Assessing Statistical Reasoning in Descriptive Statistics: A Qualitative Meta-analysis

Shiau Wei Chan*, Zaleha Ismail, Bambang Sumintono

Department of Production and Operation Management, Faculty of Technology Management and Business, Universiti Tun Hussein Onn Malaysia, 86400 Parit Raja, Batu Pahat, Johor, Malaysia
Department of Educational Mathematics, Science and Multimedia Creative, Faculty of Education, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia
Institute of Educational Leadership, Deputy Vice Chancellor(Academic & International), University of Malaya, 50603 Kuala Lumpur, Malaysia

*Corresponding author: swchan@uthm.edu.my

Article history
Received :15 June 2014
Received in revised form : 15 September 2014
Accepted :15 October 2014

Graphical abstract

Abstract
To date, there are abundant studies on statistical reasoning in descriptive statistics and inferential statistics. Nevertheless, the types of statistical reasoning assessments used in those studies are different from each other. Hence, this qualitative meta-analysis is aimed to explore the methods utilized in assessing statistical reasoning among students from all levels in descriptive statistics. A total of 36 studies on reasoning about measures of central tendency, variability and distribution were found and reviewed in this paper. It was noticed that six major types of methods were employed to assess students’ statistical reasoning in descriptive statistics, namely interview, survey or questionnaire, tasks, tests, minute paper, and teaching. This study contributes considerably to the statistical reasoning area as it provides new information on statistical reasoning in descriptive statistics. For future studies, some recommendations are proposed to improve statistical reasoning assessments.

Keywords: statistical reasoning; descriptive statistics; qualitative meta-analysis

© 2015 Penerbit UTM Press. All rights reserved.

1.0 INTRODUCTION

Educational assessment is a tool and a way of managing the educational practice, besides serving as a response and information about correct or incorrect learning methods [1]. Pellegrino, Chudowsky and Glaser [2] affirmed the three intentions of the assessment, which are to determine individual achievement, evaluate programs and support student learning. There are two types of educational assessments, i.e. formative assessment and summative assessment. Formative assessment is a planned process that regularly determines students’ understanding in the instructional activities [3]. Meanwhile, summative assessment is a cumulative assessment that may generate an ultimate grade at the end of the course [4]. In statistics education, there are several types of assessments accessible in the market, for instance the statistical reasoning assessment (SRA), Comprehensive of Assessment of Outcomes in a First Statistics Course (CAOS), Assessment Resource Tools for Improving Statistical Thinking (ARTIST), and so on. However, the instructors tend to use traditional assessments in the statistics classroom instead of alternative assessments, which are incapable of guiding students to reason statistically [5].

Since the 1990s, a significant move has occurred from procedural understanding towards engendering conceptual understanding in the statistics education. It means that the researchers and instructors began to guide students to reason statistically rather than focusing only on calculation, procedures and skills. In fact, statistical reasoning has become known from the 20th century. In the 1970s, researchers gave emphasis to the growth and testing on cognitive science theories to elucidate the misconceptions in statistical reasoning. Nevertheless, those cognitive science theories were not employed to improve the teaching and learning until the 1980s. After that, those empirical works were implemented to investigate the statistical reasoning of the students in the classroom. Beginning from the 1990s, the content of textbooks was altered to emphasize more on conceptual understanding rather than procedural understanding. Moreover, the teaching approaches had been transformed to foster students’ statistical reasoning, for instance through simulations and hands-on activities [6]. However, these transformations could not be achieved without the support from previous researches in statistical reasoning.

To date, numerous earlier studies have been conducted on statistical reasoning in descriptive statistics and inferential statistics. Different types of approaches were utilized to assess
the statistical reasoning of the students from primary, secondary, and tertiary level. Thus, the qualitative meta-analysis of this study is to explore the methods used in assessing statistical reasoning among students from all levels, particularly in descriptive statistics.

2.0 ASSESSING STATISTICAL REASONING IN DESCRIPTIVE STATISTICS

Statistical reasoning is defined 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 and have a chance to make inferences and interpret statistical results” [7]. Lovett [6] asserted that statistical reasoning involves the utilization of statistical concepts and tools to recapitulate the situation, draw conjectures and make conclusions from the data. Furthermore, Martin [8] characterized statistical reasoning as formulating judgments and conclusions based on the data from sample surveys, observational studies, or experiments.

Descriptive statistics include measures of central tendency, variability and distribution. Measures of central tendency are the main component in conjecturing data analysis and graphs as well as in comprehending the idea of distribution [9]. It comprises of mean, median, and mode. Mean is the total sum of observation divided by the overall observations. Meanwhile, the median is the middle value of a set of data and the mode is the highest frequency. Some statisticians deemed the average as a measure of central tendency, which comprises of mean, median and mode [10]. Nevertheless, Konold and Pollakstek [11,12] disputed that the term “average” has dissimilar interpretations based on the context of the problem and it could be examined as either fair share [13], data reduction [13], typical value [14], or signal in noise [14]. Mokros and Russell [15] described ‘average’ as a way to elucidate and summarize as well as to compare data sets. In mathematics curriculum, ‘average’ is viewed as a synonym for arithmetic mean [16].

On the other hand, ‘dispersion’ and ‘spread’ are the synonyms of variability. It includes range, variance, standard deviation and interquartile range. The square root of the variance is the standard deviation. The range is the subtraction of the highest value with the lowest value while the interquartile range is the subtraction of the third quartile with the first quartile. ‘Variability’ and ‘variation’ can be utilized interchangeably, but Reading and Shaughnessy [17] judged them in a different way, where variability is the apparent attribute of the entity and variation concerns demonstrating or assessing that attribute. Distribution is always associated to the attribute of the entity and variation concerns demonstrating or representing a set of data logically or not. It is also about knowing a good summary of data can make the comparison of the measures of central tendency and variability easier.

Furthermore, distribution is perceived as one of the primary and essential ‘big ideas’ in statistics [9]. Distribution can be classified into two major types, i.e., theoretical distribution and empirical distribution [19]. Theoretical distribution entails differentiating or showing a probability model including normal distribution while empirical distribution allows us to observe the variation in the data directly. Measures of central tendency, shape and spread are the general characteristics of such distribution [9]. Reasoning about distribution is defined as the analysis of compound structure including features such as measures of central tendency, spread, skewness, outliers, and density [21] as well as ideas like sampling, causality, chance, and population [22]. There are numerous methods to signify the distribution of data sets. For example, a dot plot or histogram can be employed to portray the shape of a data set, while a box plot is better utilized to demonstrate an outlier and a stem-and-leaf plot can be used to illustrate the clumps or gaps in the distribution [9]. The exploitation of graphical representations is a proficient way to enhance students’ conceptual understanding of distribution [23].

3.0 METHODOLOGY

This study intends to examine the methods utilized in assessing statistical reasoning among the students in descriptive statistics. Therefore, a qualitative meta-analysis was performed by using the literature search process of Blumberg, Cooper & Schindler [24] which are: (i) build up a pool of potential information using various databases including Google Scholar, ProQuest, Web of Science, ERIC and Science Direct; (ii) use filter to diminish pool size, such as focusing on peer-reviewed publications; (iii) make a rough assessment of sources to further diminish pool size, for example classifying studies into crucial, probably crucial and not crucial; (iv) analyze literature in pool according to theories, respondents, instruments, methods, and the findings; and (v) refine filters (try new search terms) or stop search. In this study, the inclusion criteria of the studies that were utilized are: (a) content relevancy – the studies on assessing statistical reasoning in descriptive statistics including measures of central tendency, variability and distribution; (b) Year of publication – 1988 to 2012; (c) Language – English language. As a result, a total of 36 studies were reviewed in this study, as shown in Table 1.

Table 1: Studies of assessing statistical reasoning in descriptive statistics

<table>
<thead>
<tr>
<th>Researcher</th>
<th>Level</th>
<th>Statistical Concept</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strauss &amp; Bichler [13]</td>
<td>80 children from ages 8, 10, 12 and 14</td>
<td>Arithmetic average</td>
<td>Interview (32 tasks)</td>
</tr>
<tr>
<td>Leon &amp; Zawojewski [25]</td>
<td>145 students (42 4th graders, 61 8th graders and 42 college students)</td>
<td>Arithmetic mean</td>
<td>16-item questionnaire</td>
</tr>
<tr>
<td>Mokros &amp; Russell [15]</td>
<td>21 students (7 each in 4th, 6th and 8th grades)</td>
<td>Average</td>
<td>Interview using a series of open-ended problems</td>
</tr>
<tr>
<td>Cai [26]</td>
<td>250 6th graders</td>
<td>Averaging algorithm</td>
<td>7 tasks</td>
</tr>
<tr>
<td>Watson &amp; Moritz [27]</td>
<td>88 students from Grade 3 to 9</td>
<td>Comparing two data sets</td>
<td>Interview</td>
</tr>
<tr>
<td>Watson &amp; Moritz [28]</td>
<td>94 students from Grades 3 to 9</td>
<td>Average</td>
<td>Interview</td>
</tr>
<tr>
<td>Researcher</td>
<td>Level</td>
<td>Statistical Concept</td>
<td>Method</td>
</tr>
<tr>
<td>--------------------</td>
<td>--------------------------------</td>
<td>---------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>Pfannkuch [44]</td>
<td>1 secondary teacher</td>
<td>Comparing box plot distribution</td>
<td>Questionnaire</td>
</tr>
<tr>
<td>Reid &amp; Reading</td>
<td>32 students (pre-study) and 23 students (post-study)</td>
<td>Variability, comparing data sets, sampling and probability</td>
<td>Questionnaire</td>
</tr>
<tr>
<td>Grothner &amp; Bergner [39]</td>
<td>46 preservice elementary school teachers</td>
<td>Mean, median and mode</td>
<td>Questionnaire</td>
</tr>
<tr>
<td>Leavy &amp; O’Loughlin [41]</td>
<td>263 preservice teachers</td>
<td>Mean</td>
<td>Using a questionnaire consisting of the five tasks and individual clinical interview</td>
</tr>
<tr>
<td>Reading &amp; Reid</td>
<td>57 university students</td>
<td>Exploratory data analysis, probability, sampling distributions and inferential reasoning</td>
<td>Using minute papers</td>
</tr>
<tr>
<td>Watson et al. [31]</td>
<td>746 students in grades 3, 5, 7, and 9</td>
<td>Variation</td>
<td>Questionnaire</td>
</tr>
<tr>
<td>Makar &amp; Confrey</td>
<td>22 prospective secondary math and science teachers</td>
<td>Information notions of variation and distribution</td>
<td>Interview</td>
</tr>
<tr>
<td>Reading &amp; Reid</td>
<td>6 students from primary school and 6 students from secondary school</td>
<td>Variation</td>
<td>Interview</td>
</tr>
<tr>
<td>Reading [33]</td>
<td>Students in Grades 7, 9, and 11 (aged 13 to 17)</td>
<td>Variation</td>
<td>Task set in a real world context</td>
</tr>
<tr>
<td>Makar &amp; Confrey</td>
<td>4 secondary teachers</td>
<td>Comparing two groups</td>
<td>Interview using Fathom software</td>
</tr>
<tr>
<td>Groth [35]</td>
<td>15 high school students</td>
<td>Measures of central tendency</td>
<td>Problem solving clinical interview</td>
</tr>
<tr>
<td>delMas &amp; Liu [36,37]</td>
<td>12 university students</td>
<td>Standard deviation</td>
<td>Interview using conceptually enhanced software</td>
</tr>
<tr>
<td>Reading &amp; Reid</td>
<td>46 university students</td>
<td>Exploratory data analysis, probability, sampling distributions and inferential reasoning</td>
<td>Using minute papers</td>
</tr>
<tr>
<td>Torok &amp; Watson</td>
<td>16 students from grades 4, 6, 8 and 10</td>
<td>Variation</td>
<td>Interview</td>
</tr>
<tr>
<td>Batanero, Cobo &amp; Diaz [30]</td>
<td>2 samples of 14 years old (n=168) and 16 years old (n=144)</td>
<td>Average</td>
<td>Questionnaire with 9 open-ended tasks</td>
</tr>
<tr>
<td>Watson et al. [31]</td>
<td>746 students in grades 3, 5, 7, and 9</td>
<td>Variation</td>
<td>Questionnaire</td>
</tr>
<tr>
<td>Makar &amp; Confrey</td>
<td>22 prospective secondary math and science teachers</td>
<td>Information notions of variation and distribution</td>
<td>Interview</td>
</tr>
<tr>
<td>Reading &amp; Shaughnessy [17]</td>
<td>6 students from primary school and 6 students from secondary school</td>
<td>Variation</td>
<td>Interview</td>
</tr>
<tr>
<td>Reading [33]</td>
<td>Students in Grades 7, 9, and 11 (aged 13 to 17)</td>
<td>Variation</td>
<td>Task set in a real world context</td>
</tr>
<tr>
<td>Makar &amp; Confrey</td>
<td>4 secondary teachers</td>
<td>Comparing two groups</td>
<td>Interview using Fathom software</td>
</tr>
<tr>
<td>Groth [35]</td>
<td>15 high school students</td>
<td>Measures of central tendency</td>
<td>Problem solving clinical interview</td>
</tr>
<tr>
<td>delMas &amp; Liu [36,37]</td>
<td>12 university students</td>
<td>Standard deviation</td>
<td>Interview using conceptually enhanced software</td>
</tr>
<tr>
<td>Reading &amp; Reid</td>
<td>46 university students</td>
<td>Exploratory data analysis, probability, sampling distributions and inferential reasoning</td>
<td>Using minute papers</td>
</tr>
<tr>
<td>Grothner &amp; Bergner [39]</td>
<td>46 preservice elementary school teachers</td>
<td>Mean, median and mode</td>
<td>Questionnaire</td>
</tr>
<tr>
<td>Cruz &amp; Garrett</td>
<td>94 secondary students aged 17 years old</td>
<td>Average</td>
<td>Open and multiple-choice questions</td>
</tr>
<tr>
<td>Leavy &amp; O’Loughlin [41]</td>
<td>263 preservice teachers</td>
<td>Mean</td>
<td>Using a questionnaire consisting of the five tasks and individual clinical interview</td>
</tr>
<tr>
<td>Reading &amp; Reid</td>
<td>57 university students</td>
<td>Exploratory data analysis, probability, sampling distributions and inferential reasoning</td>
<td>Using minute papers</td>
</tr>
<tr>
<td>Reading &amp; Reid</td>
<td>6 tertiary students (pre-interviews) and 4 students (post-interviews)</td>
<td>Variability, comparing data sets, sampling and probability</td>
<td>Interview</td>
</tr>
<tr>
<td>Sharma [46]</td>
<td>24 pre-service teacher education students</td>
<td>Variability</td>
<td>Questionnaire</td>
</tr>
<tr>
<td>Watson, Callingham &amp; Kelly [47]</td>
<td>73 students (18 from Grade 3, 18 from Grade 5, 15 from Grade 7, 15 from Grade 9, 7 six-year-old children)</td>
<td>Expectation and variation</td>
<td>In-depth interview tasks</td>
</tr>
<tr>
<td>Ciancetta [48]</td>
<td>275 undergraduate and graduate students</td>
<td>Comparing distribution of data</td>
<td>Task-based web survey and interview</td>
</tr>
<tr>
<td>Cruz &amp; Garrett</td>
<td>227 students (130 aged between 16 and 21 years old from secondary school and 97 aged between 22 and 49 years old from university)</td>
<td>Arithmetic mean</td>
<td>Questionnaire with open-ended and multiple-choice questions</td>
</tr>
<tr>
<td>Sharma [50]</td>
<td>29 students aged 14 to 16 years</td>
<td>Average</td>
<td>Interview using open-ended and close questions</td>
</tr>
<tr>
<td>Reid &amp; Reading</td>
<td>46 tertiary students</td>
<td>Describe and compare distributions; one-way analysis of variance, simple linear regression</td>
<td>Using class test and assignment questions</td>
</tr>
<tr>
<td>Canada [52]</td>
<td>50 middle school (24 7th graders and 26 6th graders) students and 58 pre-service teachers</td>
<td>Compare two data sets</td>
<td>Task using the aspects of average and variation</td>
</tr>
<tr>
<td>Watson [53]</td>
<td>109 students aged from 6 to 15</td>
<td>Variation</td>
<td>3 interview protocol</td>
</tr>
<tr>
<td>Sirnik &amp; Kmetic [16]</td>
<td>27 18-year-old students and 20 13-year-old students</td>
<td>Arithmetic mean</td>
<td>10-item test and 7-item test</td>
</tr>
<tr>
<td>Chatzivasiliou, Michalis &amp; Tsaliaki [54]</td>
<td>109 4th and 6th grade students</td>
<td>Arithmetic mean</td>
<td>Questionnaire</td>
</tr>
<tr>
<td>Peters [55]</td>
<td>16 secondary mathematics/statistics teacher-leaders</td>
<td>Variation</td>
<td>Semi-structured content interview with 3 main tasks</td>
</tr>
<tr>
<td>Turegun &amp; Reeder [56]</td>
<td>41 students from two introductory statistics course</td>
<td>Variability</td>
<td>9-item multiple-choice questionnaire</td>
</tr>
<tr>
<td>Jacobbe [57]</td>
<td>3 elementary school teachers</td>
<td>Mean and median</td>
<td>Interviews, questionnaire &amp; assessments</td>
</tr>
</tbody>
</table>
4.0 FINDINGS AND DISCUSSION

Based on the meta-analysis, there were six main types of methods utilized to assess students’ statistical reasoning in descriptive statistics, i.e., interview, survey or questionnaire, tasks, tests, minute paper, and teaching. Among these four methods, interview was the most used method by the researchers, as 47.2% from the studies (17 out of 36) employed the interview method. This is mostly because by using interview method, the researchers are able to investigate and probe the responses of the respondents in order to collect in-depth information about their feelings and experiences [58]. On the contrary, teaching is the least used method for researchers to assess statistical reasoning among the participants. It was only used in one out of 36 studies (0.03%). Other methods are survey or questionnaire (12 studies), tasks (3 studies), tests (3 studies), and minute paper (2 studies). On the other hand, there were only three out of 36 studies (0.08%) that involved the usage of information technology in assessing reasoning about variability and distribution. It means that a majority of the studies did not utilize information technology in the assessments.

There were some previous studies that have been carried out to assess students’ reasoning about measures of central tendency as shown in Table 1. It seems that all researchers had exploited qualitative techniques to evaluate students’ reasoning about measures of central tendency, including the usage of interview and open-ended problems. Five were executed on secondary school students, two of the studies were executed on primary school students while three were executed on both primary and secondary school students. Meanwhile, two studies were executed on preservice teachers, one was executed on teachers, one was executed on both secondary school and university students, and one was executed on primary school, secondary school and college students.

Table 1 reveals some previous studies that have been performed to assess the reasoning about variability among students and teachers. These are made up of qualitative evaluation approaches that include interview, open-ended tasks, and questionnaire. There are some other approaches utilized as well. For example, delMas and Liu [36] used a technological tool in their interview while Reading [33] made use of a real-world task. There were seven studies that involved university students as their respondents. Five studies were conducted on primary and secondary school students, one was secondary school students, one was on school teachers and one was on preservice teachers. On the other hand, Table 1 demonstrates a number of methods employed to assess reasoning about distribution among students and teachers including tasks, interview, and project. Three out of six studies involved teachers as the participants, while others involved primary school and secondary school students (one study), university students (one study), and secondary school students and preservice teachers (one study).

The current meta-analysis revealed that there were a lot of assessments used to assess students’ reasoning about measures of central tendency, variability and distribution. However, most of the assessments were traditional assessments such as paper-and-pencil tasks and multiple-choice questions. Some traditional forms of assessments were not designed to align with the recent curriculum and instructional goals; hence they cannot provide a clear picture of students’ understanding and knowledge. Not only that, they were also too restricted to assess students’ understanding [59]. Thus, it is proposed to integrate information technology in the statistical reasoning assessments to construct new and different assessments in future research. These days, the utilization of information technology in the assessment is gradually becoming crucial to improve pedagogical innovation and curriculum reformation [60]. Appropriate usage of information technology can promote students’ statistical understanding [61] as well as facilitate the statistical process, including posing questions, gathering and analyzing data as well as interpreting the findings [62].

In fact, there are many advantages in using information technology in statistics classes. One of the benefits is that it can mitigate time and burden of students to handle tedious and cumbersome calculations when dealing with a wide array of data. This enables students to have adequate time to explore, analyze and interpret data [63]. Another benefit is that information technology can assist students to understand the abstract idea of statistics. Students could display and visualize data sets in multiple graphical representation forms such as histograms and box plots by using a computer [9], thus enhancing their understanding of statistical data, analysis, and graph as well as eradicate their misconceptions [64]. Pratt, Davies and Connor [65] argued that graphical representations that are generated by computers are not merely used as presentation tools, but also as analytical tools in data investigation. Furthermore, utilization of computer in distance learning enables students to work on their own pace outside the classroom as the web-based resources are always obtainable [66]. They can simply access the resources at any time and any place they desire as well as communicate among themselves conveniently via email [67].

On the other hand, this meta-analysis indicates that almost all the studies only focused on one concept of statistical reasoning. Hence, it is recommended that three statistical reasoning topics (reasoning about measures of central tendency, variability, and distribution) are integrated into one assessment for further investigation in this area. Even though the researchers and instructors have begun to emphasize on central statistical concepts or ‘big ideas’ in teaching and learning statistics [9], the incorporation of these central statistical concepts into assessment is still inadequate and many students still cannot see how these concepts are interconnected [68]. By combining these three statistical reasoning, students can see these concepts as a whole entity rather than as isolated concepts [69]. Besides, it also promotes their conceptual understanding on statistical concepts and reduces their misconceptions in statistical reasoning.

5.0 CONCLUSIONS

This study reports a qualitative meta-analysis with 36 studies on assessing students’ statistical reasoning in descriptive statistics. There were six main types of methods used to assess students’ statistical reasoning in descriptive statistics including interview, survey or questionnaire, tasks, tests, minute paper, and teaching. However, most of studies involved the usage of traditional assessments which did not utilize any technological tool. Hence, it is suggested to integrate information technology in statistical reasoning assessments in future exploration. In addition, the inclusion of three statistical reasoning in descriptive statistics, i.e., reasoning about measures of central tendency, variability and distribution, is recommended to be combined in a single assessment as well.
Acknowledgement

The authors would like to thank for the support given to this research by Universiti Tun Hussein Onn Malaysia (UTHM), under grant Vot: U365.

References


