PRE-SERVICE MATHEMATICS TEACHERS' SUBJECT MATTER KNOWLEDGE OF THE MODE AS A MEASURE OF CENTRAL TENDENCY

Komathi Jairaman
komraj1998@yahoo.com

Sharifah Norul Akmar Binti Syed Zamri
snorul@um.edu.my

Suzieleez Syrene Binti Abdul Rahim
suzieleez@um.edu.my
University of Malaya

Abstrak: Perubahan terkini dalam pendidikan statistik menekankan kepada pengembangan literasi statistik kerana dunia secara keseluruhannya pada masa ini memerlukan maklumat statistik. Walau bagaimanapun, kajian sejak beberapa dekad yang lalu telah menunjukkan bahawa literasi statistik pelajar kita tidak begitu baik. Perkara ini telah menarik perhatian guru dan peranan mereka dalam pengembangan literasi statistik. Artikel ini melaporkan pengetahuan subjek enam guru matematik pra-perkhidmatan tentang mod. Data dikumpulkan melalui teknik temu bual klinikal berasaskan soalan jenis terbuka berkaitan dengan konsep mod. Hasil kajian menunjukkan bahawa kebanyakan guru matematik pra-perkhidmatan mengenal pasti mod berdasarkan nombor terbesar yang juga dipercayai sebagai menunjukkan kekerapan tertinggi. Mereka juga mempamerkan kekurangan dari segi pengetahuan bahawa mod boleh menjadi satu ukuran biasa dalam sesuatu set data. Ciri mod ini boleh digunakan sebagai satu kaedah pantas dalam pelaporan purata yang merupakan satu idea yang penting berkaitan dengan literasi statistik.

Kata Kunci: *pengetahuan subjek, literasi statistik, mod*

INTRODUCTION

The development of statistical literacy starts in schools. Teachers are responsible for incorporating enhancement of statistical literacy in their statistics classes and preparing their students to enter society. Watson (2006) firmly believed that teachers should offer their students productive experiences using real-world examples that demonstrate the utility of statistical concepts in real life situations. Thus, incorporating enhancement of statistical literacy in their teaching of statistics. However, certain concerns were brought to highlight when research found that students were unable to relate statistics effectively in their daily lives (Garfield & Ben-Zvi, 2007).

This drew the attention of teachers with a big question on whether our teachers have the knowledge to enhance statistical literacy in their teaching. Teachers need more than adequate knowledge of mathematics. "Teachers must have an in-depth knowledge of mathematics they are going to teach" (Wun, 2010, p. 68). This will ensure that they can create meaningful teaching experiences for students. This point has been emphasized and elaborated by The National Council of Teachers of Mathematics (NCTM) (2000) which documented that:

> *Students learn mathematics through the experiences that teachers provide. Thus, students' understanding of mathematics, their ability to use it to solve problems, and their confidence in, and dispositions toward, mathematics are all shaped by the teaching they encountered in school. The implementation of mathematics education for all students requires effective mathematics teaching in all classrooms. To be effective, teachers must know and understand the mathematics they are teaching and able to draw on that knowledge flexibly in their teaching tasks*. (pp. 16-17)

While teachers learn more about the subject matter when they come into teaching, they also require a good background about their subject prior to their induction to teaching (Grossman, Wilson, & Shulman, 1989). Therefore, teachers need a solid foundation to start off and build upon their subject matter competence as they gain exposure and

experience in their profession. Similar applies to the subject matter knowledge involved in enhancing statistical literacy. Beginning teachers need to have a solid foundation in the subject matter knowledge involved in enhancing statistical literacy.

Several topics in statistics constitute the subject matter knowledge involved in enhancement of statistical literacy such as inference, variation, probability and measures of central tendency. However, among these topics, measures of central tendency occupy a fundamental portion in most mathematics curricula including in Malaysia. In relation to statistical literacy, measures of central tendency are also found to be quite dominant because of the frequent appearances of these measures in the media (Jacobbe & Carvalho, 2011). This necessitates pre-service mathematics teachers to have the necessary subject matter knowledge of measures of central tendency involved in enhancing statistical literacy.

Subject matter knowledge

Grossman et al. (1989) mentioned that four dimensions are involved in the subject matter knowledge of beginning teachers, namely content knowledge, substantive knowledge, syntactic knowledge, and beliefs. A teacher's understanding of the subject matter should not only cover knowing "that" but should also cover the knowing "why" (Shulman, 1986). This definition was further widened by Ball, Hill, and Bass (2005) as "mathematical knowledge for teaching" which focused not only on the content that teachers need to know for teaching but also the need to understand the ways of knowing that content for teaching. Therefore, these definitions of the subject matter knowledge helped to conceptualize that in the present study, the subject matter knowledge is defined as knowledge of the content and organization of the topic of measures of central tendency involved in enhancing statistical literacy.

Statistical literacy

On the other hand, *Statistical literacy* carries many interpretations too. It was interpreted as the ability to understand and evaluate critically statistical information in daily life along with the ability to appreciate the contributions of statistical thinking in various life aspects (Wallman, 1993). Gal (2002) elaborated on this interpretation and proposed two components for the requirements of statistical literacy; the first component is people's ability to interpret and critically evaluate statistical information and the second is the ability to discuss or communicate their reactions to such statistical information (Gal, 2002, p. 3). On the other hand, Watson (2006) interpreted statistical literacy as the meeting point between the statistics curriculum and real life. Meanwhile Garfield and Ben-Zvi (2004) have viewed it as a set of skills acquired from statistics classes. However, in the present study, statistical literacy is defined as the ability to relate central tendency concepts to real life situations.

The investigation of the subject matter knowledge of measures of central tendency in the present study begins with Gal's model of statistical literacy which illustrates the subject matter knowledge components involved in enhancing statistical literacy. The model draws upon the notion that the subject matter knowledge for the enhancement of statistical literacy includes two major components which are the knowledge component and the dispositional component. The knowledge component consists of five cognitive elements, namely literacy skills, statistical knowledge, mathematical knowledge, context knowledge, and critical questions whereas the dispositional component consists of two elements namely, critical stance, and belief and attitudes.

Nevertheless, in this paper, the discussion of the investigation of the subject matter knowledge of measures of central tendency is focused on Gal's statistical knowledge of measures of central tendency which includes measures of central tendency and data, familiarity with basic terms and ideas related to measures of central tendency and conclusions reached based on the measures of central tendency (Gal, 2002, p. 10).

The investigation was carried out on four different constructs, namely measures of central tendency with reference to context, measures of central tendency in handling bias, measures of central tendency and problem solving, and measures of central tendency in making inference. Watson (2006) highlighted that these constructs involving the measures of central tendency are important for enhancing statistical literacy. The following diagram illustrates the conceptual framework of the investigation discussed in this article:

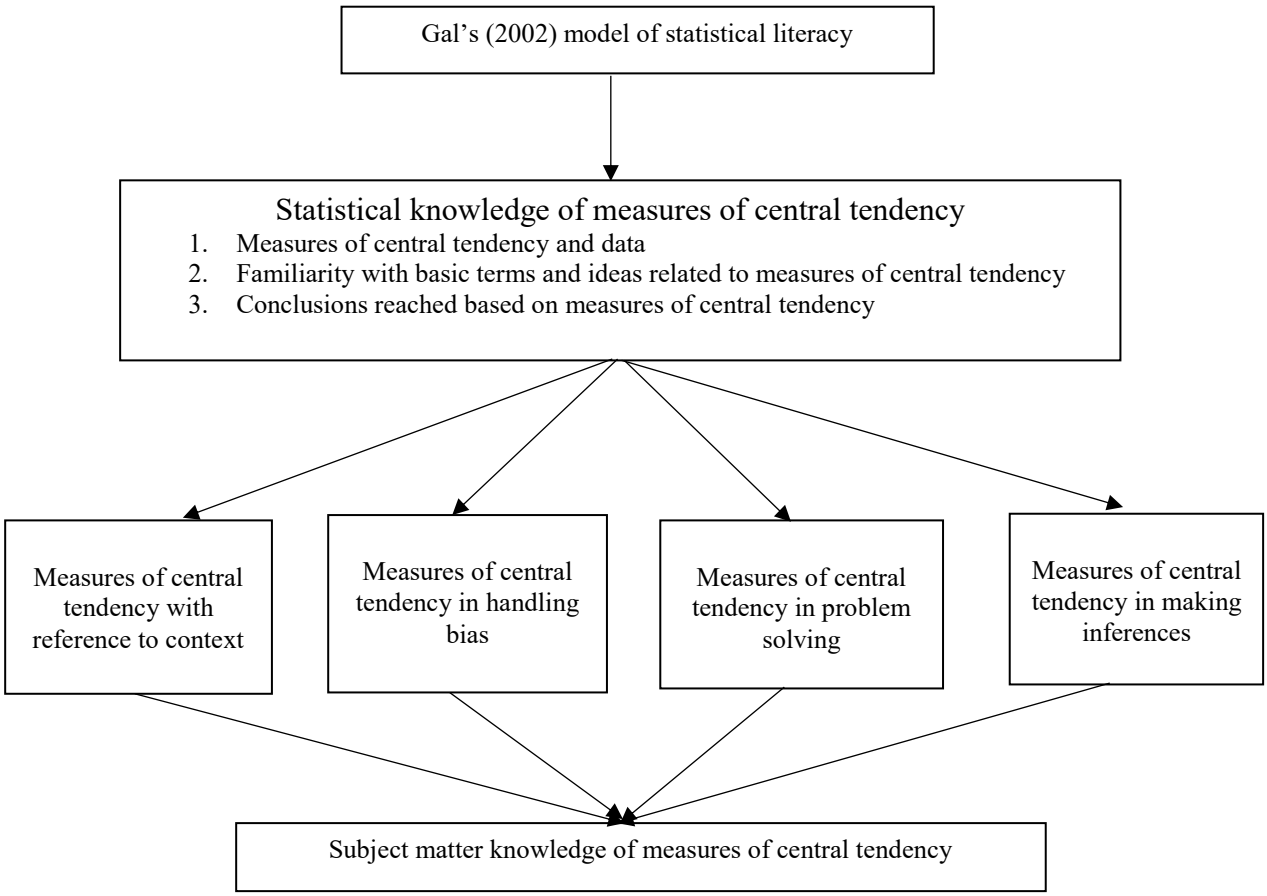

Figure 1. Conceptual framework of the study.

Therefore, the investigation sought to answer the central question - What types of subject matter knowledge of measures of central tendency do pre-service mathematics teachers have?

However, this article reports only on one part of the larger study which is focused on the concept of the mode. The mode has not drawn much attention in the literature as compared to the other two measures of central tendency, the mean and the median. One reason for the mode having a less prominent position in the mathematics curriculum may be the perception that the procedure involved in obtaining the mode is far less complicated than the procedures involved in obtaining the mean or the median (Groth & Bergner, 2006).

The mode is referred to the most frequent value in a data set and is obtained based on the highest frequency. One important characteristic of the mode that distinguishes this measure from the other measures of central tendency is that the mode is not necessarily numerical all the time and can be categorical too. Another characteristic of the mode is that it is possible to have more than one mode for a set of data. This is unlike for the mean or median. The mode is also a measure of what is typical in a data set. This characteristic of the mode is sometimes used as a quick method in the reporting of central tendency or an average, an important idea related to statistical literacy.

METHODOLOGY

A case study research design was employed to investigate and understand, in-depth, pre-service mathematics teachers' subject matter knowledge of measures of central tendency. "A case study design is used to gain an in-depth understanding of the situation and meaning for those involved" (Merriam, 1998, p. 19). Several researchers (e.g., Chew, 2007; Sharifah Norul Akmar, 1997; Wun, 2010) have employed the case study design to study Malaysian students, pre-service teachers, and lecturers.

Purposeful sampling was used to select the six subjects for this study. The subjects were pre-service mathematics teachers enrolled in a 4-year Bachelor of Science with Education (B. Sc. Ed.) program from a public university in Peninsular Malaysia. The subjects either majored or minored in mathematics. At the point of data collection, they were in their final semester of their program. The researchers had selected three B.Sc.Ed. program students who majored in mathematics, and three B.Sc.Ed. program students who minored in mathematics for the purpose of this study. The subjects have no teaching experience prior to this study. Each of the pre-service mathematics teachers was given a pseudonym, namely Lina, Bella, Harry, Amy, Anna, and Joan in order to protect the anonymity of all the interviewees. The following Table 1 illustrates the brief background information about the subjects.

Table 1
Subjects' Gender, Age, Major, and CGPA

Subject	Gender	Age	Major	Minor	CGPA
Lina	Female	24 Years 1 Month	Mathematics	Chemistry	3.62
Bella	Female	23 Years 4 Months	Mathematics	Chemistry	2.98
Harry	Male	24 Years 5 Months	Mathematics	Chemistry	3.18
Amy	Female	24 Years 2 Months	Biology	Mathematics	3.46
Anna	Female	23 Years 0 Month	Physics	Mathematics	2.93
Joan	Female	24 Years 0 Month	Chemistry	Mathematics	2.98

A total of fifteen tasks were devised for the larger study. The content validity of the tasks was achieved using experts' reviews on the tasks. The panel of experts consisted of three university lecturers from both public and private universities who specialized in mathematics and statistics education. The panel of experts was asked to judge the relevance and the content coverage of the tasks that represented the subject matter knowledge of measures of central tendency.

However, this paper will focus on the reports on the responses of the subjects on one particular task, Task 3 (given in the Appendix) related to the mode used to explore the subject matter knowledge with reference to the context. Task 3 was adapted from a previous study by Konold and Garfield (1992). This task is used to explore the subjects' knowledge of identifying the mode. The task is also used to ascertain if the subjects identify the mode based on the highest frequency and if so how they justify their answers. In addition, this task is also used to explore the subjects' knowledge of the mode in representing a typical behavior of a set of data in which this characteristic of the mode can be used as a quick method of reporting any central tendency or average. The value judgement assigned by the three experts for Task 3 were either 4 or 5. This indicated that the task was within the range of quite to most relevant and comprehensive in terms of its content.

Data were collected using clinical interview techniques based on open-ended questions as instrument. The interviews were conducted in the Discussion Room in the Main Library of the university. The physical setting was arranged with a table and two chairs. The interviews were recorded using audio voice recorder and a digital video camera. Blank papers, pencils and a calculator were accessible to all the subjects throughout the interviews. Materials collected for the data analysis consisted of the audio recordings, video recordings, subjects' notes in the provided task sheets, and the researcher's interview notes.

The data analysis process consisted of three levels. In the first level, both the audio and video recordings of the clinical interviews were verbatim transcribed into a written form. The transcription included the researcher-subject interaction during the interviews as well as the subjects' nonverbal behaviors. At the second level, the raw data in the form of transcription were coded, categorized, and analyzed according to the specific themes to produce protocol

related to the description of the subjects' subject matter knowledge of the mode. At the coding stage, codes were developed using both predefined codes and the ones emerged from the data. Coding rubrics for determining the subjects' subject matter knowledge of the mode were devised to facilitate the coding, categorization, and analysis processes. Analytic memos and analytic meetings among the researchers were done to reflect on the initial ideas, to examine these ideas in relation to new information and further pose ideas about developing relationships and patterns. Finally, at the third level, a case study for the subjects was constructed based on the information from the written protocol where the analysis was carried out to describe the subjects' task solving behaviors.

In order to ensure the credibility of the findings, the researchers triangulated the data with the subjects' notes in the task sheets and the researcher notes. Respondent validation involved the checking of the preliminary analysis with the subjects so that what is "said" matched the "intended" during the interviews and the researcher's interpretation "ring true". Debriefing sessions among the researchers to discuss the approaches, the interpretations and recognize the biases in ensuring reliability of findings were also carried out.

FINDINGS

Findings of the study were presented in terms of the pre-service mathematics teachers' knowledge of the mode.

Mode Based on the Highest Frequency

Findings of the study suggest that five out of six pre-service mathematics teachers have actually identified the mode incorrectly. Bella, Harry, Amy, Anna and Joan have identified the mode by looking at the biggest number. These pre-service mathematics teachers knew that the mode has to be based on the highest frequency. However, they incorrectly based the highest frequency by looking at the biggest number of comments. These pre-service mathematics teachers held the knowledge that the biggest number indicates the highest frequency. On the other hand, Lina who initially identified the mode based on the biggest number, later realized her mistake and changed her response. Lina was the only pre-service mathematics teacher in this study who identified the mode based on the highest frequency and justified her answer correctly.

Whereas, Amy, Harry and Joan identified the mode as Student D. They knew that the mode can be a category and not necessarily numerical all the time. Even though they had identified the mode incorrectly by looking at the biggest number of comments made but they believed that this value is actually indicating the highest frequency. Therefore, they obtained the mode by looking at the category that carried the highest frequency (in their case it was the biggest value). However, Bella and Anna who identified the mode as 22 had lack of knowledge that the mode should be referring to the category that carries the highest frequency and not the highest frequency itself. Table 2 summarizes these findings.

Table 2
Pre-service Mathematics Teachers' Mode Identification and Their Justification

Mode	Justification		Subjects
	Appropriate	Inappropriate	
2	By looking at the number of comments, there are three students who made 2 comments, so the mode is 2.		Lina
22		Because that is the highest comment among those students. So it is 22. Because what I understand about the mode is the highest frequency so based on the table, student D gives 22 comments while the others gave at most 5. So 22 is the mode.	Bella
		The highest value from the table.	Anna

D	Straight away find the highest frequency from the data and you get the mode.	Harry
	Because it has the highest number of comments as compared to the other students.	Amy
	Because student D has the highest frequency or highest number of comments.	Joan

Mode as a Form of Data Representation

In relation to the knowledge of the mode as a form of data representative, almost all six pre-service mathematics teachers in this study lacked knowledge in this aspect. Lina mentioned that the mode that she identified can only be used as a form of data representative when all the other data values in the data set are close to the mode. The following excerpt illustrates this point:

> Lina: It can be used. (Refers to the mode)
> R: Why?
> Lina: Because the difference between the numbers of comments is not that extreme (for the case where 6 is used) compared to the previous one which is 22.
> R: What if for the case of 22?
> Lina: In this case, I would not use this value because the difference is very extreme. One is zero and the other 22.

Bella, on the other hand explained that the mode that she obtained, 22, is unsuitable for use as a form of data representative because the value is different compared to all the other values in the data set. When the researcher probed her further by substituting the value 22 with 6 and asked her if she would use the mode now to represent the data, Bella explained that she will base it on the average that she calculated. If the average calculated is of the same value as the mode, then she would use the mode. If not she would take the average as a form of data representative as shown in the following excerpt.

> Bella: I must find the average first if it is 6 than yes can. I add the total up everything 5 plus 2 plus 6…. and divide by the number of students. If I get 6 than I say yes.
> R: If you do not get 6. How then?
> Bella: I take the average. And same reason from the previous, because it is referring to number of comments so it cannot be in decimal so must round off.

Harry totally disagreed on using the mode as a form of data representative. In fact, he believed the mean is a better form of data representative. Harry explained that the mode represents only one particular student and not the whole group and he believed that the mean is a better form of data representative because this measure balances up all the other values in the data set. The following excerpt illustrates Harry's point on this:

> Harry: Because from the data that some of students never give any comments and some of the students give a lot of comments until 22. So we cannot straight away say that the mode to represent the number of comments for all the students.
> R: So would you be using the mode to represent the number of comments made by the students on that day?
> Harry: I will not use the mode but I will choose the mean.
> .
> .
> R: Why mean?
> Harry: Because to balance up the zero comments from student A and some of them who commented a lot. So we just take it as average. So the mean will represent the number of comments made by the class on that day.

> Because in mode, each number is not representing the whole class but representing a particular student. So if we take mode as to represent the number of comments of the class then it is not suitable.

Similarly, Amy also disagreed on using the mode as a form of data representative and explained that the mean is a better form of data representative because the mean involves all the other comments too as shown in the following excerpt:

> Amy: Because if to say the number is used to represent the number of comments that means you are going to represent or take account all the students in the class for the number of comments that they made. So it is better to involve everyone. So it is better to use the mean.

Whereas Anna who initially mentioned that the mode can be used to represent the data later changed her response to the mode not being a suitable form of data representative because it is based on one particular value. Anna elaborated that the measure which is suitable as a form of data representative should consider all the values in the data set and that the mean is a better measure for this as described in the following excerpt:

> Anna: Mmmmm …One student only. It is based on one student only [looked uncertain]. In that case, I think cannot.
> .
> .
> Researcher: Why?
> Anna: Because that value represents only one student. By right it should consider all students. That is why we need to consider the mean.

Finally, Joan mentioned that the mode can be used to represent the number of comments made on that day. However, her justification for using mode as a form of data representative was found to be inappropriate. Joan mentioned that the mode can be a form of data representative because it can be counted. The following excerpt illustrates on this point:

> Joan: Yes, I would. The number 22 can be used.
> R: Why?
> Joan: Because it can be counted.
> R: Can you explain further?
> Joan: It can be counted. That is all.

The following Table 3 summarizes the pre-service mathematics teachers' knowledge of the mode as a form of data representative and their justification.

Table 3
Pre-service Mathematics Teachers' Knowledge of the Mode as a Form of Data Representation and Their Justification

Mode as a form of data representation	**Justification**		**Subjects**
	Appropriate	**Inappropriate**	
Suitable		Suitable to use mode only for cases where the mode is close to the other values in the data set	Liana
		Because the mode can be counted	Joan
Not suitable		The mode is different than the other comments and would be based on the average to decide.	Bella
		Because the mode represents only one particular data and not the whole data. The mean is a better form of data representative because this measure takes into account all the data.	Amy, Anna, Harry

DISCUSSION AND CONCLUSION

In conclusion, only one out of the six pre-service mathematics teachers had the knowledge to identify mode correctly based on the most frequent or highest frequency. Three of them had the knowledge that the mode can be categorical. However, these pre-service mathematics teachers had a misconception that the biggest number of comments made on that particular day would indicate the most frequent number of comments. Two of them had limited knowledge on how to identify the mode because they have actually taken the biggest number of comments (a misconception that this value is indicating the most frequent number of comments) as the mode. These teachers lacked knowledge that the mode is not necessarily numerical all the time and can be categorical in certain situations. Similarly, most pre-service elementary teachers in Groth and Bergner (2006) also implied that the mode is only of use for numerical data sets. In relation to the pre-service mathematics teachers' knowledge of the mode as a representative of a data set from which they arise (Groth & Bergner, 2006), it can be concluded that all the six pre-service mathematics teachers had very limited knowledge. None of them have actually utilized the mode as a form of data representative with proper justification for its use. The mode can describe a typical behavior of a data set and therefore it can be used as a form of data representative. Moreover, this characteristic of the mode is used as a quick method in reporting central tendency or an average. The general term average may mask which of the three measures of central tendency is being used (Watson, 2006), therefore having a deep knowledge of these measures of central tendency, namely the mode, will help in enhancing statistical literacy.

Although this article reports on the six pre-services mathematics teachers' responses to Task 3 only but we will further investigate their attempts at other tasks related to the mode which will help us to understand their subject matter knowledge of the mode. The final findings of this study are expected to provide some substantive results not only on pre-service mathematics teachers' subject matter knowledge of the mode but also measures of central tendency. We hope that these findings can encourage a review on the current Mathematics teacher education programs at our local universities or teacher training institutes so that the fundamental ideas of central tendency concepts which are important for the enhancement of statistical literacy can be taught at a deeper level.

ACKNOWLEDGEMENT

A special thank you to the Ministry of Education, Malaysia for granting the Post-Graduate Scholarship Scheme and University of Malaya for the professional support.

REFERENCES

Chew, C. M. (2007). *Form One students' learning of solid geometry in a phase-based instructional environment using the Geometer's Sketchpad*. (Unpublished doctoral thesis, University of Malaya, Kuala Lumpur).

Gal, I. (2002). Adults' statistical literacy: Meanings, components, responsibilities. *International Statistical Review*, *70*(1), 1–25.

Garfield, J., & Ben-Zvi, D. (2004). *The challenge of developing statistical literacy, reasoning and thinking*. Dordrecht, The Netherlands: Kluwer.

Garfield, J., & Ben-Zvi, D. (2007). How students learn statistics revisited: A current review of research on teaching and learning statistics. *International Statistical Review*, *75*(3), 372-396.

Grossman, P. L., Wilson, S. M., & Shulman, L. S. (1989). Teachers of substance: Subject matter knowledge for teaching. In M. Reynolds (Ed.), *Knowledge base of beginning teachers* (pp. 23-36). Oxford, UK: Pergamon Press.

Groth, R. E., & Bergner, J. A. (2006). Pre-service elementary teachers' conceptual and procedural knowledge of mean, median, and mode. *Mathematical Thinking and Learning*, *8*, 37-63.

Jacobbe, T., & Carvalho, C. (2011). Teachers' understanding of averages. In C. Batanero, G. Burrill, C. Reading (Eds.), *Teaching Statistics in School Mathematics-Challenges for Teaching and Teacher education: A Joint ICMI/IASE Study*

Konold, C., & Garfield, J. (1992). Statistical reasoning assessment: Part 1. Intuitive thinking. Amherst, MA: University of Massachusetts, Scientific Reasoning Research Institute.

Merriam, S. B. (1998). *Qualitative research and case study applications in education.* San Francisco, CA: Jossey-Bass.

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

Sharifah Norul Akmar, S. Z. (1997). *Skim penolakan integer pelajar tingkatan dua* [*Integers subtraction schemes of form two students*]. (Unpublished doctoral thesis, University of Malaya, Kuala Lumpur).

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*(2), 4-14.

Watson, J. M. (2006). *Statistical literacy at school: Growth and goals*. Mahwah, NJ: Erlbaum.

Wun, T. H. (2010). *Pre-service secondary school mathematics teachers' subject matter knowledge of perimeter and area*. (Unpublished doctoral thesis, University of Malaya, Kuala Lumpur).

APPENDIX

Task 3

The table below shows the number of comments made by eight students during a class period on a particular day.

Student	Number of comments
A	0
B	5
C	2
D	22
E	3
F	2
G	1
H	2

a. What is the mode?
Probe: If number of comments 22 is changed to 6

b. How did you obtain this value as the mode?
Probe: If number of comments 22 is changed to 6

c. If you were to represent one value for the number of comments made by the students on that day, would you use this value?
Probe: Why?