AMERICAN SCIENTIFIC PUBLISHERS

Advanced Science Letters
Vol. 21, 2352-2355, 2015

A Refined Technology-based Statistical Reasoning Assessment Tool in Descriptive Statistics

Shiau Wei Chan[1*], Zaleha Ismail[2], Bambang Sumintono[3]
[1]Faculty of Technology Management and Business, Universiti Tun Hussein Onn Malaysia, 86400 Parit Raja, Batu Pahat, Johor, Malaysia
[2]Faculty of Education, Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia
[3]Institute of Educational Leadership, Universiti Malaya, 59990 Jalan Pantai Baru, Kuala Lumpur, Malaysia

Statistical reasoning is an essential field in statistics education, but its assessment has yet to be fully established in Malaysia, especially in terms of alternative assessments. This study has been conducted to refine an initial technology-based statistical reasoning assessment tool from Chan and Ismail's (2014) study in descriptive statistics for upper secondary school students. The assessment tool was constructed upon five levels of statistical reasoning from Garfield's (2002) model and four constructs from the framework of Jones et al. (2000). The content validity and inter-rater reliability of this assessment tool were measured. 10 upper secondary students participated in a task-based interview using this assessment tool. Amendment was made after the first task-based interview to refine the assessment tool to become more reliable and realistic in assessing the statistical reasoning ability among secondary students.

Keywords: Alternative assessments, Statistical reasoning, Secondary education

1. INTRODUCTION

Assessment plays a vital role in both teaching and learning (Lau et al., 2011). According to Muirhead (2002), the term 'assessment' refers to the collection of information concerning the achievement of students to determine their progress. Despite statistics education's gradual shift from traditional assessment to alternative assessment, the assessment of statistical reasoning is still at its infancy in Malaysia. In addition, it has been found that the elements of statistical reasoning have remained largely underexposed in the Malaysian mathematics textbooks; most of the existing problems in the textbooks are routine, traditional, and close-ended problems. Although there are some statistical reasoning assessments available in the market, they are not suitable for Malaysian secondary students because not all topics have been covered. Such holds true for topics such as causation and correlation in the Statistical Reasoning Assessment (SRA). Moreover, some assessments fall short in meeting the modern education development since the questions are still mostly true/false and multiple choice questions (Wang et al., 2012). Hence, an alternative assessment of statistical reasoning is required to gauge students' reasoning, discuss their ways of thinking, and support their judgments and conclusions (Garfield and Ben-Zvi, 2008) as well as to encourage more efficient learning and teaching practice (Lin & Lai, 2013).

With rapid development of information and Internet technologies, the integration of computers in academic assessment is becoming increasingly popular (Lau et al., 2011) to complement traditional methods. Earlier studies also proved that the usage of computer can improve the assessment process (Eyal, 2012) and develop students' statistical reasoning (Chan and Ismail, 2012). Various types of technological tools can be used in the statistics classroom such as Minitools, TinkerPlots, graphing calculator, and so on. In this study, the

*Email Address: swchan@uthm.edu.my

1936-6612/2015/21/400/008 doi: 10.1166/asl.2015.6273

GeoGebra software, a dynamic mathematics software, was employed to solve the statistical reasoning tasks. Its superior performance to other software has made it a great tool to combine the characteristics of a spreadsheet, dynamic geometry software, and computer algebra systems altogether. This software can be downloaded freely from website since it is open source software (Hohenwarter & Lavicza, 2007). In this study, the features of GeoGebra spreadsheet have been incorporated into the statistical reasoning assessment to produce a technology-based assessment. The authors' paper - Chan and Ismail (2014) has discussed about the initial development of this assessment tool. The research objective of this study is to refine the initial technology-based statistical reasoning assessment tool in the study of Chan and Ismail (2014) after the first task-based interview in order to assess upper secondary students' statistical reasoning levels across four constructs.

2. STATISTICAL REASONING

Garfield and Chance (2000) described statistical reasoning as 'the way people reason with statistical ideas and make sense of statistical information. It involves making interpretations based on sets of data or statistical summaries of data where students need to combine ideas about data, make inferences, and lastly interpret the statistical results' (pp. 101). In the context of this study, 'statistical reasoning' refers to an ability which can be examined using Garfield's (2002) five levels of statistical reasoning and Jones et al.'s (2000) four constructs. The five levels of statistical reasoning are idiosyncratic reasoning, verbal reasoning, transitional reasoning, procedural reasoning, and integrated process reasoning. On the other hand, the four constructs are describing data; organizing and reducing data; representing data; and analyzing and interpreting data. These five levels and four constructs will be further discussed in the next section.

The initial technology-based statistical reasoning assessment tool was created based on an initial statistical reasoning framework containing five levels of statistical reasoning and four constructs as mentioned in the study of Chan and Ismail (2014). Thus far, not much study has been conducted based on these five levels. Additionally, the types of participants involved and topics covered were inadequate. So, this study bridged the gap by using these five levels for secondary students. After the second task-based interview, the initial statistical reasoning framework was revised to become refined statistical reasoning framework. A student was said to have achieved the idiosyncratic reasoning level when he or she could describe data; organize and reduce data; represent data; and analyze and interpret data, but had reached a totally false conclusion. Besides, verbal reasoning is perceived as the capability to describe data; organize and reduce data; represent data; and analyze and interpret data orally, but the reactions could be incomplete or partially accurate. A student who had reached the transitional reasoning could describe data; organize and reduce data; represent data; and analyze and interpret data, but incapable of relating the ideas to actual data or graph. Furthermore, a student with procedural reasoning capability would be proficient enough at describing data; organizing and reducing data; representing data; and analyzing and interpreting data, but not provide a complete justification. Integrated process reasoning referred to students who were adept at describing data; organizing and reducing data; representing data; and analyzing and interpreting data as well as giving elucidation and relation to the actual data or graph.

On the other hand, the four constructs from Jones et al.'s (2000) framework were employed in this study because they were crucial in the data handling, but still difficult to be mastered by the students. In this study, describing data involved reading information from the data or a graph, demonstrating awareness towards the displayed features of the graphical representations, and identifying the general characteristics of the graphical representation. Moreover, organizing and reducing data involves organizing the data into a computer system as well as reducing data using measures of central tendency and measures of variability. Representing data includes presenting the data sets graphically using a computer, identifying the dissimilar representations for the same data set, and judging the efficacy of two dissimilar representations for the same data. Analyzing and interpreting data involved making comparisons within the same data set and between two different data sets other than making inferences, predictions or conclusions for the data or graph.

3. METHODOLOGY
3.1 RESEARCH DESIGN

This study adapted the framework validation processes from the earlier studies of Jones et al. (2000) and Mooney (2002). The first step was constructing the initial technology-based statistical reasoning assessment tool based on the initial statistical reasoning framework as explained in the study of Chan and Ismail (2014). The second step was interviewing the students using the initial technology-based statistical reasoning assessment tool. The third step was analyzing the students' responses to the initial technology-based statistical reasoning assessment tool. The fourth step was refining the initial statistical reasoning framework and the initial technology-based statistical reasoning assessment tool. The fifth step was interviewing the students for the second time using the refined technology-based statistical reasoning assessment tool. This paper mainly focuses on the fourth and fifth step. The results of second interview were discussed in the findings section.

3.2 RESPONDENTS

In this study, the respondents were ten upper secondary students from the same secondary school in Johor, Malaysia. The sample size was appropriate as the respondents in previous studies for framework validation (e.g. Jones et al., 2000) was between 6 to 20 students. The students took part in the task-based interview which lasted for two to three hours. Throughout the interview, the students were requested to answer five tasks in the technology-based statistical reasoning assessment tool. They sat in front of a laptop and exploited the GeoGebra software when essential. The entire interview was recorded using a camcorder. The students' participation was entirely voluntary and all identities were kept confidential due to ethnical concerns. Thus, every participant was assigned with a pseudonym (S1 to S10).

3.3 INSTRUMENTATION

The instrument in this study was a refined technology-based statistical reasoning assessment tool for assessing students' statistical reasoning levels across the four constructs during the task-based interview. The descriptive statistics topics covered were measures of central tendency, variability, and distribution. At first, it was created upon the initial statistical reasoning framework (Chan and Ismail, 2014). There were 56 items in this assessment tool with each item corresponding to the sub-processes of the four constructs. The content validity and inter-rater reliability of this assessment tool were measured. Three experts validated the contents of this instrument to ensure that the content matched the area being measured. They were lecturers from foreign universities who had contributed significantly in the statistical reasoning area. This assessment tool was amended according to the comments, recommendations, and views given by the experts as discussed in detail in study of Chan and Ismail (2014). On top of that, the inter-rater reliability obtained by calculating the percentage of agreement between two raters from a local university (Cozby & Bates, 2012) was 96.4%, which meant that the assessment tool was rationally consistent since this had surpassed the 70% suggested by Boyatzis (1998). Nonetheless, after the first task-based interview, it was noticed that the responses given by the students in the initial technology-based statistical reasoning assessment tool had some missing statistical reasoning characteristics owing to certain unsuitable questions. Hence, the assessment tool was revised and then deployed in the second task-based interview. The number of items was reduced to 51 after the amendment. Detailed information on the amendment will be discussed in the findings section.

4. FINDINGS

4.1 AMENDMENT OF ITEMS FOR DESCRIBING DATA

This study is aimed to refine the technology-based statistical reasoning assessment tool. This is because there were many items in the technology-based statistical reasoning assessment tool that were unable to evaluate students' statistical reasoning in the first task-based interview. Therefore, each of items which categorized under the four constructs were refined. For instance, the word of 'center' was changed to 'measures of central tendency' because some students confused with the word of 'center' when solving the tasks. Consequently, those items were revised to solve such problem. Besides, the phrase 'Explain your answer' was added to each item for describing data. For example, when the initial item questioned, 'What are the highest and lowest amount of protein (in grams) for various fast food sandwiches?', the students merely identified the highest and lowest values, so the researchers cannot determine their statistical reasoning. Therefore, with the phrase 'Explain your answer', the students would elucidate further, so their reasoning on how they made sense of the statistical concepts could be traced.

4.2 AMENDMENT OF ITEMS FOR ORGANIZING AND REDUCING DATA

For organizing and reducing data, the phrase 'Explain how' was inserted to the item that asked the students to organize the data into the GeoGebra spreadsheet. This was designed to ask the students to explicate on how they keyed in the data into the computer step-by-step. Furthermore, the items of O2 (reducing the data using measures of central tendency) and O3 (reducing the data using measures of spread) were removed because these items only required the students to perform practical steps, meaning their reasoning could not be detected as well. Besides, the phrase 'Explain your answer' was included in the other items to identify the students' statistical reasoning. For instance, when the item required the students to find the interquartile range, the students had to carry out the computation to obtain the answer. After that, they needed to give the explanation on how to get the answer and the meaning of interquartile range.

4.3 AMENDMENT OF ITEMS FOR REPRESENTING DATA

Majority of the items for representing data were changed by inserting the phrase 'Explain how'. By doing so, the students were required to elucidate the way they produced the graphical representation in detail. For instance, when the students were asked to create a frequency polygon using GeoGebra spreadsheet, they had to draw it first before they could explain on its construction.

4.4 AMENDMENT OF ITEMS FOR ANALYZING AND INTERPRETING DATA

The researchers moved one of the initial items of A1 (making comparisons within the same data set) to A2 (making comparisons between two different data sets), i.e., 'Compare the results in Question 15 with Question 14. What do you observe? Explain why.' This was done since they were not the same data set, but two different data sets after the two additional students were added in the data. Moreover, some items were altered by adding the phrase 'Explain why' wherever the students had to provide their reason. For example, for the question 'Are there any similarities or differences between the two graphs produced on the computer? Explain why', the students asserted the similarities and differences between the two graphs and then gave their rationale.

5. DISCUSSION

An initial technology-based statistical reasoning assessment tool was constructed with its content validity and inter-rater reliability measured in Chan and Ismail's (2014) study. After that, it was revised after the task-based interview. Such had increased the credibility or trustworthiness of this assessment tool for assessing secondary students' statistical reasoning.

Although mathematical reasoning is within the Malaysian secondary mathematics curriculum, it is not being widely practiced and has failed to develop the reasoning ability among the students. As we can observe from the TIMSS results, the average score of Malaysian students for the items of reasoning had decreased greatly from year 1999 to 2011. This is due to the fact that most of the questions in Malaysian mathematics and additional mathematics textbooks have focused too intensely on calculation and procedural skills. Hence, statistical reasoning ought to be integrated into the assessments in order to enhance students' statistical reasoning. This problem is overcome by developing a statistical reasoning assessment for secondary students.

The newly constructed technology-based statistical reasoning assessment tool is unique and different from any textbook and other traditional assessments. First of all, this assessment tool is more advanced than other traditional assessments since it involves the usage of a technological tool - the GeoGebra spreadsheet. The features of information technology are not commonly found in a textbook; the students have to draw the graphical representation manually. Secondly, the questions used in this assessment tool are non-routine, non-traditional, and are open-ended, which can help to augment statistical reasoning. On the contrary, in traditional assessment, the students are only required to compute the mean, mode, and median. In this assessment tool, however, the students have to explain on their computation as well as the meaning of mean, mode and median other than finding their values. Thirdly, this assessment tool has integrated three statistical reasoning altogether, which are the measures of central tendency, variability, and distribution. These three statistical reasoning are generally taken as three different topics and are taught separately in a statistics classroom. As such, many students have failed to see their interconnections (Garfield & Ben-Zvi, 2007), but this can be overcome by combining them together in a meaningful manner. Fourthly, three out of the five tasks in this assessment tool contained real data. In the textbook, most of the data are usually contrived. By using real data, the students are able to relate the concepts more realistically to their daily lives, and thus are more driven to actively interact in the activities where they investigate, and analyze the data; consequently, this deepens their understanding on statistical ideas (Ben-Zvi, 2004). Finally, when students are introduced to new graphical representations like box plot and stem and leaf plot in a more comprehensive manner, the students will become more aware of different ways to meaningfully present their data graphically.

In future studies, the research community can use the refined technology-based statistical reasoning assessment tool for different grade levels to compare and contrast their statistical reasoning ability. Also, it can be applied to different schools, gender, country, and cultural and educational background. Nonetheless, the topics in this assessment tool ought to be expanded to inferential statistics, probability, and other relevant topics. Besides, instructors ought to be given ongoing professional development training on how to employ the GeoGebra software. This is to ensure that they are well-prepared and well-equipped before teaching their students.

6. CONCLUSION

To conclude, an initial technology-based statistical reasoning assessment tool in descriptive statistics was refined. The refined technology-based statistical reasoning assessment tool in this study has contributed notably to the statistical reasoning area. Future investigations should be continuously carried out in order to improve the statistical reasoning framework along with this assessment tool.

REFERENCES

Ben-Zvi, D. (2004). *Reasoning about data analysis*. In D. Ben-Zvi & J. Garfield (Eds.), The Challenge of Developing Statistical Literacy, Reasoning, and Thinking (pp. 121-145). Springer.

Boyatzis, R. E. (1998). *Transforming qualitative information: Thematic analysis and code development*. London: SAGE Publication.

Chan, S. W., & Ismail, Z. (2012). The role of information technology in developing students' statistical seasoning. *Procedia-Social and Behavioral Sciences*, 46, 3660-3664.

Chan, S. W., & Ismail, Z. (2014). A technology-based statistical reasoning assessment tool in descriptive statistics for secondary school students. *The Turkish Online Journal of Educational Technology*, 13(1), 29-46.

Cozby, P. C., & Bates, S. C. (2012). *Methods in behavioral research* (7^{th} Ed.). New York: McGraw-Hill.
delMas, R. C. (2004). *A comparison of mathematical and statistical reasoning*. In D. Ben-Zvi & J. Garfield (Eds.), The Challenge of Developing Statistical Literacy, Reasoning, and Thinking (pp. 79-95). Springer.
Eyal, L. (2012). Digital assessment literacy – the core role of the teacher in a digital environment. *Educational Technology & Society*, 15(2), 37-49.
Garfield, J. (2002). The challenge of developing statistical reasoning. *Journal of Statistics Education*, 10 (3).
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.
Garfield, J., & Chance, B. (2000). Assessment in statistics education: Issues and challenges. *Mathematical Thinking and Learning*, 2(1and2), 99-125.
Garfield, J. B., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning: Connecting research and teaching practice*. Springer.
Hohenwarter, M., and Lavicza, Z. (2007). Mathematics teacher development with ICT: towards an International GeoGebra Institute. In D. Küchemann (Ed.), *Proceedings of the British Society for Research into Learning Mathematics. 27*(3). University of Northampton, UK: BSRLM.
Jones, G. A., Thornton, C. A., Langrall, C. W., Mooney, E. S., Perry, B., & Putt, I. A. (2000). A framework for characterizing children's statistical thinking. *Mathematical Thinking and Learning*, 2, 269-307.
Lau, P.N.K., Lau, S.H., Hong, K.S., & Usop, H. (2011). Guessing, partial knowledge, and misconceptions in multiple-choice tests. *Educational Technology & Society*, 14(4), 99-110.
Lin, J-W., & Lai, Y-C. (2013). Harnessing collaborative annotations on online formative assessments. *Educational Technology & Society*, 16(1), 263-274.
Mooney, E. S. (2002). A framework for characterizing middle school students' statistical thinking. *Mathematical Thinking and Learning*, 4(1), 23-63.
Moore, D. S. (1997). *Statistics: Concepts and controversies* (4^{th} Ed.). New York: Freeman.
Muirhead, B. (2002). Effective online assessment strategies for today's college and universities. *Educational Technology & Society*, 5(4), 1-7.
Rossman, A. J., Chance, B. L., & Medina, E. (2006). *Some important comparisons between statistics and mathematics, and why teachers should Care*. In G. F. Burrill (Ed.) Thinking and Reasoning about Data and Chance: Sixty-eighth NCTM Yearbook (pp. 323-334), National Council of Teachers of Mathematics.
Wang, Y., Li, H., Feng, Y., Jiang, Y., & Liu, Y. (2012). Assessment of programming language learning based on peer code review model: Implementation and experience report. *Computers & Education*, 59, 412–422.

Received: 22 September 2010. Accepted: 18 October 2010