International Journal of Assessment Tools in Education
2023, Vol. 10, No. 4, 672–689

https://doi.org/10.21449/ijate.1255859

Published at https://ijate.net/ https://dergipark.org.tr/en/pub/ijate ***Research Article***

Determining the psychometric properties of middle school statistical thinking testlet-based assessment tool

Lim Hooi Lian[1,*], Wun Thiam Yew[1]

[1]School of Educational Studies, University Sains Malaysia, 11800 Minden, Penang, Malaysia

ARTICLE HISTORY

Received: Feb. 24, 2023
Revised: Oct. 16, 2023
Accepted: Oct. 23, 2023

Keywords:
Content validity,
Rasch model,
SOLO model,
Statistical thinking,
Construct validity.

Abstract: The majority of students from elementary to tertiary levels have misunderstandings and challenges acquiring various statistical concepts and skills. However, the existing statistics assessment frameworks challenge practice in a classroom setting. The purpose of this research is to develop and validate a statistical thinking assessment tool involving form one (Grade 7) students' statistical thinking. The SOLO model was applied to develop five testlet tasks. Each testlet task involved four components. This study employed the survey methodology to assess the statistical thinking of 356 form one students. Content validity was determined using the Content Validity Index (CVI). The construct validity was determined using Rasch analysis. The results demonstrated that the instrument for assessing the statistical thinking of the form one students was valid and trustworthy. This finding of the study also revealed new evidence that the instrument allowed the teachers to identify the students' progress effectively based on the hierarchical manner of item levels in the testlet format. The instrument was useful in identifying students' statistical thinking levels. The students' ability to respond appropriately to a task at a particular level reveals their degree of cognitive development. Testlet task was also easy to diagnose the strengths and weaknesses in learning statistics topics.

1. INTRODUCTION

Quantitative information is everywhere. Everyone depends on statistical information to make interpretations and decisions. For instance, the outbreak of the coronavirus (COVID-19) pandemic is one of the obvious examples that require government, authorities, and citizens to make informed decisions based on the latest statistics data (Bilgin, Bulger & Fung, 2020). In the current world dominated by quantitative information, it is impossible to avoid tables, charts, raw scores, central tendency, and dispersion values. Consequently, the study of statistics offers a crucial tool for educating the population to respond wisely and logically to quantitative information in their environment.

All citizens should develop this important skill as a part of their daily lives. As stated by Kerka (1995), in order to become rational, creative, and dynamic citizens in today's society, we should equip ourselves with the ability to make use of quantitative information. Consequently, statistics have emerged as one of the core subjects included in the mathematics curriculum of most countries. For instance, statistics is a key component of the secondary as well as

*CONTACT: Lim Hooi Lian ✉ hllim@usm.my School of Educational Studies, University Sains Mallaysia, Pulau Pinang, Malaysia

primary mathematics curricula in Malaysia. In Malaysia, teaching and learning statistics start at age 8, namely from year two and continuing through year six. Primary students are taught how to organize and group data, read data from charts, construct pictographs, pie charts, and bar charts, by using the fundamental concept of central tendency to interpret data and solve problems involving data handling in everyday life (Malaysia Ministry of Education, 2017; 2018; 2019). In form one of middle school, students learn data representation and data interpretation with the intention of resolving complex routine problems. In form two, the students study and creatively use the concept of central tendency in the context of non-routine problem-solving (Malaysian Ministry of Education, 2018; 2019).

1.1. Review of Related Studies

However, past research has shown that the majority of students from primary school to tertiary level have misunderstandings and difficulties to acquire numerous statistical concepts and skills. These include statistical graph complexity and reading (Arteage et al., 2015), interpretation of box plots (Pierce & Chick, 2013), data representation (Chick & Pierce, 2012; Saidi & Siew, 2019; Ibnatul et al., 2021; Subanji et al., 2021), and descriptive statistics (Chan, et al., 2013), variability (delMas & Liu, 2005; Matthew & Clark, 2007), distribution (Lee & Meletiou-Mavrothesis, 2003), measures of central tendency (Cooper & Shore, 2008; Olani et al., 2011). Malaysian form four students tend to have inferior statistical reasoning abilities, claimed Chan and Zaleha (2014), Krishnan and Noraini (2014). On the other hand, asserted that Malaysian school-level examinations placed a greater emphasis on students' computational proficiency than on their growth as thinkers. As a result, the students unconsciously form a mental habit that highlights statistics' emphasis on mathematics, but rarely gives equal importance to the knowledge development in their statistical learning, which entails higher-order thinking ability. Obviously, this confusing circumstance suggests that the students are having troubles in mastering statistics.

Consequently, the effort to strengthen teaching and learning statistics has been thoroughly examined in a number of research (e.g., Tishkovskaya & Lancaster, 2012; Krishnan & Noraini, 2014; English & Watson, 2015; Setambah et al., 2019). One of them includes the need to reform statistics teaching and learning is about assessing students' learning outcomes, notably statistical reasoning. Consequently, assessment is crucial to the process of learning and teaching. By creating a more suitable teaching and learning environment, the teachers will be better equipped to lead and direct the students in light of the assessment's findings.

On the creation of the assessment framework, several studies have concentrated on statistical thinking. As an illustration, Jones et al. (2000), Mooney (2002), and Watson (2005) developed a framework for determining the level of middle school students' statistical thinking based on the analysis of their responses to the assigned tasks. In contrast, the methodology for assessing the level of statistical reasoning regarding descriptive statistics among high school students was established by Groth (2003), and Chan and Zaleha (2013). Aoyama (2007) used the paradigm developed by Callingham and Watson (2005) to examine how well high school through to university students could comprehend graphs. Aisah et al. (2018) developed an assessment framework based on the updated Bloom Taxonomy to assess the statistical reasoning of engineering students.

Nevertheless, the current assessment methods make it difficult to put theory into reality in a classroom environment. The application of the qualitative approach, particularly in interviews, is extremely time-consuming for the teacher. This assessment approach can only be used to assess a small number of students. Secondly, the implementation primarily relies on the teacher' understanding of the guiding principles of the model, including the SOLO model, to correctly categorize the responses of their students. Subsequently, it is challenging to identify the interconnection of the statistical concepts, particularly when using fundamental statistics skills to produce the new dimensional information (provide the data with an explanatory context or a set of conditions, like making predictions and providing a reason). Furthermore, the existing assessment frameworks have been specially designed for certain curriculum levels in the certain nations. Hence, the newly developed statistics assessment framework and instrument are crucial in establishing the high degree of validity and reliability of the mathematics assessment framework in Malaysia, particularly at the middle school level.

In order to address these issues, this study offered valuable pointers on the development of statistical thinking assessment practice via two-pronged approaches. The assessment framework might first be developed. Then, it is applied to design the hierarchical structure for the statistics tasks based on the descriptors of the framework. The framework can also be easily adopted or adapted to assess other topics of statistics as well. The students' ability to respond appropriately to the task at a particular level reveals their level of cognitive development. Instead of using a qualitative approach, it demonstrates a more complex use of the SOLO model in reverse (testlet task format). It means that the teachers are able to identify their students' strengths and weaknesses easily by referring to the item level that their students managed to solve. It is very convenient and meaningful to be applied in formative assessment. Besides, the overall performance of the topic can be determined effectively as the data are analyzed quantitatively. Thus, summative assessment can also be implemented using the newly developed assessment framework.

Nevertheless, it is more beneficial to practice this testlet task in the combination of formative and summative assessment. Before giving insightful comments to students, teachers can use the testlet task to properly determine the students' strengths and weaknesses. In the meantime, the summative assessment could be utilized to assess the comprehensive performance of the students after the process of teaching and learning for such a topic.

Hence, the purpose of this was to develop a statistical thinking assessment framework and assessment tool involving form one (Grade 7) students' statistical thinking. Second, the reliability and validity of the newly developed instrument were determined to ensure the application was valid in a classroom setting.

1.2. Development of Statistical Thinking Assessment Framework

Jones et al. (2000) developed an assessment framework to characterize the statistical reasoning of elementary and middle school students. The framework identifies four statistical processes, namely: (1) describing data: which involves trying to establish clear and specific information shown visually, identifying graphical conventions, and making important linkages between the display and the original data; (2) managing and minimizing data which involves mental activities on data such as rating, summarizing and grouping, (3) representing data: the construction of visual representations that demonstrate diverse organizational patterns of data; (4) evaluating and interpreting data: the interpretation of statistical results.

Contrarily, Garfield (2002) created an assessment model for statistical reasoning that took into account the five different types of reasoning: procedural, verbal, idiosyncratic, transitional, and integrated process. Students exhibit knowledge of several statistical elements and indicators at the level of idiosyncratic reasoning. Nonetheless, they frequently employed them without completely comprehending them. Thus, their perception of the meaning is frequently wrong. Consequently, students frequently mix them with irrelevant content. At the verbal reasoning level, students demonstrate their verbal comprehension of specific statistical principles, but they are incapable of applying these concepts to actual behavior. In other words, students are able to provide an accurate definition of a concept, but they may not understand it fully. Further, they could not understand how to combine statistical ideas or methods to reach the level of transitional thinking. They could, however, be able to tell them apart with accuracy.

Students frequently demonstrate the capacity to comprehend and adequately grasp the dimensions of statistical ideas or methods at the procedural reasoning level, but they are unable to integrate them properly. If the students are able to fully understand the statistics concepts and confidently decide on the statistics rules and concepts to be applied, it could be determined that the students have reached this level.

In the meantime, Chan and Ismail (2014) developed a tenth-grade assessment framework that incorporated three types of statistical reasoning employing information technology, namely centre, spread, and distribution reasoning. The students understood these three features as whole entities rather than isolated and separated features. The framework was based on the previous framework developed by Mooney's assessment framework. To assess the students' skills in statistical reasoning, a task-based interview was utilized.

The aforementioned assessment frameworks have substantially contributed to the assessment of students' statistical thinking. Although their applicability is confined to qualitative methods of assessment, such as an interview, they appear less applicable in the actual classroom context. Consequently, the purpose of this work is to address this constraint.

1.3. SOLO Model as Reference in Developing Assessment Framework

This assessment framework of this study was devised using the Structure of the Observed Learning Outcome model (SOLO model). Biggs and Collis created this assessment method in 1982. This model has highlighted cumulative cognitive components and latent hierarchy. It reveals that when students respond to a task given, their structure responses can be analyzed and categorized into five levels, from pre-structural to extended abstract levels. These are the characteristics of the hierarchical levels:

a. Prestructural - the students provide their answers without addressing the issue. They lack comprehension of the purpose of the specified task.
b. Unistructural - the students will answer by concentrating on one or a small number of information that direct tangible aspects offered and allotted to them. The information can be accessed via the stem (problem scenario) or the given graphic.
c. Multistructural - the response of the students is to collect more or all the pertinent information provided to acquire the answer. The information may be used as a recipe in which a series of steps are performed sequentially to complete the task. However, they are not integrated by the students.
d. Relational - the students' response entails integrating all parts of the provided information into a cohesive framework or structure. In other words, the information offered is insufficient to tackle the problem immediately. Instead, it must be interconnected carefully to generate an acceptable answer.
e. Extended abstract - the students reply by generalizing the framework or implying a distinct and more abstract context.

The SOLO model and the idea of the testlet were combined to construct an assessment framework and assessment tool for assessing the statistical thinking skills of middle school students. The purpose of this combination was to provide a more user-friendly and practical instrument for assessing students' level of statistical thinking, diagnosing their strengths and shortcomings in relation to this issue.

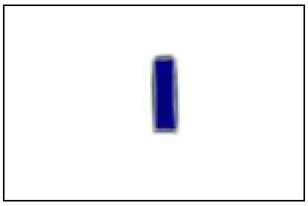
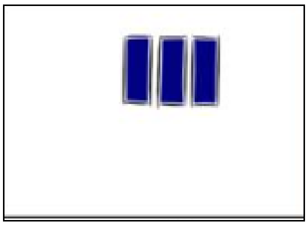
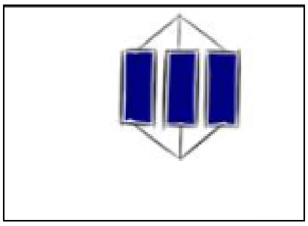
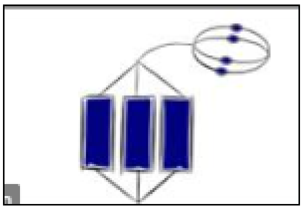
The task's structure consisted of two components. The stem is the initial part of the structure. It describes the situation or issue in the format of paragraphs. The second component is comprised of the four-level items that correspond to the four main levels of the SOLO model. The students' ability to react accurately at a specific level of the item implies that they have attained a particular level of cognitive capacity. The teacher is capable of recognizing the deficiencies of their students depending on the level achieved.

The following criteria were utilized to develop the items for each level:

a. Unistructural – apply directly one observable set of information, source, or material presented in the stem to deliver the answer. When interpreting the information provided, just a single aspect of the information will be considered.
b. Multistructural – employ a more prominent piece of information, or source supplied in the stem to react. At this level, the interpretation of the information may involve more information points without establishing a relationship between them.
c. Relational – use the additional information provided to create a framework by integrating the given information. All accessible information will be integrated at this stage to generate a cohesive meaning and structure.
d. Extended abstract - make deductions and predictions based on knowledge, critical and logical thinking, not simply referring the information in the stem.

Table 1 displays the requirements for the various levels.

Table 1. *Criteria of levels based on SOLO Model.*

Level	The pattern of the response structure	Description
Unistructural		One aspect of the information provided in the task is used to give a response.
Multistructural		Several relevant aspects of the information provided in the task are used to give a response.
Relational		Integrate all the information provided in the task is used to give a response.
Extended abstract		Make hypotheses or predictions for the new situation of the task.

In order to assess middle school students' development in statistical thinking, this study tried to create a thorough and all-encompassing assessment framework. This study specifically focused on two major views in developing an assessment framework: (a) determining the content domains that are essential to the development of statistics thinking in middle school students, and (b) defining and categorizing the levels of statistical thinking across the content domains.

Van de Walle et al. (2014) asserted that in doing statistics, the information analysis process involves three main stages, namely classification, graphical representations, and interpretation of results. Classification refers to deciding how to categorize things. The students normally learn this skill during their early grades. Next, graphical representations and interpretation of findings were the primary focus of this study, as they are taught in middle schools in most nations. In the Malaysian Form One (Grade 7) Secondary School Standard Curriculum, for example, the curriculum requirement of statistics themes comprises the process of gathering, organizing, and representing information, as well as the interpretation of that representation. The following are the key learning standards that encompass the content standards:

a. Construct or convert information representations, such as numerous forms of stem-and-leaf plots, bar charts, dot plots, line graphs, and pie charts.
b. Interpret various information representations, including making inferences or predictions.

In this study, the content domain of the curriculum was utilized to construct the statistical processes for the assessment framework. The cognitive processes of Form one students engaged in the information handling process were represented by four statistical processes: understanding the information provided, calculating, and comparing the value of information, representing the information into various types and making inferences and predictions. On the basis of the four levels of the SOLO model, which are unistructural, relational, multistructural, and extended abstract, the four processes were assessed on four thinking levels. The views of Curcio and Mooney were revised to categorize statistical processes hierarchically, according to the SOLO model. The four levels of information handling implemented by Curcio and Mooney's are as follows:

a. Reading the data: Reading data involves a significantly low cognitive level of data processing. The students merely select the data or facts that are clearly presented in the table or graph, such as the axis labels and titles. At this stage, analysis is unnecessary.
b. Reading between the data: Reading between the data demands a higher level of data handling skill. The interpretation of the data in the table, chart, or graph is included in this level of data handling. It calls for the students' capacity to compare values such as the most significant value, lowest value, minimum points or calculate the quantities such as differences, mean, and range using the fundamental mathematical concepts and skills (such as addition, subtraction, multiplication, and division).
c. Representing data: Graphical representation of data, such as a chart, graph, or dot plot, is known as representing data. This level requires the capacity to organize, interpret and analyze the data. This stage is necessary for displaying the data in a graphical format that reflects the main features of the original data.
d. Reading beyond the data: Reading beyond the data demands students to infer, predict, or make inference using their prior knowledge (or knowledge that they can recall) for the data given. The inference or prediction is neither explicitly nor implicitly stated in the data given.

Table 2. *General statistical thinking assessment framework.*

Category	Unistructural (reading the data)	Multistructural (reading between the data)	Relational (representing data)	Extended abstract (reading beyond the data)
Content Domain of Statistics	The readers simply apply a single data that is explicitly stated in the stem to respond. There is no interpretation at this level.	The readers refer to more or all data in the stem, then: i) apply the basic mathematical concept and skills (e.g., basic mathematics operation) to calculate the total value, differences or percentage; ii) compare the values given.	The readers link all the relevant aspects of data, and incorporate various aspects of his/her statistical thinking so that he/she can convert or represent the data in the appropriate graphical form.	The readers are perceived as having the ability to examine the data from more than one perspective. He/she makes inferences or predictions based on analytic and logical thinking, as well as his/her existing knowledge, not just the data in the stem.

Table 2 displays the basic assessment framework that combines the SOLO model levels and the viewpoints of Curcio and Mooney to classify the hierarchy of statistical thinking to be assessed. Unistructural, multistructural, relational, and extended abstract are the four major levels. At the unistructural level, the student responds by applying a single piece of data expressly provided in the stem. At the multistructural level, the student refers to additional data in the stem and employs some elementary mathematical skills in order to react. The student connects all the pertinent data points at the relational level and converts or represents the data in a suitable graphical representation. At the highest level, the students make inference and predictions using logical and analytical thinking based on both the data in the stem and his or her prior knowledge.

Table 3. *Statistical thinking assessment framework.*

Content Domain	Unistructural (reading the data)	Multistructural (reading between the data)	Relational (representing data)	Extended abstract (reading beyond the data)
Line graph	Apply single data in the stem (the number of students enrolled in the year 2016) to give the response.	Apply two data in the stem to find the difference between the two values.	All the data in the stem is analyzed and converted into a line graph	Make a prediction and provide a logical reason based on his/her existing knowledge and the data in the line graph developed.
Bar chart	Apply single data in the stem (the number of students who go to school by car) to give the response.	Apply all data in the stem and the basic mathematics operation to find the value in percentage.	All the data in the stem is analyzed and converted into a bar chart	Make an inference and provide a logical reason based on his/her existing knowledge and the data in the stem.
Pie chart	Apply single data in the stem (the number of Myvi cars sold in 2017) to give the response	Refer to all data in the stem and the basic mathematics operation to find the highest value.	All the data in the stem is analyzed and represented in a pie chart	Make an inference and provide a logical reason based on his/her existing knowledge and the data in the stem.
Dot Plot	Apply single data in the stem (the number of athletes who weigh 46 kilograms) to give the response	Apply all data in the stem and the basic mathematics operation to find the total value.	All the data in the stem is analyzed and converted into a dot plot	Make an inference and provide a logical reason based on his/her existing knowledge and the data in the stem.
Histogram	Apply single data in the stem (the number of students who spend 30-34 hours for their individual study) to give the response	Apply all data in the stem and the basic mathematics operation to find the value in percentage.	All the data in the stem is analyzed and represented in a histogram.	Make an inference and provide a logical reason based on his/her existing knowledge and the data in the stem.

The comprehensive assessment framework describing the features of each level for analyzing the relevant content domains is shown in Table 3. The four item levels are designed in accordance based on the assessment framework for the respective content domains. The capacity of students to accurately respond to a specific level of the items indicates their level of statistical thinking, as shown in Table 3.

2. METHOD

2.1. Research Design

In this study, the survey method was utilized. The purpose of using this method was to obtain authentic data from a large population. Besides, this method provides accurate and appropriate data regarding the characteristics of a particular individual, such as feelings, perceptions, attitudes, and knowledge.

2.2. Sample of Study

According to the Department of Education, there are 3720 (76%) form one students studying in the North-East District and 1200 (24%) form one students studying in the South-West District of the National Secondary Schools, in Penang Island. The desired sample size of form one students is 356, namely 270 form one students from six secondary schools of the North-East District and 86 form one students from two secondary school of the South-West District. The eight schools were randomly selected from a list of schools. To ensure that the results should accurately reflect the population's average performance, form one students from high, middle, and low-performing classes were included in the sampling from each school. The latest school mathematics test result was referred to identify the students' mathematics performance.

2.3. Instrumentation

In the development of statistical thinking assessment tool, the standard manner of instrument development stages which was suggested by Miller and Lovler (2018) was applied, namely:

1. Define the constructs and purpose of assessment. Literature review and document analysis were the focus at this stage. The concept and definition of statistical thinking was identified through the document analysis and curriculum content analysis. The purpose of the assessment was determined by the issues detected in the current teaching and learning classroom. Then, an accurate assessment model was searched to assist in the development of the instrument.
2. Develop the test plan and item format – The SOLO model had been applied in developing general statistical thinking framework (refer to Table 2). Then, the item format (refer to Figure 1) and number of items for each content domain was designed.
3. Compose the test item for each content domain based on the statistical thinking assessment framework (refer to Table 3)
4. Pilot and review the test item – The first draft of the test items was reviewed and judged by five experts. The result of the content validity evaluation was quantified using the CVI (Content Validity Index) formula (refer to Table 4). Then, the test was piloted to ensure the language and instructions were clear and easy to understand, as well as to revise the poor items detected.
5. Validate the test and conduct item analysis – After the test was revised. the validation and item analysis processes were gone through based on the data collected from 356 samples.

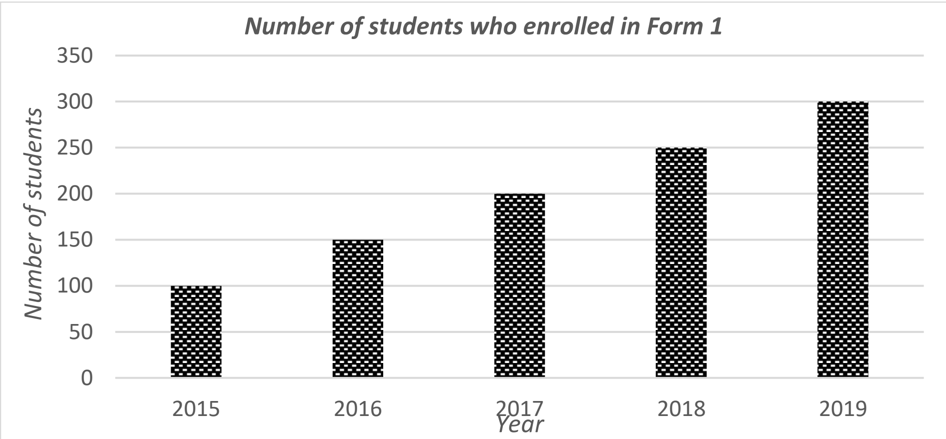
The assessment framework was referred to in creating the testlet task (refer to Table 3). The four items comprising the four levels of the SOLO Model were placed following the stem in the testlet task. Item 1 asked students to refer to a value on the displayed bar chart to produce the right response, which was the number of students enrolled in the 2016 academic year. In item 2, students were required to find the difference between the number of students registered in 2015 and 2019 using the bar chart provided. In item 3, all the values displayed in the bar graph were analyzed and converted into a new graphical form, namely a line graph. In the final item, students were required to make a prediction based on the line graph they created. Afterwards, students were required to produce an explanation based on their logical thinking and prior knowledge, as well as the data in the line graph.

The instrument of this study was composed of five testlets: a line graph, a bar graph, a pie chart, a dot plot, and a histogram. These five testlets reflected the five content domains of the framework. Each testlet included four items. Consequently, there were 20 items in all. All the items were developed in an open-ended format. An example of a line graph testlet created for this study is shown in Figure 1.

The most important procedure after developing the new instrument is to determine its validity. According to AERA (American Educational Research Association) et al. (2014), validity is the degree of evidence supporting the interpretations of the assessment score obtained from the suggested assessment tool. Thus, in this study, the newly developed framework and testlet task required a clear and explicit definition of the domain to be assessed. The validity of the evidence based on the assessment framework and the testlets must be obtained before it can be implemented. In a non-statistical form of validity known as content validity, the content of the test is systematically examined to ensure that it represents a representative sample of the behavior domain being assessed (Anastasi & Urbina, 1997).

Figure 1. *Example of a line graph testlet.*

The number of form one students who registered at SMK Sungai Pasir during a five-year period is depicted in the accompanying bar chart.



a. How many students were enrolled in 2016?
b. Calculate the difference in the number of students enrolled in 2015 and 2019.
c. Convert the bar chart into a line graph.
d. Is it possible to expand your line graph to forecast student enrollment for 2020? Give a reason.

2.4. Data Collection

Each testlet task has four levels of items. Consequently, there were 20 items in all. The items in the five testlet were in an open-ended format. Researchers administered the test to the students after getting the approval of the State Education Department, the participating schools, teachers, and students. The students were given one hour and thirty minutes to answer the 20 items.

It has always been rather arbitrary to choose the appropriate number of experts. According to Zamanzadeh (2015), it was advised that the chance of agreement be sufficiently controlled by at least five experts. In this study, the coverage of all the testlet tasks and the relevance of each item in terms of content validity were assessed by five experts. Three out of the five experts were secondary school teachers. They had been mathematics teachers for middle school students for more than five years. Two more experts were university lecturers in mathematical education. They had been educators of mathematics education courses for more than eight years. Independently, each expert assessed the tasks involved in the assessment. An independent judgement was necessary to guarantee that the experts were not affected by one another, according to AERA et al. (2014). The created assessment tool was revised and improved based on the experts' comments and opinions.

2.5. Data Analysis

The process of validation should involve gathering evidence that provides a sound scientific basis for the assessment score interpretation. The score interpretation included identifying the construct that the assessment tool was supposed to assess (AERA et al., 2014). To quantify the judgement data for content validity, the item-CVI (I-CVI) and scale-level CVI (S-CVI) were chosen. The construct-based validity evidence was also determined for the assessment tool. The Partial Credit Analysis Rasch Model was used to examine the data gathered from 356 form one students. In order to determine whether the assessment tool which were valid and reliable in terms of item dimensionality, item fit, item polarity, reliability, and separation, it was necessary to examine the construct validity.

3. RESULTS

3.1. Content Validity

In calculating CVI, two features must be determined, namely I-CVI and S-CVI. I-CVI are the proportion of the items of the instrument that receives ratings of three or four for their relevance and scope. The proportion of all items that are deemed to have content validity is known as S-CVI.

The significance of each item was rated by experts, often on a 4-point scale in order to determine an item-level CVI (I-CVI) in terms of relevance to the construct or domain assessed. Then, for each item, the I-CVI, was calculated by dividing the number of experts who gave it a score of three or four by the total number of experts. Each item having an I-CVI of 0.78 or above would be regarded as outstanding (Hair et al., 2014; Polit et al., 2007; Price et al., 1995). Researchers should be aware that when the number of panels increases, the probability of chance agreement decreases (Zamanzadeh et al., 2015).

The testlet-based assessment tool consisted of 5 tasks with 20 items. The tasks showed the high content validity of individual items (I-CVI range: 0.80 to 1.00), 13 items (65%) of the items scored 1 while the rest 7 items (35%) scored 0.80, which means all the items were appropriate, and there was no need for revision or elimination. The test generally showed high degree of content validity.

When there are more than two experts, as is most frequently the case, there are a few approaches to compute the scale-level content validity S-CVI. The most common calculation of the S-CVI/Ave is the average I-CVI across items (Polit et al., 2007).

Based on the feedback from five experts, the content validity for scale (S-CVI) was 0.93, which means the testlet tasks have achieved acceptable S-CVI, as shown in Table 4. The designed framework and the assessment tool achieved an acceptable level of content validity. The assessment tool's scale-level content validity and overall content validity index were both scored highly.

Table 4. *Content validity for an Item I-CVI.*

Question	Sub- question	rating 3 or 4	rating 1 or 2	I-CVI	Interpretation
1	(a)	5	0	1*	appropriate
	(b)	5	0	1*	appropriate
	(c)	5	0	1*	appropriate
	(d)	5	0	1*	appropriate
2	(a)	4	1	0.8*	appropriate
	(b)	5	0	1*	appropriate
	(c)	4	1	0.8*	appropriate
	(d)	4	1	0.8*	appropriate
3	(a)	4	1	0.8*	appropriate
	(b)	5	0	1*	appropriate
	(c)	5	0	1*	appropriate
	(d)	5	0	1*	appropriate
4	(a)	4	1	0.8*	appropriate
	(b)	4	1	0.8*	appropriate
	(c)	5	0	1*	appropriate
	(d)	5	0	1*	appropriate
5	(a)	4	1	0.8*	appropriate
	(b)	5	0	1*	appropriate
	(c)	5	0	1*	appropriate
	(d)	5	0	1*	appropriate

*I-CVI is higher than 79%

3.2. Construct Validity

3.2.1. *Item dimensionality*

The Rasch Model's unidimensionality is one of the most crucial factors in determining the construct validity of the assessment. To ensure that each item is associated with the same latent variable, the item dimensionality was determined. The identical latent variable was assessed, namely statistical thinking in this study (Bond, 2015). Principal Component Analysis (PCA) was analyzed for this component. In Figure 2, the variance explained by the measure was 44.2%. It was close to the model expected, which was 46.9%. The value of the unexplained variance in the first contrast was 6.7%, which was less than 15%, and the eigenvalue for the unexplained variance in the first contrast was 2.3, which is less than 5, confirming that there was no secondary dimension of latent variable that appears in this assessment despite being lower than the expected value and it was also less than 5 (Bond, 2015).

Figure 2. *Standard residual variance.*

```
     Table of STANDARDIZED RESIDUAL variance (in Eigenvalue units)
                                                 -- Empirical --    Modeled
Total raw variance in observations     =         34.0 100.0%        100.0%
  Raw variance explained by measures   =         15.0  44.2%         46.9%
    Raw variance explained by persons  =          6.2  18.3%         19.5%
    Raw Variance explained by items    =          8.8  25.8%         27.4%
  Raw unexplained variance (total)     =         19.0  55.8% 100.0%  53.1%
    Unexplned variance in 1st contrast =          2.3   6.7%  12.0%
```

3.2.2. *Item fit*

The item fit has to be determined to ensure all the data collected from the instrument fits the models that contribute to the unidimensionality. The reported fit statistics focused on two aspects: infit and outfit. The acceptable range for the value of infit and outfit was between 0.6 to 1.4 (Lincare, 1994). Figure 3 shows the related result. Every item fell within the acceptable range. Meanwhile, the mean infit MNSQ value for the item was 1.04. The model was able to predict the data too well, as seen by the item's mean outfit MNSQ score of 1.03 According to Linacre (2012), items with an MNSQ value closer to 1.00 was considered a better fit. In conclusion, the finding of the data had the overall fit and was accepted by the Rasch model.

3.2.3. *Item polarity*

The item was able to differentiate between the students' abilities, according to the item's positive PTMEA CORR value. Based on Figure 3, all the items had a positive value of PTMEA CORR, except for Item 5, in which the PTMEA CORR value was 0. It might be that the correlation could not be calculated due to the structure of the data or an extreme item (Bond, 2015). However, almost all the items were able to contribute to the assessment of the latent variable. According to Bond and Fox (2015), the items' good PTMEA CORR values demonstrated their ability to assess the construct that the testlet was supposed to assess and contribute to a high level of construct validity.

Figure 3. *Item statistic: Measure order.*

ENTRY NUMBER	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	PT-MEASURE EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	ITEM	G
12	143	356	2.41	.08	1.39	3.5	1.36	2.5	.22	.31	62.9	63.8	Q3D	Z
11	386	356	1.42	.06	.78	-4.0	.75	-3.7	.52	.44	41.3	33.3	Q3C	Z
8	464	356	1.18	.06	.73	-5.2	.75	-4.1	.42	.46	41.6	33.3	Q2D	Z
15	476	356	1.14	.06	.85	-2.7	.85	-2.4	.55	.46	34.0	34.4	Q4C	Z
19	484	356	1.11	.06	1.28	4.5	1.24	3.4	.52	.46	27.0	34.4	Q5C	Z
20	530	356	.98	.06	.73	-5.1	.73	-4.5	.45	.47	41.9	34.7	Q5D	Z
4	537	356	.95	.06	.92	-1.4	.96	-.6	.36	.47	41.0	35.6	Q1D	Z
16	603	356	.75	.06	1.17	2.6	1.14	2.1	.44	.48	33.4	36.6	Q4D	Z
18	477	356	.43	.07	.87	-2.4	.86	-1.7	.45	.43	46.1	43.6	Q5B	Y
9	232	356	.22	.12	1.05	1.3	1.06	1.2	.17	.26	64.6	67.7	Q3A	X
7	771	356	.15	.06	1.36	4.2	1.32	3.4	.46	.46	36.0	43.0	Q2C	Z
6	570	356	-.04	.08	1.03	.4	.92	-.5	.43	.41	66.9	66.6	Q2B	Y
3	824	356	-.09	.07	1.39	4.1	1.24	2.3	.44	.45	46.6	47.5	Q1C	Z
14	620	356	-.39	.09	1.02	.2	.92	-.4	.45	.38	82.6	80.1	Q4B	Y
2	665	356	-.90	.12	1.26	1.4	1.75	2.3	.15	.31	89.3	89.5	Q1B	Y
13	321	355	-1.50	.18	.94	-.4	.78	-1.3	.33	.19	90.7	90.5	Q4A	X
10	693	356	-1.54	.19	1.10	.4	.90	-.1	.29	.22	96.6	96.1	Q3B	Y
17	345	356	-2.74	.31	.94	-.1	.73	-.8	.27	.12	96.9	96.9	Q5A	X
1	351	356	-3.55	.45	1.00	.2	1.38	.8	.07	.09	98.6	98.6	Q1A	X
5	356	356	-6.39	1.83	MINIMUM MEASURE				.00	.00	100.0	100.0	Q2A	X
MEAN	492.4	355.9	-.32	.20	1.04	.1	1.03	-.1			59.9	59.3		
S.D.	171.4	.2	1.99	.39	.21	2.9	.28	2.4			24.5	24.8		

Table 5. *Person- item reliability and separation indices.*

Person/Item	Separation	Reliability
Person	1.37	0.70
Item	9.15	0.99

Based on Table 5, the reliability of a person was 0.70, while the separation of a person was 1.37. The lower value of the person separation value indicated that the limited data available to estimate the students' ability resulted in the lower value of the person reliability value. In contrast, the reliability of the item was 0.99, while the separation of items was 9.15. The value of 0.99 indicated that the item was very consistent. According to Fisher (2007), a reliability value greater than 0.94 is regarded as excellent. A good separation value was more than

2 (Linacre, 2012). It implies that the items might be divided into 9 categories based on the students' levels of ability.

The logit scale was an interval-level measurement scale that was used to both the person's ability and the level of difficulty as shown in Table 6. The four level items in each testlet (except testlet 3: Unistructual to multistructural level, testlet 4 and 5: Relational to extended abstract level) were ordered from the easiest to the most difficult, based on the lowest to the highest value of logit scales shown. It demonstrated that the hierarchy of items were meaaningful. Unistructural level was much easier than the extended abstract level. The relational level was more difficult than multistructural level.

The person locations were plotted to represent that if the students had a 50% probability of correctly answering the item that located at the same point on the logit scale. The students had greater than 50% probability of correctly answering the item with less difficulty, namely the item difficulty level was lower than the students' ability estimate. The bigger gap between the item difficulty and ability estimate, the greater probability of correctly answering the item.

Testlet 1 was an easy item. The four levels of SOLO model were displayed in a hierarchical manner. The majority of students (183) had greater than 50% probability of correctly answering the highest level of item. Meanwhile 152 students had 50% probability of correctly answering the relational level of item.

The four levels of the SOLO model were also displayed in a hierarchical manner in testlet 2. The majority of students (299) had greater than 50% probability of correctly answering the relational level of item. Only 47 students had greater than 50% probability of correctly answering the highest level of item.

The easiest item for testlet 3 was the second level of item, namely multistructural level. Majority of students (270) have greater than 50% probability of correctly answering this item. There was a gap in difficulty level between unistructural level and multistructural level, namely 1.76 logit. Nobody achieved the highest level of this item.

The most difficult item was at the relational level for testlet 4 and 5. There was a small gap between relational and extended abstract levels, namely .39 logits for testlet 4 and .13 logits for testlet 5. However, the majority of students (123) had greater than 50% probability of correctly answering this item.

Table 6. *Statistical thinking level.*

Level/ Testlet	Unistructural (logit scale)	The total number of students who had greater than 50% probability of correctly answering the item	Multistrutural (logit scale)	The total number of students who had greater than 50% probability of correctly answering the item	Relational (logit scale)	The total number of students who had greater than 50% probability of correctly answering the item	Extended abstract (logit scale)	The total number of students who had greater than 50% probability of correctly answering the item
1	-3.55	3	-.90	18	-.09	152	.95	183
2	-6.39	0	-.04	10	.15	299	1.18	47
3	.22	40	-1.54	270	1.42	46	2.41	0
4	-1.50	12	-.39	114	1.14	123	.75	107
5	-2.74	64	.43	109	1.11	123	.98	60

4. DISCUSSION and CONCLUSION

Validity is a unitary concept. It concerns the degree to which various pieces of evidence can be used to support the assessment score (AERA et al., 2014). Thus, evidence about the content of assessments and analyses from the individual response can be gathered to provide sound scientific validity evidence. Content-oriented evidence is the heart of the validation process in educational assessment. It concerns the alignment between content domains to be assessed and the item relevance as well as the content coverage represented by the items (AERA et al., 2014). Content validity index (CVI) is a well-known and widely used formula for quantifying content validity data analysis. There are two features to be determined in applying CVI, namely validity for the item (I-CVI) and content validity for scale (S-CVI). This type of validity offers initial proof of an instrument's construct validity. Additionally, it offers details on the representativeness and clarity of the items. The items might be improved based on the comments and suggestions from a panel of experts (Polit & Beck, 2006; Zamanzadeh et al., 2015).

I-CVI and S-CVI/Ave achieved an acceptable level based on the findings, and the assessment tool's content validity has been attained to a satisfactory level. The development of the testlet task demonstrated high item-content validity for assessing students' statistical thinking. Upon further validation, it is expected that this assessment tool might be used for the development and potential improvement.

The degree of association between the items that conformed to the constructs to be assessed may be determined by analyzing the evidence of validity based on the internal structure (AERA et al., 2014). It might indicate the single dimension of the trait to be assessed. In this study, the unidimensionality, item fit, item polarity, and reliability separation indices were the four primary factors that were required to be justified in order to determine the construct validity based on the principles of the Rasch Model. The newly created assessment tool satisfied the four key requirements stated in the Rasch Model analysis, according to the findings. Besides, Rasch analysis estimated the item difficulty and person ability along a standard scale, namely the logit scale, which provided a more accurate estimation between the difficulty of the item and the ability of a person. In short, the power of Rasch Model analysis enabled the estimation of construct validity regardless of the dependence on person ability (Teh & Lim, 2016). It was an established technique to improve the precision and accuracy in determining psychometric properties for the developed instrument and analyze the respondents' performance characteristics in a more sophisticated manner (Boone, 2016; Teh & Lim, 2016).

As the statistical thinking instrument met the analytical criterion, the results of the content validity and construct validity assessments indicated that it was trustworthy and valid. Therefore, the newly developed assessment framework and assessment tool showed a significant contribution to the new knowledge in statistics assessment in terms of practicality and usefulness. The well-hierarchically organized item level and the quantitative application method of analyzing student's scores proved that the assessment tool could be applied by mathematics teachers either informally (formative assessment) or formally (summative assessment) in their classroom setting.

This study also showed a fresh insight: if students are unable to perform at a higher level, the hierarchical items can assist teachers in identifying the students' deficiencies. This data is essential for teachers to create the most effective remediation strategies for their students. In addition, this testlet task is beneficial for practice in both formative and summative assessments. Before offering helpful feedbacks, teachers will be able to properly assess the students' problems and strengths utilizing the testlet tasks during formative assessment. In the meantime, the summative assessment might be utilized to assess the overall performance of the students after the teaching and learning process for this topic. Both purposes of the assessment can be administered in traditional format (paper-and-pencil test) or the computer mode. In conclusion, the establishment of this assessment framework and assessment tool had significantly contributed to a more effective and efficient assessment of statistical thinking.

Although there are some existing statistical thinking teaching tools developed specifically to the classroom environment such as AutoStat and Code-Driven Tool (Alston-Knox et.al., 2019; Fergusson, 2022), they have some limited fuctions in terms of the content and student level. AutoStat alleviaties the need for coding, emphasizes on statistical models, uses computers and coding to understand the models and algorithms (Alston-Knox et al., 2019). This software is only applicable in higher education level. Meanwhile Code-Driven Tool is a task design framework for developing statistical and computational thinking. The students may need to have some basic knowledge about computer coding in solving the tasks. In sum, if the teachers need to identify the students' strength and weakness in detail during the monitoring and teaching process in statistics topic, these both tools may not be appropriate.

This study provides a more systematic and effective assessment framework as well as a validated assessment tool to assess the middle school students' ability to think critically about the key concept of statistics, namely data handling and representation. Prior to understanding the more advanced concepts of statistics and probability, such as discrete and continuous probability distributions, sampling and estimation (point and interval), and hypothesis testing for population mean and proportion, students must grasp these essential statistical concepts. This assessment framework and assessment tool can be applied in any statistical topic. In addition, it could be utilized as a tool to enhance the students' statistical thinking and to identify their strengths and weaknesses.

Although the statistical thinking assessment tool has been validated, there are some limitations that need to be addressed. First, the assessment tool only included one testlet for each content domain. In future studies, researchers could expand the assessment framework by adding more items for the four levels of the data handling process. Furthermore, the assessment framework was developed to assess only one of the form one mathematics topics, namely statistics. It can be adapted and extended by including the statistics content of form two and form three in the assessment of the students' statistical thinking across the forms.

Acknowledgments

This research was supported by Bridging Grant (304.PGURU.6316212) from Universiti Sains Malaysia, Penang, Malaysia.

Declaration of Conflicting Interests and Ethics

The authors declare no conflict of interest. This research study complies with research publishing ethics. The scientific and legal responsibility for manuscripts published in IJATE belongs to the authors. **Ethics Committee Number**: Universiti Sains Malaysia, 12345-678.

Authorship Contribution Statement

Lim Hooi Lian: Investigation, Resources, Visualization, Software, Formal Analysis, and Writing-original draft. **Wun Thiam Yew**: Methodology and Validation.

Orcid

Lim Hooi Lian https://orcid.org/0000-0002-9089-2262
Wun Thiam Yew https://orcid.org/0000-0002-2714-9636

REFERENCES

Alston-Knox, C.L., Strickland, C.M., Gazos, T., & Mengersen, K.L. (2019). Teaching and Learning in Statistics: Harnessing the power of modern statistical software to improve students statistical reasoning and thinking, *Proceedings of the 5th International Conference on Higher Education Advances (HEAd'19)*. http://dx.doi.org/10.4995/HEAd19.2019.9239

American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME) (2014). *Standards for Educational and Psychological Testing.* American Educational Research Association.

Aisah, M.N., MazJamilah, M., Khatijahhusna, A.R., & Safwati, I. (2018). *Developing statistical reasoning and thinking assessment for engineering students: Challenges and new direction.* https://pdfs.semanticscholar.org/c899/c375341f3045d3bd0a5bac0536ce18694fa2.pdf

Anastasi, A., & Urbina, S. （1997）. *Psychological testing (7th ed.).* Prentice Hall.

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

Arteage, P., Batanero, C., Contreras, J.M., & Canadas, G.R. (2015). Statistical graphs complexity and reading levels: A study with prospective teachers. *Statistique et Enseignement*, *6*(1), 3-23.

Barham, A.I., Ihmeideh, F., Al-Falasi, M., &Alabdallah, A. (2019). Assessment of first grade students' literacy and numeracy levels, and the influence of key factors. *International Journal of Learning, Teaching and Educational Research*, *18*(12), 174-195. https://www.ijlter.org/index.php/ijlter/article/view/1840/pdf

Biggs, J.B., & Collis, K.F. (1982). *Evaluating the quality of learning: The SOLO taxonomy (Structure of the Observed Learning Outcome).* Academic.

Bilgin, A.A.B., Bulger, D., & Fung, T. (2020). Statistics: Your ticket to anywhere. *Statistics Education Research Journal, 19*(1), 11-20.

Bond, T.G., & Fox, C.M. (2015). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. Routledge.

Callingham, R., & Watson, J.M. (2005). Measuring statistical literacy. *Journal of Applied Measurement*, *6(*1), 29, 19-47.

Chan, S.W., & Zaleha, I. (2012). The role of data technology in developing students' statistical reasoning. *Procedia-Social and Behavioral Sciences*, *46*, 3660-3664.

Chan, S.W., & Zaleha, I. (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.

Chan, S.W., Zaleha, I., & Bambang, S. (2013). A Rasch model analysis on secondary students' statistical reasoning ability in descriptive statistics. *Procedia-Social and Behavioral Sciences 59*[Online], 133 – 139. http://www.sciencedirect.com/science/article/pii/S1877042814028407

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

delMas, R., & Liu, Y. (2005). Exploring students' conceptions of the standard deviation. *Statistics Education Research Journal*, *4*(1), 55-82.

English, L., & Watson, J. (2015). Exploring variation in measurement as a foundation for statistical thinking in the elementary school. *International Journal of STEM Education*, *2*(3). https://doi.org/10.1186/s40594-015-0016-x

Fergusson, A-M., G. (2022). *Towards an integration of statistical and computational thinking: Development of a task design framework for introducing code-driven tools through statistical modelling* [Doctor of Philosophy thesis, The University of Auckland]. https://researchspace.auckland.ac.nz/handle/2292/64664

Garfield, J. (2002). The challenge of developing statistical reasoning. *Journal of Statistics Education*, *10*(3). https://doi.org/10.1080/10691898.2002.11910676

Groth, R.E. (2003). *Development of a high school statistical thinking framework* (Unpublished doctoral dissertation), Illinois State University.

Ibnatul, J.F., Adibah, A.L., & Hawa, S.S. (2021). Assessing statistical literacy level of postgraduate education research students in Malaysian research universities. *Turkish Journal of Computer and Mathematics Education, 12*(5), 1318-1324.

Kerka, S. (1995). *Not just a number: Critical numeracy for adults* (ERIC Digest No. 163, Rep. No.EDO–CE–95–163). Columbus, OH: ERIC Clearinghouse on Adult, Career, and Vocational Education. (ERIC Document Reproduction Service No. ED 385 780).

Krishnan, S., & Idris, N. (2014). Investigating reliability and validity for the construct of inferential statistics. *International Journal of Learning, Teaching and Educational Research, 4*(1), 51-60.

Linacre, J.M. (1994). Reasonable mean-square fit values. *Rasch Meas. Trans.*, 8 (3), p. 370.

Linacre, J.M. (2012). *A user's guide to Winsteps Ministeps Rasch-model computer programs* [version3.74.0]. http://www.winsteps.com/index.htm

Mairing, J.P. (2020). The effect of advance statistics learning integrated Minitab and Excel with teaching teams. *International Journal of Instruction, 13* (2), 141-150.

Malaysia Ministry of Education (2017). *KSSM Mathematics Form One*. Putrajaya: MOE.

Malaysia Ministry of Education (2018). *KSSM Mathematics Form Two*. Putrajaya: MOE.

Malaysia Ministry of Education (2019). *KSSM Mathematics Form Three*. Putrajaya: MOE.

Matthews, D., & Clark, J. (2007). *Successful students' conceptions of mean, standard deviation and the central limit theorem*. http://www1.hollins.edu/faculty/clarkjm/stats1.pdf

Mooney, E.S. (2002). A framework for characterizing middle school students' statistical. *Mathematical Thinking and Learning, 4*(1). 23-63.

Olani, A., Hoekstra, R., Harskamp, E., & van der Werf, G. (2011). Statistical reasoning ability, self-efficacy, and value beliefs in a reform-based university statistics course. *Electronic Journal of Research in Educational Psychology, 9*(1), 49-72.

Pierce, R., & Chick, H. (2012). Workplace statistical literacy for teachers: Interpreting boxplots. *Mathematics Education Research Journal, 25*, 189-205. http://dx.doi.org/10.1007/s13394-012-0046-3

Polit, D.F., & Beck, C.T. (2006). The content validity index: Are you sure you know what's being reported? Critique and recommendations. *Res Nursing Health, 29*(5), 489-497.

Polit, D.F., Beck, C.T., & Owen, S.V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in Nursing & Health, 30*(4), 459–467. https://doi.org/10.1002/nur.20199

Saidi, S.S., & Siew, N.M. (2019). Assessing students' understanding of the measures of central tendency and attitude towards statistics in rural secondary schools. *International Electronic Journal of Mathematics Education, 14*(1), 73-86. https://doi.org/10.12973/iejme/3968

Setambah, M.A.B., Tajudin, N.M., Yaakob, M.F.M., & Saad, M.I.M. (2019). Adventure learning in basics statistics: Impact on students critical thinking. *International Journal of Instruction, 12*(3), 151-166. https://doi.org/10.29333/iji.2019.12310a

Subanji., Nusantara, T., Rahmatina, D., & Purnomo, H. (2021). The Statistical Creative Framework in Descriptive Statistics Activities. *International Journal of Instruction, 14*(2), 591-608. https://doi.org/10.29333/iji.2021.14233a

Tishkovskaya, S., & Lancaster, G.A. (2010). Teaching strategies to promote statistical literacy: Review and implementation. In data and context in statistics education: towards an evidence-based society. *Proceedings of the Eighth International Conference on Teaching Statistics*. International Statistical Institute.

Tishkovskaya, S., & Lancaster, G. (2012). Statistical education in the 21st century: A review of the challenges, teaching innovations and strategies for reform. *Journal of Statistics Education*, *20*(2), 1–55.

Van de Walle, J.A., Karp, K.S., & Bay-Williams, J.M. (2014). *Elementary and middle school mathematics: Teaching developmentally* (8th ed.). Pearson.

Wild, C.J., & Pfannkuch, M. (1999), Statistical thinking in empirical enquiry, *International Statistical Review*, *67*, 223-265.

Zamanzadeh, V., Ghahramanian, A., Rassouli, M., Abbaszadeh, A., Alavi-Majd, H., & Nikanfar, A.-R. (2015). Design and implementation content validity study: development of an instrument for measuring patient-centred communication. *Journal of Caring Sciences*, *4*(2), 165–178. https://doi.org/10.15171/jcs.2015.017

Zamanzadeh, V., Rassouli, M., Abbaszadeh, A., Majd, H.A., Nikanfar, A., & Ghahramanian, A. (2015). Details of content validity and objectifying it in instrument development. *Nursing Practice Today*, *1*(3), 163–171.