International Journal of Science Academic Research
Vol. 04, Issue 06, pp.5777-5780, June, 2023
Available online at http://www.scienceijsar.com

ISSN: 2582-6425

Research Article

DEVELOPMENT OF STATISTICAL REASONING ABILITY TEST INSTRUMENTS TEACHER CANDIDATE STUDENTS

***Laila Hayati, Ketut Sarjana, Harry Soeprianto and Nani Kurniati**

Mathematics Education Study Program, Faculty of Teacher Training and Education, University of Mataram, Indonesia

Received 20th April 2023; **Accepted** 14th May 2023; **Published online** 30th June 2023

Abstract

This research aims to develop a test instrument that measures the statistical reasoning ability of teacher candidate students. Statistical reasoning is the main goal in learning statistics, where more attention is give to descriptive statistics concepts. However, assessment of the concept of descriptive statistics has not been fully considered in education in Indonesia, especially in terms of summative and formative assessments oriented to the development of student competencies. There are relatively few statistical reasoning ability test instruments in the context of teacher candidate students. Therefore, this research intends to develop a statistical reasoning ability test instrument based on the framework of Chan & Ismail in the context of prospective teacher students. The development model used is a 4D model: *Define, Design, Develop,* and *Disseminate*. Information collection was carried out by validating test instruments by 3 validators of Mathematics and Statistics Education lecturers. The results of the study stated that the instrument developed was declared valid and could be used to measure the statistical reasoning ability of teacher candidate students.

Keywords: Test instrument, Descriptive statistics, Statistical reasoning, Teacher candidate students, Statistics education.

INTRODUCTION

In the Regulation of the Minister of ducation and Culture Number 23 concerning educational assessment standards (Kemendikbud, 2016), it is said that assessments that measure the achievement of student learning outcomes are an important part of the teaching and learning process This can give an idea of what quality of learning learners have learned and the problems they face (Magdalena *et al.,* 2020). Learning objectives in statistics today have focused on higher-order thinking skills, namely reasoning and statistical thinking (NCTM, 2000). Garfield, del Mas & Change (Ben-Zvi & Garfield, 2004), and Garfield & Ben-Zvi (2008) putting forward statistical reasoning be ability understand and be able to explain statistical processes, relate one concept to another, or be able to combine ideas about data and change including making interpretations based on data sets, data representations, or statistical summaries of data, and being able to fully interpret statistical results. Another opinion emphasizes that the verbs used in statistical reasoning assessments are why, how, and explain about the process (DelMas, 2002). In addition, it defines statistical reasoning as what students are capable of doing with statistical material (e.g. recalling, recognizing and discriminating among statistical concepts), and the skills demonstrated by students in using statistical concepts in certain problem-solving steps (Chervaney *et al.,* 1977). In practice, statistical reasoning involves assessing how data is collected, describing the data, making inferences from the data, and the nature of uncertainty in the results of using samples. Furthermore (Garfield, 2003) suggests several types of statistical reasoning abilities that must be developed in students, namely reasoning about data, reasoning about data representation, reasoning about statistical measures, reasoning about uncertainty, reasoning about samples and reasoning about associations.

Corresponding Author:** ***Laila Hayati,
Mathematics Education Study Program, Faculty of Teacher Training and Education, University of Mataram, Indonesia

However, most educators still assess learners at the procedural and computational levels, thus not providing a comprehensive picture of learners' ability to reason and think statistically. Moreover (Martadiputra & Suryadi, 2012), in his research revealed that prospective teacher students demonstrate statistical thinking skills that is still low. Thus, educators do not Just focus guiding learners at the procedural stage only, but train to Statistical reasoning and thinking. When learning is coaching higher order thinking skillsin learning, to measure the success of learning, it is necessary to assess (Jailani *et al.,* 2018). Therefore, summative and formative assessments are needed that measure the development of student competencies, especially at the higher level of thinking (reasoning and statistical thinking). Summative and formative assessment is needed for three reasons (Malawi & Maruti, 2016), first, to examine the conceptual and meaningful understanding of learners, second, to emphasize the learning process rather than the final result alone. Third, to stimulate a more effective learning process. Assessments that measure reasoning and statistical thinking can also encourage innovative learning and curriculum reform (Drijvers, Kodde-Buitenhuis, & Doorman, 2019). To be able to conduct assessments that measure student development, especially statistical reasoning, valid test instruments are needed and in accordance with the level of students. Until now, there are several types of assessment instruments to measure statistical reasoning and thinking skills (Ziegler, 2014), like *Statistical Reasoning Assessment (SRA), Comprehensive Assessment of Outcomes Test (CAOS), Assessment Resource Tools for Improving Statistical Thinking Scale (ARTIST).* These assessment instruments have been widely used in various countries, measuring two major concepts of statistics, namely descriptive statistics and inferential statistics, measuring elementary and secondary level students and using English. For this reason, assessment instruments are needed in accordance with the context in Indonesia and measure prospective teacher students. This is necessary because some topics are not suitable for the level of students in Indonesia, because they are not in the

syllabus/Semester Learning Plan. Thus, standardized test instruments are needed to fill this gap by formulating a new statistical reasoning ability test instrument according to the level of prospective teacher students in Indonesia, with special emphasis on descriptive statistical material.

RESEARCH METHODS

The test instrument framework developed in this study is based on the statistical reasoning framework proposed by (Chan & Ismail, 2014); namely describing data, organizing and reducing data, data representation, and analyzing and interpreting data. The material developed focuses on descriptive statistics with the context of prospective teacher students. The test instrument was then developed into 30 questions, consisting of 20 multiple-choice questions and 10 description questions. The Head of the Education Assessment Center of the Ministry of Education and Culture (Kemendikbud, 2020) revealed that the varied forms of questions aim to measure the level of cognition of students more deeply. The variety of questions is also expected to encourage students to have higher-order thinking skills that are the demands of the competence of the 21st century generation. (Jones *et al.,* 2000). In accordance with the needs of this study, adaptation and modification are carried out into 3 stages of research, in accordance with the design of 4D model development research, namely Define, Design, Develop and Dissemination. In this research carried out until the development stage. (Thiagarajan *et al.,* 1974). The flow of development research can be shown in figure 1 below.

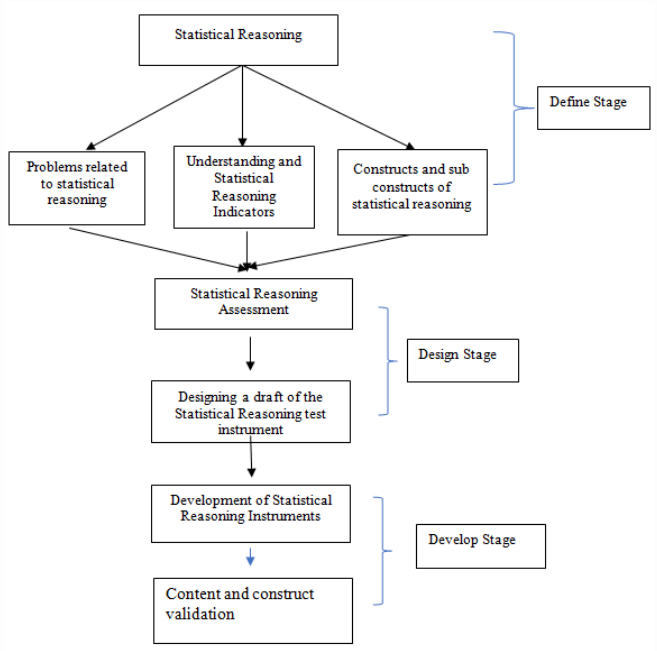

Figure 1. Development Research Flow

The first stage is defining, which is the stage to find problems, weaknesses, or conditions that are the root driving force of development activities. This stage is useful for collecting various information related to the product to be developed. In this stage, several analyses are carried out, including: initial final analysis to find out the basic problems in the development of test instruments. This initial analysis was carried out by observation and interviews with 1 lecturer who taught basic statistics courses, then student analysis, and material analysis. Furthermore, in the second stage of design, researchers designed a draft test instrument in Indonesian to measure students' statistical reasoning ability based on the statistical reasoning framework by Chan & Ismail (2014), and the third stage is development. At this stage, the draft test instrument that has been developed is validated by experts. The validation of the instrument carried out, is the validity of the content and the validity of the construct. The instrument developed was validated by 3 validators, namely expert lecturers in the field of mathematics and statistics education. Suggestions and inputs from validators will then be followed up to improve the developed instruments. The validity of the instrument by experts indicates the quality of the instrument in terms of content and construct. Content validity measures the degree of ability of the test to measure the scope of the substance of the element to be measured while construct validity measures how far the test items are able to measure what is really to be measured according to a predetermined conceptual definition (Sugiyono, 2014)

To measure content and construct validity, the formula used is Aiken Validity: (Aiken, 1980)

$$V = \frac{\sum s}{n(c-1)}$$
$$s = r - l_0$$

V= validity index of Aiken's V
l_0= lowest validity assessment number,
c= highest validity assessment rate,
r= number given by validator

The validity criterion that the researchers used was a modification of the criteria proposed by Aiken(Annabil, Wulandari, Yamtinah, Ariani, & Ulfa, 2022) as presented in Table 1 below.

Table 1. Aiken's V Assessment Criteria

V Price	Interpretation
0,80-1,00	High
0,40-0,80	Moderate
0, 00-0,40	Low

The test instrument can be used if the validator's assessment shows a V value in the moderate category. If the validity category is low, then the instrument will be revised first according to validator suggestions.

RESULTS

The statistical reasoning test instrument was developed in accordance with the characterization of teacher candidate students and the Semester Learning Plan in Indonesia based on four key constructions developed by Chan & Ismail (2014). Statistical reasoning is included in higher-order thinking skills. For this reason, to assess statistical reasoning ability, the characteristics of assessment instruments that measure higher-order thinking skills are needed (Ariyana *et al.,* 2018). The characteristics of assessment instruments that measure higher-order thinking skills (Kemendikbud, 2017). is: (1) assessing higher-order thinking skills; (2) leads to contextual problems; (3) The questions used vary.

This research has determined indicators of statistical reasoning ability, namely describing data, organizing data, representing data, and analyzing and interpreting data (Chan &; Ismail,

2014). In terms of describing data, it consists of several sub-constructs, namely obtaining and producing information from data or graphs, showing awareness of the characteristics/ properties/ symbols displayed by a representation graph, recognizing the characteristics of a representation graph. In terms of describing data, it consists of organizing data using a central measure, by calculation, organizing data using spread size by calculation. In terms of representing data, it consists of presenting data using various representations: tables, diagrams, and graphs, as well as identifying different representations for the same data set, assessing the effectiveness of two different representations for the same data. In terms of analyzing and interpreting data consists of Making comparisons within the same data set and making comparisons between two different data sets, as well as making predictions, inferences or inferences from data or graphs.

Here is one example of a statistical reasoning test instrument developed:

Wahana Lingkungan Hidup (WALHI) is an organization of environmentalists in Indonesia who want to know whether residents in a certain area consume foods that contain harmful substances (such as lead) in levels that are categorized as unhealthy. For this reason, Walhi conducted a study with lead level data from a random sample of 23 people as follows:

2,8
2,8; 3,0; 3,5; 3,8; 3,8; 4,1; 4,1; 4,2; 4,6; 4,8; 4,8; 5,0; 5,1; 5,2; 5,2; 5,3; 5,4; 5,9; 5,9; 6,3; 6,4; 6,6; 6,8

Biologists classify lead levels as unhealthy if their content is greater than 6.0 ppm.

1. What proportion of the population in the sample has lead levels classified by biologists as unhealthy?
2. Make a matching diagram from the data!
3. Write down the conclusions from the data.

The question assesses statistical reasoning skills because learners must relate and reason about how to classify unhealthy, choose diagrams that match the type of data, and make conclusions from the data presented.

The statistical reasoning ability test instrument that has been prepared has been tested for the validity of its content and construct. A valid instrument must have both internal and external validity. The internal validity of the instrument in the form of a test must meet the validity of the construction and the validity of the content. Instruments that have construction validity, if they can be used to measure symptoms as defined. Instruments that have content validity are instruments that are arranged based on the material that has been taught (Sugiyono, 2014). To test the validity of the content, it is done by comparing the content of the instrument with the material that has been taught. To test the validity of construction, instruments are constructed about aspects to be measured based on statistical reasoning constructions on descriptive statistical material, and then consulted with mathematics and statistics education experts. The test results of the validity test of the developed test instruments are presented in Table 2 and Table 3.

Table 2: Content and Construct Validity Test Results of Essay Questions

Indicators	Question items	Content Validity					Construct Validity				
		1	2	3	V	Information	1	2	3	V	Information
Describe the data	1	5	4	4	0.83	High	5	5	4	0.92	High
	2	5	4	4	0.83	High	5	5	4	0.92	High
Organize data	3	5	4	4	0.83	High	5	5	4	0.92	High
	4	5	4	4	0.83	High	5	5	4	0.92	High
Represent data	5	4	4	4	0.75	Moderate	5	4	4	0.83	High
	6	4	4	4	0.75	Moderate	5	4	4	0.83	High
	7	4	4	4	0.75	Moderate	5	5	4	0.92	High
Analyze and interpret data	8	4	3	4	0.67	Moderate	4	4	4	0.75	Moderate
	9	4	3	4	0.67	Moderate	4	4	4	0.75	Moderate

Table 3: Content and Construct Validity Test Results of Multiple Choice Questions

Indicators	Question items	Content Validity					Construct Validity				
		1	2	3	V	Information	1	2	3	V	Information
Describe the data	1	5	4	4	0.83	High	5	5	4	0.92	High
	2	5	4	4	0.83	High	5	5	4	0.92	High
	3	5	4	4	0.83	High	5	5	4	0.92	High
Organize data	4	5	4	4	0.83	High	5	5	4	0.92	High
	5	5	4	4	0.83	High	5	5	4	0.92	High
	6	5	4	4	0.83	High	5	5	4	0.92	High
	7	5	4	4	0.83	High	5	5	4	0.92	High
Represent data	8	4	4	4	0.75	Moderate	5	5	4	0.92	High
	9	4	4	4	0.75	Moderate	5	5	4	0.92	High
	10	4	4	4	0.75	Moderate	5	5	4	0.92	High
	11	4	4	4	0.75	Moderate	5	4	4	0.83	High
	12	4	4	4	0.75	Moderate	5	4	4	0.83	High
	13	4	4	4	0.75	Moderate	5	4	4	0.83	High
Analyze and interpret data	14	4	4	4	0.75	Moderate	5	4	4	0.83	High
	15	4	4	4	0.75	Moderate	5	4	4	0.83	High
	16	4	4	4	0.75	Moderate	4	4	4	0.75	Moderate
	17	4	4	4	0.75	Moderate	4	4	4	0.75	Moderate
	18	4	4	4	0.75	Moderate	4	4	4	0.75	Moderate
	19	4	4	4	0.75	Moderate	4	4	4	0.75	Moderate
	20	4	4	4	0.75	Moderate	4	4	4	0.75	Moderate

From Table 2 it can be seen that all essay questions are valid, are in the high and moderate categories. Furthermore, Table 3 is the test result of the validity of multiple-choice questions. From Table 3, it is concluded that all multiple-choice questions are valid with high and moderate categories. Thus, description and multiple-choice questions can be used to measure the statistical reasoning ability of teacher candisate students. There are several inputs given by validators, including related to the allocation of time adjusted to the form of assessment, and the form of the test. The questions should not be too long, should be concise and clear, the numerical data should also not be too large and should be in whole numbers, the context of the questions must also be adjusted to the context of the environment in Indonesia and the cases presented according to the level of students. Input and suggestions from validators, then revised and consulted again. In drawing up an instrument, it is not enough just to determine the topic or material to be assessed, but it is also necessary to determine more specifically what capabilities will be assessed for a particular material. For this reason, a form of description and multiple-choice tests is needed. This is in accordance with the results of the study (Oktaviani, Rahayu, & Sutisna, 2019). The results showed that students who have higher abilities will show better learning outcomes when taking tests with essay forms, while students who have lower abilities will show better learning outcomes when taking tests with multiple choice forms. Questions in the form of descriptions in the form of explanations can provide more valid information about students' statistical reasoning abilities, so that educators can provide interpretations of students' statistical reasoning levels (Wulansari, Putra, Rusliah, & Habibi, 2019).

Conclusion

Test instruments to measure statistical reasoning skills are developed with criteria; a) the suitability of the content domain to be measured based on statistical reasoning indicators, and b) test instruments developed tailored to the context of prospective teacher students, and the Indonesian context. The statistical reasoning skill measurement instrument developed has been valid in content and construct, based on the assessment of the validator instrument. Therefore, it can be used to measure students' statistical reasoning ability on descriptive statistical material.

REFERENCES

Aiken, L. R. (1980). Content Validity and Reliability of Single Items or Questionnaires. *Educational and Psychological Measurement, 40*(4), 955-959.

Annabil, N. R., Wulandari, I., Yamtinah, S., Ariani, S. R., & Ulfa, M. (2022). ANALISIS INDEKS AIKEN UNTUK MENGETAHUI VALIDITAS. *Paedagogia: Jurnal Penelitian Pendidikan, 25*(2), 184-191.

Ariyana, Y., Pudjiastuti, A., Bestary, R., & Zamroni. (2018). *Buku Pegangan Pembelajaran Berorientasi pada Keterampilan Berpikir Tingkat Tinggi.* Jakarta: Dirjen Guru dan Tenaga Kependidikan, Kementerian Pendidikan dan Kebudayaan.

Ben-Zvi, D., & Garfield, J. (2004). *The Challenge of Developing Statistical literacy, Reasoning, and Thinking.* The Netherlands: Kluwer Academic Publishers (Springer).

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.

Chervaney, N., Collier, R., Fienberg, S., Johnson, P., & Neter, J. (1977). A framework for the development of measurement instruments for evaluating the introductory statistics course. *The American Statistician*, 17-23.

delMas , R. C. (2002). Statistical Literacy, Reasoning, and Learning: A Commentary. *Journal of Statistiks Education, 10*(3), 1-11.

Drijvers, P., Kodde-Buitenhuis, H., & Doorman, M. (2019). Assessing mathematical thinking as part of curriculum reform in the Netherlands. *Educational Studies in Mathematics*, 435-456.

Garfield, J. (2003). Assessing Statistical Reasoning. *Statistics Education Research Journal, 2*(1), 22-38.

Garfield, J., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning: Connecting research and teaching practice.* New York: Springer.

Jailani, Sugiman, Retnawati, H., Bukhori, Apino, E., Djidu, H., & Arifin, Z. (2018). *Desain Pembelajaran Matematika Untuk Melatihkan Higher Order Thinking Skill.* Yogyakarta: UNY PRESS.

Jones, G., Thornton, C., Langrall, C., Moone, E., Perry, B., & Putt, I. (2000). A Framework for Characterizing Children's Statistical Thinking. *Mathematical Thinking and Learning, 2*(4), 269-307.

Kemendikbud. (2016). *Permendikbud No. 23 Tahun 2016 tentang Standar Penilaian Pendidikan.* Jakarta: Kementerian Pendidikan dan Kebudayaan.

Kemendikbud. (2017). *Modul Penyusunan Soal Higher Order Thinking Skill (HOTS).* Jakarta: Dirjen Pendidikan Dasar dan Menengah.

Magdalena, I., Fauzi, H. N., & Putri, R. (2020). Pentingnya Evaluasi dalam Pembelajaran dan Akibat Memanipulasinya. *BINTANG, 2*(2), 244-257.

Malawi, I., & Maruti, E. S. (2016). *Evaluasi Pendidikan.* Jawa Timur: CV. AE Media Grafika.

Martadiputra, B. A., & Suryadi, D. (2012). Peningkatan Kemampuan Berpikir Statistis Mahasiswa S1 Melalui Pembelajaran MEAs yang dimodifikasi. *Infinity: Jurnal Ilmiah Prodi Matematika STKIP Siliwangi Bandung, 1*(1), 79-89.

NCTM. (2000). *Principles and Standards for School Mathematics.* USA: The National Council of Teachers of Mathematics, Inc.

Oktaviani, M., Rahayu, W., & Sutisna, A. (2019). Kemampuan Koneksi Matematis Peserta Didik ditinjau dari Bentuk Tes dan Disposisi Matematis. *Jurnal Penelitian dan Pembelajaran Matematika, 12*(2), 213-225.

Sugiyono. (2014). *Metode Penelitian Kombinasi (Mixed Methods).* Bandung: Alfabeta.

Thiagarajan, S., Semmel, D., & Semmel, M. (1974). *Instructional Development for Training Teachers of Exceptional Children.* Minnesota: University of Minnesota.

Wulansari, T., Putra, A., Rusliah, N., & Habibi, M. (2019). Pengaruh Model Pembelajaran Berbasis Masalah pada materi Statistika terhadap Kemampuan Penalaran Statistis Siswa. *AKSIOMA: Jurnal Matematika dan Pendidikan Matematika, 10*(1), 35-47.

Ziegler, L. (2014). RECONCEPTUALIZING STATISTICAL LITERACY: DEVELOPING AN ASSESSMENT. *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS 9)* (pp. 1-6). USA: International Statistical Institute.

****