.g. Watson, 1997; Gal, 2002). Because of advancing access to various types of information technology, students and citizens at present, and in the future, will be exposed increasingly to information that is based on statistical surveys and/or represented using various graphical methods. It is very important for all students and citizens to have an insight into how such information is presented and to be able to ask critical questions about how the information is presented and the conclusions that are drawn from it. This is an important task and role of statistics education in the 21st century.

Level 4 in graphical understanding is important to all information consumers or receivers. Assisting more young people at the end of their secondary education to operate effectively at this level is a huge challenge in itself. On the other hand, Level 5 focuses attention on information producers; that is, those who are required to act on information presented to them,

such as those who are involved in total quality management in an assembly line production or those who have to respond to unpredictable shifts in demand in an energy plant. These people need to be able to propose alternative explanations to account for variations in the data, and to act accordingly. Statistics education must help people such as these to learn how to extract valuable knowledge from data (Kimura, 1999). As areas of manufacturing and services become increasingly data driven, more and more citizens will be expected to operate at Level 5. A capacity to operate at Level 4 will be not sufficient.

The hierarchy of interpretations of graphs, as identified in this study, will be used to prepare further guidelines for teaching statistical literacy, although it is likely that more differences can be found within the levels identified in this study. More interviews and new data, will serve to produce a more sophisticated elaboration of the hierarchy described in this paper.

There may be other possibilities for the construction of a hierarchy. However, in this study, the emergence of high reliability indexes does appear to support the existence of a single dimensional scale for graph interpretation ability. It also supports the idea of progressive levels of ability in interpretation of graphs, described by a one-dimensional sequence (e.g. Level 1 to Level 5) as attempted in this study. But this should not rule out the possibilities of a two-dimensional or even more complex construction. This study opens up a challenge to investigate these possibilities.

ACKNOWLEDGEMENT

My sincere thanks and appreciation are extended to Professor Max Stephens for his feedback on various drafts of this paper. This research was funded by Grant-in-Aid for Scientific Research in Japan, No. 18830014.

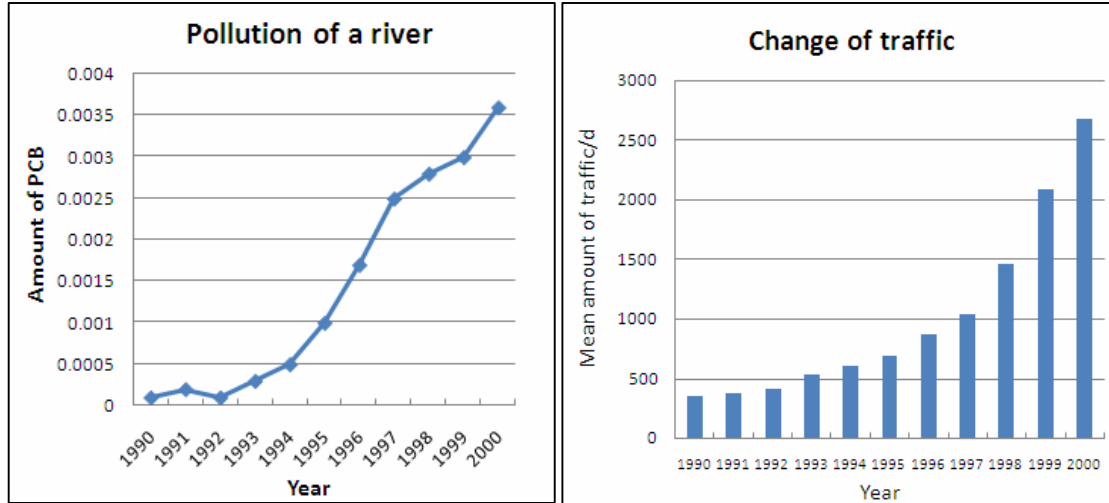
REFERENCES

- Adams, R. J., & Khoo, S. T. (1996). *Quest: Interactive item analysis system. Version 2.1* [Computer software]. Melbourne, Australia: Australian Council for Educational Research.
- Aoyama, K.m & Stephens, M. (2003). Graph interpretation aspects of statistical literacy: A Japanese perspective, *Mathematics Education Research Journal* 15(3), 3-22.
- Biggs, J., & Collis, K. (1982). *Evaluating the quality of learning: The SOLO taxonomy*. New York: Academic Press.
- Biggs, J., & Collis, K. (1991). Multimodal learning and the quality of intelligent behavior. In H. Rowe (Eds.), *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.
- Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch model. Fundamental measurement in the human sciences*. Mahwah, NJ: Lawrence Erlbaum.
- 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.
- Friel, S., Curcio, F., & Bright, G. (2001). Making sense of graphs: Critical factors influencing comprehension and instructional implications. *Journal for Research in Mathematics Education*, 32(2), 124-158.
- Gal, I., & Garfield, J. (1997). Curricular goals and assessment challenges in statistics education. In I. Gal, & 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 & Company.
- Kimura, S. (1999). Toukeizyohoukyouikuno kariyuramoto 5-dankaino toukeitekikankyu purosusu (Curriculum of statistics education and five phases of statistical inquiry process). In Zentouken (Ed.), *Toukeizyohoukyouikuno Rironto Zyugyouzissenno Tenkai* (pp. 33-46). Tsukuba: Syuppankai.
- Kodera, T., & Shimizu, Y. (2007). *Sekaiwo hiraku suugakuteki riterasi* (Mathematical literacy to open the world). Akashi: Syoten.
- Ministry of Education. (1998). *Tyuugakkou gakusyusidouyouryou* (The national course of study of elementary schools). Okurasyou: Insatsukyoku.

- Monteiro, C., & Ainley, J. (2006). Student teachers interpreting media graphs. In A. Rossman & B. Chance (Eds.), *Proceedings of the Seventh International Conference on Teaching Statistics, Salvador, Brazil*: International Statistical Institute and International Association for Statistical Education. Online: <http://www.stat.auckland.ac.nz/~iase>.
- National Council of Teachers of Mathematics (2000). *Principles and standards for school mathematics*. Reston, VA: NCTM.
- Organisation of Economic Cooperation and Development (1999). *Measuring student knowledge and skills -A new framework for assessment*. Paris: OECD.
- Organisation of Economic Cooperation and Development (2004). *Learning for tomorrow's world -First results from PISA 2003*. Paris: OECD.
- Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests*, The University of Chicago Press.
- Steen, L. (1990). *On the shoulders of giants: New approaches to numeracy*, National Academy Press.
- Watson, J. (1997). Assessing statistical literacy using the media. In I. Gal, & J. Garfield (Eds.), *The assessment challenge in statistics education* (pp. 107-121). Amsterdam: IOS Press and The International Statistical Institute.
- Watson, J., & Moritz, J. (1999). The beginning of statistical inference: Comparing two data sets. *Educational Studies in Mathematics* 37, 145-168.
- Watson, J., & Callingham, R. (2003). Statistical literacy: A complex hierarchical construct, *Statistics Education Research Journal*, 2(2), 3-46.
- Watson, J. M., & Moritz, J. B. (2000). Developing concepts of sampling. *Journal for Research in Mathematics Education*, 31 (1), 44-70.

APPENDIX

Item 1 (URB). The following graphs are the results of a research on “Changes in a river’s pollution in town A from 1990 to 2000” and “Changes in traffic in town A from 1990 to 2000”. PCB is an indicator of the pollution level?



1.1. How has the state of this river’s pollution changed from 1990 to 2000? Please choose the most suitable option and explain your response

- a: The river has been polluted
- b: The river has been cleaned
- c: There is no change
- d: I don’t know

Code 2: a
Code 1: a (with wrong reasoning), b, c, d
Code 0: No response

1.2. How has the traffic flow changed from 1990 to 2000? Please choose the most suitable option and explain your response.

- a: It has increased
- b: It has decreased
- c: There is no change
- d: I don’t know

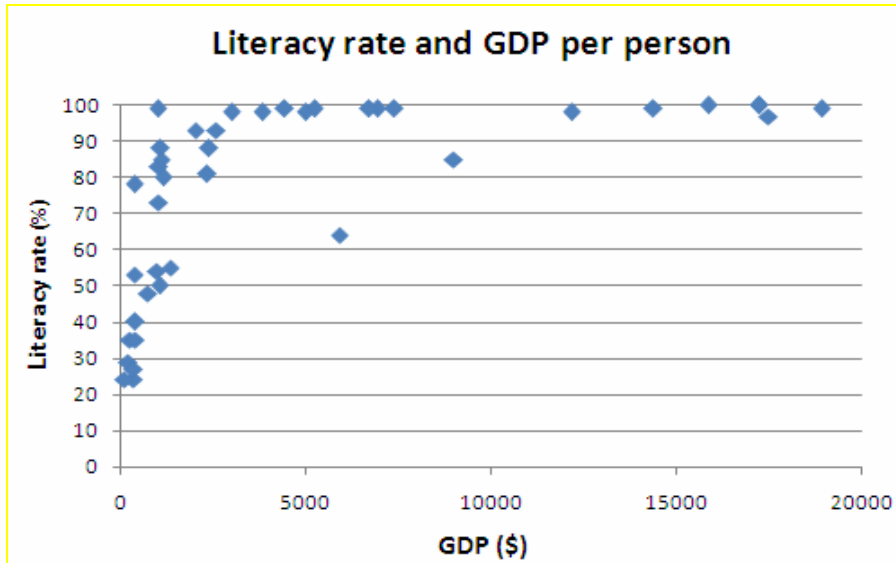
Code 2: a
Code 1: a (with wrong reasoning), b, c, d
Code 0: No response

1.3. Some people concluded from these two graphs that Town A has urbanised from 1990 to 2000. Do you agree with this opinion? Or disagree? Please explain your response

- a: Agree
- b: Disagree
- c: Neither
- d: I don’t know

Code 3: b, c, With recognition of the possibility of urbanization but referring to other possibilities
Code 2: a, With giving a reasonable explanation of urbanization
Code 1: a, b, c, d, With wrong reasoning
Code 0: No response

Item 2 (GDP). The following graph shows the literacy rate and GDP(\$) per person for 40 countries that were chosen at random.



2.1. How would you describe the literacy rates of GDP higher countries? Please explain your answer.

- a: High
- b: Low
- c: It is not clear
- d: I don't know

Code 2: a
Code 1: b, c
Code 0: d, No response

2.2. Do you think there is a relationship between GDP per person and literacy rate? If so, what type of relationship do you find? Please explain your answer.

- a: There is a relationship
- b: There is no relationship
- c: Not so clear
- d: I don't know

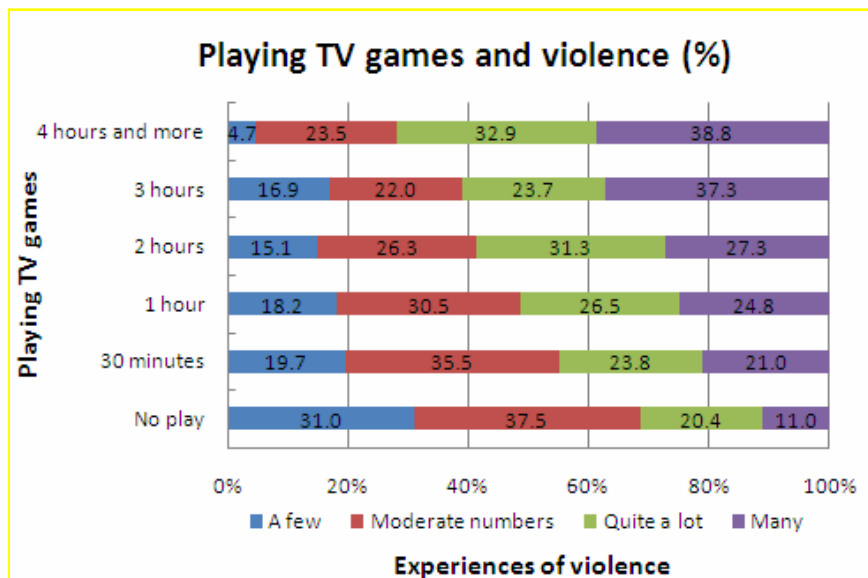
Code 3: a, With giving reasonable explanation for the distribution in graph
Code 2: a, With just referring correlation but not enough to explain the distribution
Code 1: a, b, c, With inadequate explanation
Code 0: d, No response

2.3. There are some countries with high literacy rate and low GDP per person. What do you think might be the reason for such phenomena? Please explain your answer.

- a: The data are dubious, or biased.
- b: It is not so strange because those two variables are not related.
- c: I think there are some reasons.
- d: I don't know.

Code 3: c, With giving a reasonable variable to explain
Code 2: c, With inadequate explanation? Just referring to the possibility of other variables
Code 1: a, b, c, With wrong reasoning
Code 0: d, No response

Item 3 (VLN). The following graph is the result of research that investigates how many hours per day elementary students play TV games at home and how many experiences of violence (e.g. hitting or pushing a classmate, or pulling someone's hair) they have.



3.1. What is the percentage of people experiencing “Quite a lot” of violence 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: Neither
- d: I don't know

Code 2: b
Code 1: a, c
Code 0: d, No response

3.3. As the hours of playing a TV game increase, is the proportion of “many” also increasing?

- a: Increasing
- b: Decreasing
- c: Neither
- d: I don't know

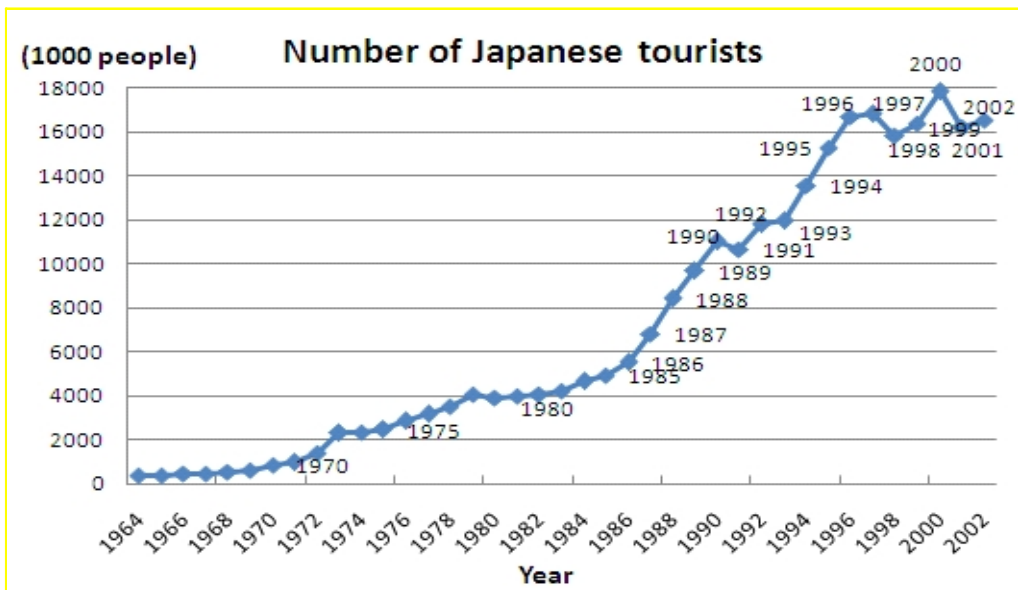
Code 2: a
Code 1: b, c
Code 0: d, No response

3.4. It is suggested that a long time spent playing TV games will cause violence. Do you agree with this opinion?

- a: Agree
- b: Disagree
- c: Neither
- d: I don't know

Code 4: b, c, With recognition of correlation but not to conclude with causality, referring other possible variables
Code 3: b, c With recognition of correlation but not to conclude with causality, not to refer any other possible variables
Code 2: a, With recognition of correlation, and to conclude with causality
Code 1: a, b, c, With wrong reasoning
Code 0: d, No response

Item 4 (TRV). The following graph shows the number of Japanese people travelling abroad from year 1964 to 2002.



4.1. How has the number of Japanese travellers changed over this period?

a: Increasing

b: Decreasing

c: Neither

d: I don't know

Code 2: a, b, With reasonable explanation
Code 1: b, c, With wrong reasoning
Code 0: d, No response

4.2. Do you find that some points are strange? How many points and which year?

a: Yes (Number: Year:)

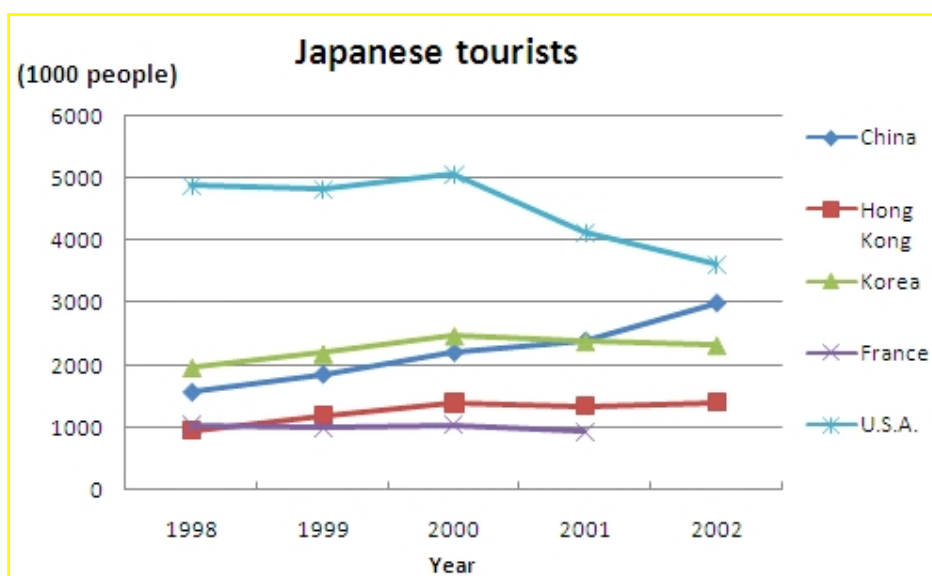
b: No

c: I don't know

Code 2: a, Referring to some appropriate turning points
Code 1: b, c, Referring to some wrong points
Code 0: c, No response

4.3. There are some decreasing points. Can you explain why?

The following graph shows the major destinations from Japan from year 1998 to 2002.



4.4. For which countries has the number of tourists changed dramatically? Do you have any explanation such change?

Code 3: Referring both of America and China with reasonable explanations for each
 Code 2: Referring either of America or China with reasonable explanation
 Code 1: Wrong reasoning
 Code 0: No response

4.5. In the previous graph, there are some decreasing points. Can you explain why?

Code 3: Referring several aspects related to this data (ex. Influence of terrorism, economic climate in Japan and so on)
 Code 2: Giving just one reason (ex. Tourists to America have been decreased)
 Code 1: Wrong reasoning
 Code 0: No response

Author : **Kazuhiro Aoyama**

E-mail : aoyamak@gmail.com

Address : INCET, Naruto University of Education

748 Nakashima Takashima Naruto-cho Naruto-shi 772-8502 Japan

Phone : +81-(0)88-687-6546

Fax : +81-(0)88-687-6546