The New Educational Review

DOI: 10.15804/tner.2024.78.4.09

Hodiyanto
(corresponding author)
Universitas Negeri Surabaya, Indonesia
Mega Teguh Budiarto
Universitas Negeri Surabaya, Indonesia
Rooselyna Ekawati
(corresponding author)
Universitas Negeri Surabaya, Indonesia

Psychometric Evidence of Geometric Ability on Quadrilateral Using Iteman: Test Development Study

Abstract

Quadrilaterals are prerequisites for learning space geometry, so it is necessary to know the geometric abilities of quadrilaterals to learn space geometry successfully. These tests will also help identify students' level of mastery of the quadrilateral concept and provide information for teachers and researchers in planning teaching, learning and research. This research method adapted test development research: test conceptualisation, test construction, test tryout, and item analysis. There were three validators. 120 8th-grade students in West Kalimantan, Indonesia, participated. Item analysis was carried out with Iteman 3.0 software. The geometric ability instrument is valid and reliable because the Scale Content Validity Index (S-CVI) is 1, and the reliable coefficient is 0.757. 21 items have a discrimination index above 0.3 and 3 items below 0.3. The distractors worked fine; only a few distractors did not work fine. Thus, this instrument can be used for geometric ability tests.

Keywords: Discrimination Index, Geometric Ability, Iteman, Test Development, Validity

Introduction

Geometric ability is part of mathematics ability (Xie et al., 2020). Mathematics ability is not single but encompasses all aspects inherent in mathematics, such as reasoning, problem-solving, and communication. Mathematical ability can be understood as acquiring, analyzing, and retaining mathematical information (Karsenty, 2014). Geometric ability is understanding, explaining, and describing geometric phenomena and solving geometric problems creatively (Schoevers et al., 2020). In this research, mathematical ability is linked to geometry, so the term geometric ability (hereafter GA) is used because researchers only use geometry as an object of study. Therefore, GA is defined as the ability to understand the concept of geometry and solve problems related to geometry.
Geometry is an essential aspect of high school mathematics because failure in this aspect can cause one to fail mathematics in general (Adeleke, 2012). Students worldwide experience difficulties in learning geometry (Al-Salahat, 2022) and students in Indonesia also experience such difficulties (Retnawati et al., 2017). One of the geometric concepts is a quadrilateral. Students can recognize the parts of a quadrilateral but need help finding the relationship among the quadrilaterals (Ersoy et al., 2019). Therefore, a good instrument needs to be developed so that teachers have a reliable and effective evaluation tool for measuring GA in quadrilaterals and the relationships among quadrilaterals. By knowing students' GA, teachers can design learning especially suitable for subsequent learning. Good instruments are needed to minimize errors in measurement and assessment (Ramadhan et al., 2020). Thus, this instrument development will be essential in measuring how students master quadrilateral concepts.

The quadrilaterals are the prerequisite for learning space geometry, so knowing the GA of quadrilaterals before studying space geometry is necessary. Based on the results of previous research, students still had difficulty solving space geometry problems (Ismail, 2020). Therefore, this instrument can be a prerequisite for knowing the extent of students' understanding of quadrilaterals before studying space geometry. Prerequisite knowledge is critical in designing curriculum and learning (Scheines & Silver, 2014), and the prerequisite affects material depth (Lestari & Dwi, 2019). Teachers need to know students' GA before they design lessons that require a geometry understanding (Balasa & Mohammed, 2021).

In addition, this research is a recommendation from the research results conducted by Hodiyanto et al. (2024) on the importance of studying abstrac-

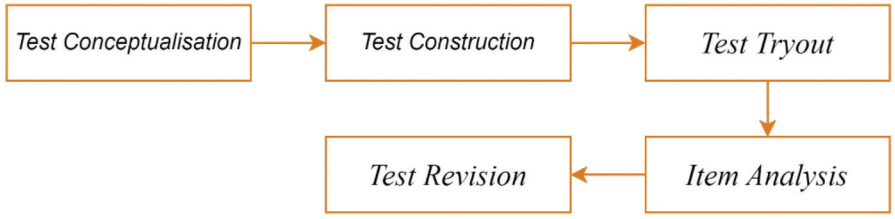
tion in constructing relationships among quadrilaterals based on GA so that it is necessary to develop GA tests on quadrilaterals. Several studies have developed GA in quadrilaterals, but these instruments were based on van Hiele's theory (Baffoe & Mereku, 2011; Decano, 2017). However, the GA instrument development in this research is based on the cognitive domain of Bloom's revised Taxonomy or Anderson and Krathwohl's version. Bloom's taxonomy is a pedagogical tool that can help teachers design learning objectives, effective and student-centered learning, and assess students (Ramirez, 2016). The cognitive domains consist of 6 levels: remembering (C1), understanding (C2), applying (C3), analyzing (C4), evaluating (C5), and creating (C6). C1: Recognizing, recalling or repeating information learned earlier. C2: Construct meaning from various functions, such as interpreting, categorising, summarising, deducing, comparing, or elucidating. C3: using a process and carrying it out through implementation. C4: deconstructing ideas into constituent parts, discerning the relationships among them, exploring their interconnections, and understanding their relevance. C5: Forming judgments using criteria and standards by examining and criticizing. C6: merging components to create a unified or functional entity and rearranging components to create a new pattern (Wilson, 2016).

The validity and reliability of the instrument are also the focus of developing GA. Apart from that, distractor analysis also needs to be considered when preparing multiple-choice items. In addition to calculating difficulty and discriminant indices in analyzing items, instrument developers must also check for distractors and see whether distractors work fine (Finch & French, 2019). The item analysis in the research was not carried out manually but using Iteman 3.0. Iteman is one of the oldest commercial item analysis packages. Iteman's main advantage over similar packages is its ability to handle up to 750 items and no limit on the number of students (Clauser & Hambleton, 2017). Iteman can also report the results of distractor options in multiple-choice. Therefore, this research aims to develop an instrument for the GA of junior high school students, specifically to report the psychometric characteristics and distractors' functions of each item in the instrument's development.

Research Methodology

This research method adapted the instrument development research proposed by Cohen et al. (2022): test conceptualization, test construction, test tryout, item analysis, and test revision. The stages in instrument development can be seen in Figure 1.

Figure 1. Stages of Instrument Development



1. Test conceptualization

Test conceptualization aims to examine the importance of developing GA instruments for junior high school students. Researchers reviewed the literature to produce conceptual and operational definitions of GA. The GA in this research was limited to quadrilaterals, including parallelograms, rectangles, rhombi, squares, kites, and trapezoids. The preparation of the GA tests was based on Anderson and Krathwohl's taxonomy (C1-C3).

2. Test construction

At this stage, the researchers create items and validate them. The number of items was 24. The instrument validators were two mathematics education lecturers and a junior high school mathematics teacher. The selection of validators varies between lecturer and teacher, so the results complement each other. Usually, the lecturer understands more about concepts, and the teacher will help review the instrument to ensure its suitability for the curriculum and students' ability level.

The content validity is first analyzed before the instrument is conducted on participants. This content validity is analyzed by calculating the content validity index (CVI). CVI can be used to assess validity related to an element's precision, clarity, and importance. Analysis of instrument validity with CVI can be done in two ways: (1) Item Content Validity Index (I-CVI), the proportion of agreement regarding the relevance of each item, and (2) Scale Content Validity Index (S-CVI), the content validity of the overall scale or the total proportion of items and a research instrument is considered relevant if it has an I-CVI value of 0.875 to 1 (Nasir et al., 2022).

3. Test tryouts

After the instrument was valid, it was tested on participants. The characteristics of the participants were: 1) they had studied quadrilaterals, 2) they were around 13–14 years old, 3) the school is accredited A, and 4) the school allowed it to be used as a testing place. Next, the test was given to class VIII junior high school, consisting of 120 participants.

4. Item analysis

This analysis includes instrument and item analysis: instrument reliability, item discrimination, difficulty index, and analysis of the distractor function. The reliability test used the Alpha technique because the GA test was multiple-choice. Usually, the value of the reliability coefficient taken is ≥ 0.70. Analysis of items was carried out after testing the instrument on respondents. The GA items were analyzed by looking at the discrimination index, difficulty index, and whether or not the item distractor works fine (Finch & French, 2019). To test the discrimination index (D) of multiple choice items, we can do this by looking for the point biserial correlation coefficient or biserial correlation (Finch & French, 2019). A good discrimination index is D ≥ 0.30 (Demirel & Cetin, 2023). The difficulty level of the items was measured by looking at the item difficulty index (DI), whether they are classified as difficult, medium, or easy. However, the difficulty index in research is generally 0.3 ≥ DI ≥ 0.7 (Demirel & Cetin, 2023).

5. Test Revision

The revision will continue if the item analysis results show that items are revised. In this study, the researchers only carried out the item analysis stage, so this research needs to continue revising and retesting items that could be better. However, the authors only report the item analysis results in this paper.

Results

Before the instrument was tested on research participants, it was first validated on three validators (V). The validation sheet given to validators contains three aspects: material, construction, and language. The validation results for the CVI will be analyzed by looking for the I-CVI and S-CVI values (Table 1 and Table 2).

Table 1. Validation Result Data

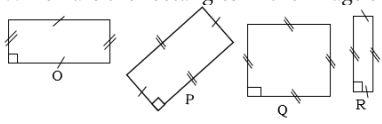
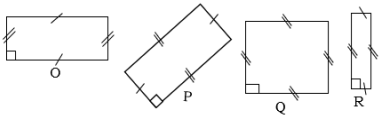
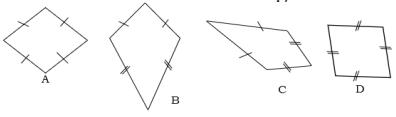
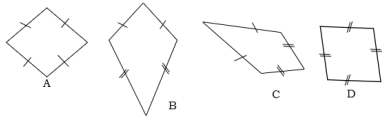
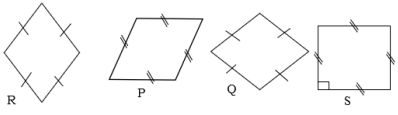
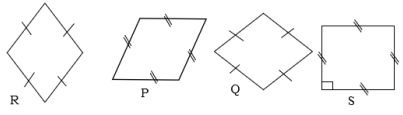
Aspect	V1	V2	V3	I-CVI
Material	1	1	1	1
Construction	0.98	1	1	0.993
Language	1	1	1	1
S-CVI				0.998

Table 2. Item Validation Result Data

Item	V1	V2	V3	I-CVI	Item	V1	V2	V3	I-CVI
1	1	1	1	1	13	1	1	1	1
2	1	1	1	1	14	1	1	1	1
3	1	1	1	1	15	1	1	1	1
4 1 1 1				1	16	1	1	1	1
5	1	1	1	1	17	1	1	1	1
6	1	1	1	1	18	1	1	1	1
7	1	1	1	1	19	1	1	1	1
8	1	1	1	1	20	1	1	1	1
9	0	1	1	0.979	21	1	1	1	1
10	0	1	1	0.979	22	1	1	1	1
11	0	1	1	0.979	23	1	1	1	1
12	1	1	1	1	24	1	1	1	1
S-CVI									1

Table 1 shows that the I-CVI value for the material and language aspects is 1, so these two aspects are relevant to measure the concept. The I-CVI value for the construction aspect is 0.993, which is less than 1 but close to 1, so several components in the construction aspect must be improved. Based on Table 2, items 9, 10, and 11 need to be revised, but not many need to be revised because they are based on the I-CVI values of these items close to 1. Acceptable content validity is attained when the S-CVI of the instrument is ≥ 0.90 and the I-CVI of the items is 1.00 (Kartika et al., 2023). Other items are also revised based on suggestions from the validators even though the value is 1, and it does not change the question's meaning. Several examples of item revisions before and after can be seen in Table 3.

Table 3. Results Before and After the Items are Revised

No	Before	After
6	What is a parallelogram whose diagonals a re perpendicular to each other called? Rhombus Right trapezoid Kite Rectangles	A parallelogram whose diagonals are perpendicular to each other is called a ... Rhombus Kite Rectangle Right trapezoid
9	Which are the rectangles in the image below?  Only O, P, and Q Only O, P, and R Only O, Q, and R All rectangles	The images below, which include rectangles, are ...  Only O, P, and Q Only O, P, and R Only O, Q, and R All rectangles
10	Which are kites from the image below?  Only B Only B and C Only A, B, and C All kites	The images below, which include kites, are ...  Only B Only B and C Only A, B, and C All kites
11	Which are rhombus from the image below?  Only Q Only R and P Only R, P, and Q All rhombuses	The images below, which include rhombus, are ...  Only Q Only R and P Only R, P, and Q All rhombuses

Based on Table 3, items 6 and 10 are only revisions in the language aspect, making them more accessible for students to understand. In items 9 and 11, the researchers revise the language and construction (the answer key). After the instrument was revised and valid from content validity, it was tested on par-

ticipants totaling 120 students. The trial results were analyzed using Iteman 3.0. Figure 2 and Figure 3 are the results of analyzing the items using Iteman 3.0.

Figure 2. Results of the Item Analysis of Using Iteman 3.0

		Item Statistics			Alternative Statistics				
Seq. No.	Scale -Item	Prop. Correct	Biser.	Point Biser.	Alt.	Prop. Endorsing	Biser.	Point Biser.	Key
1	0-1	0.750	0.634	0.465	A	0.750	0.634	0.465	*
					B	0.025	-0.362	-0.136	
					C	0.017	0.200	0.065	
					D	0.200	-0.678	-0.474	
					Other	0.008	0.048	0.012	
2	0-2	0.667	0.412	0.318	A	0.117	-0.338	-0.206	
					B	0.125	-0.405	-0.252	
					C	0.092	-0.002	-0.001	
					D	0.667	0.412	0.318	*
					Other	0.000	-9.000	-9.000	
3	0-3	0.808	0.461	0.319	A	0.808	0.461	0.319	*
					B	0.025	-0.397	-0.149	
					C	0.158	-0.403	-0.267	
					D	0.008	-0.221	-0.055	
					Other	0.000	-9.000	-9.000	
4	0-4	0.750	0.544	0.400	A	0.067	-0.484	-0.251	
					B	0.750	0.544	0.400	*
					C	0.000	-9.000	-9.000	
					D	0.175	-0.409	-0.277	
					Other	0.008	-0.221	-0.055	
5	0-5	0.792	0.643	0.454	A	0.017	-0.586	-0.190	
					B	0.075	-0.305	-0.164	
					C	0.792	0.643	0.454	*
					D	0.108	-0.592	-0.354	
					Other	0.008	-0.311	-0.078	
6	0-6	0.717	0.624	0.469	A	0.717	0.624	0.469	*
					B	0.092	0.035	0.020	
					C	0.175	-0.788	-0.535	
					D	0.017	-0.340	-0.110	
					Other	0.000	-9.000	-9.000	
7	0-7	0.242	0.662	0.482	A	0.075	-0.247	-0.133	
					B	0.617	-0.429	-0.337	
					C	0.058	-0.040	-0.020	
					D	0.242	0.662	0.482	*
					Other	0.008	-0.131	-0.033	
8	0-8	0.600	0.808	0.637	A	0.308	-0.776	-0.592	
					B	0.033	-0.216	-0.089	
					C	0.058	-0.197	-0.098	
					D	0.600	0.808	0.637	*
					Other	0.000	-9.000	-9.000	
9	0-9	0.300	0.733	0.556	A	0.092	-0.360	-0.205	
					B	0.592	-0.504	-0.398	
					C	0.017	0.003	0.001	
					D	0.300	0.733	0.556	*
					Other	0.000	-9.000	-9.000	
10	0-10	0.042	0.861	0.384	A	0.142	-0.480	-0.309	
					B	0.775	0.282	0.203	
					C	0.042	-0.602	-0.268	
					D	0.042	0.861	0.384	*
					Other	0.000	-9.000	-9.000	
11	0-11	0.100	0.248	0.145	A	0.108	-0.220	-0.131	
					B	0.033	-0.024	-0.010	
					C	0.758	-0.003	-0.002	
					D	0.100	0.248	0.145	*
					Other	0.000	-9.000	-9.000	
12	0-12	0.658	0.616	0.477	A	0.192	-0.483	-0.335	
					B	0.083	-0.341	-0.189	
					C	0.658	0.616	0.477	*
					D	0.067	-0.326	-0.169	
					Other	0.000	-9.000	-9.000	

Seq. No.	Scale -Item	Item Statistics: Prop. Correct	Biser.	Point Biser.	Alt.	Alternative Statistics: Prop. Endorsing	Biser.	Point Biser.	Key
13	0-13	0.708	0.381	0.288	A	0.708	0.381	0.288	*
					B	0.108	0.043	0.026	
					C	0.108	-0.373	-0.223	
					D	0.075	-0.492	-0.264	
					Other	0.000	-9.000	-9.000	
14	0-14	0.217	-0.674	-0.480	A	0.008	-0.311	-0.078	
					B	0.092	-0.162	-0.093	
		CHECK THE KEY			C	0.217	-0.674	-0.480	*
		C was specified, D works better			D	0.683	0.650	0.498	?
					Other	0.000	-9.000	-9.000	
15	0-15	0.608	0.556	0.438	A	0.075	-0.132	-0.071	
					B	0.117	-0.452	-0.276	
					C	0.200	-0.379	-0.266	
					D	0.608	0.556	0.438	*
					Other	0.000	-9.000	-9.000	
16	0-16	0.300	0.546	0.414	A	0.300	0.546	0.414	*
					B	0.033	-0.271	-0.112	
					C	0.567	-0.350	-0.278	
					D	0.092	-0.237	-0.135	
					Other	0.008	0.317	0.079	
17	0-17	0.608	0.741	0.583	A	0.158	-0.445	-0.295	
					B	0.608	0.741	0.583	*
					C	0.108	-0.778	-0.465	
					D	0.125	-0.158	-0.098	
					Other	0.000	-9.000	-9.000	
18	0-18	0.575	0.575	0.456	A	0.192	-0.595	-0.412	
					B	0.192	-0.222	-0.154	
					C	0.042	-0.030	-0.014	
					D	0.575	0.575	0.456	*
					Other	0.000	-9.000	-9.000	
19	0-19	0.692	0.447	0.341	A	0.692	0.447	0.341	*
					B	0.108	-0.241	-0.144	
					C	0.108	-0.504	-0.302	
					D	0.092	-0.113	-0.064	
					Other	0.000	-9.000	-9.000	
20	0-20	0.150	0.005	0.003	A	0.542	0.327	0.260	?
					B	0.142	0.155	0.100	
		CHECK THE KEY			C	0.150	0.005	0.003	*
		C was specified, A works better			D	0.158	-0.664	-0.440	
					Other	0.008	-0.221	-0.055	
21	0-21	0.700	0.678	0.514	A	0.017	-0.242	-0.078	
					B	0.092	-0.249	-0.142	
					C	0.700	0.678	0.514	*
					D	0.192	-0.677	-0.469	
					Other	0.000	-9.000	-9.000	
22	0-22	0.242	0.668	0.487	A	0.392	-0.016	-0.012	
					B	0.242	0.668	0.487	*
					C	0.233	-0.180	-0.130	
					D	0.133	-0.685	-0.434	
					Other	0.000	-9.000	-9.000	
23	0-23	0.708	0.696	0.525	A	0.058	0.065	0.032	
					B	0.708	0.696	0.525	*
					C	0.033	-0.024	-0.010	
					D	0.200	-0.874	-0.612	
					Other	0.000	-9.000	-9.000	
24	0-24	0.533	0.605	0.482	A	0.058	-0.232	-0.115	
					B	0.533	0.605	0.482	*
					C	0.342	-0.339	-0.262	
					D	0.067	-0.688	-0.357	
					Other	0.000	-9.000	-9.000	

Figure 3. Statistical Results of Using Iteman 3.0

Scale Statistics

Scale:	0
N of Items	24
N of Examinees	120
Mean	12.467
Variance	16.766
Std. Dev.	4.095
Skew	0.035
Kurtosis	-0.500
Minimum	3.000
Maximum	22.000
Median	13.000
Alpha	**0.757**
SEM	2.019
Mean P	0.519
Mean Item-Tot.	0.382
Mean Biserial	0.520

The reliability coefficient value using alpha obtained a coefficient of 0.757 (Figure 3), classified as reliable. Therefore, the instrument will be reliable or constant when used at different times (Pandian et al., 2023). Based on Figure 2, the option's asterisk (*) indicates the item's answer key. Not all analysis results from the 24 items are discussed in this paper, but only sample items representing the other items are taken. Based on Figure 2 of item 2 obtained: (1) the difficulty index is 0.667, the item is classified as a medium; (2) the discrimination index is D = 0.318 (with point biserial point). The discrimination index of this item is good because it is more than 0.3. A good discrimination index is D ≥ 0.30 D ≥ 0.30 (Demirel & Cetin, 2023); (3) all distractors work fine because 5% of respondents chose them. The distractor does not work fine if no respondent chooses them (Finch & French, 2019). Their discrimination indexes are negative; the distractors are selected more in the lower group than in the upper group. Thus, all the distractors work fine. These results indicate that item 2 is classified as good.

Based on Figure 2, the Iteman results of item 14 are (1) the difficulty index is 0.217; the item is classified as difficult; (2) the discrimination index is not good because it is negative, less than 0.3; (3) respondents choose all the distractors. The discrimination indexes of distractors A and B are negative, but A is less than 5%; (4) Note that the item provides suggestions for the test developer to review answer key D because the discrimination index is high and positive, 0.498. The upper group chooses the distractor more than the lower ability. These results indicate that item 14 is classified as not good, and distractors A and D need to be checked and revised again.

If we look at the discrimination index of each item (Figure 2), it is found that the items with a discrimination index below 0.3 are items 11, 13, 14, and 20. However, the discrimination index of item 13 is greater than 0.3 (the biserial coefficient = 0,381), and the discrimination index of items 11, 14, and 20 is below 0.3 in both the point biserial and biserial coefficients. For all the items, the discrimination index of the biserial coefficient is greater than the point biserial coefficient. Generally, the biserial coefficient is greater than the point biserial coefficient (Finch & French, 2019). So, items 11, 14, and 20 need to be checked and revised again to use them as a GA instrument. Based on the results of these items, items 2, 7, 12, 15, 17, 19, 22, and 24 are good and do not need revision. Items 1, 3, 4, 5, 6, 8, 9, 10, 16, 18, 21, and 23 are good, but several distractors need to be revised because they have a positive discrimination index, or less than 5% of respondents chose them. Items 11, 14, and 20 are not good because they have a discrimination index below 0.3, so revision is necessary.

Discussion and Conclusions

The development of this instrument is essential, especially by mathematics education teachers, to determine students' GA in understanding the concept of quadrilaterals because quadrilaterals are a requirement for studying space geometry. A teacher needs to know students' GA before they design lessons that require an understanding of geometry (Balasa & Mohammed, 2021). Therefore, it is essential to have a valid and reliable test so that teachers can use it to determine the prerequisites for space geometry. The validation result shows the instrument is valid (Tables 1 and 2), and the trial analysis result shows reliability (Figure 3). Thus, the instrument is valid and reliable as an instrument of GA. Some items are not good because they have a discrimination index below 0.3. After all, a good discrimination index is $D > 0.3$ $D \geq 0.3$ (Demirel & Cetin, 2023). This research shows that 21 items can be used as test instruments and 3 items cannot. However, some distractors need to be revised because a distrac-

tor works fine if it has a negative discrimination index and is chosen by 5% of respondents (Testa et al., 2018).

This result provides recommendations to education stakeholders. For teachers, this finding can be a guide in designing learning that considers GA, especially quadrilaterals. This instrument can be used as a formative assessment to help identify students' mastery of the quadrilateral concept and provide information for teachers in planning teaching and learning. By giving it to students, the teacher will know their GA. Then, the teacher can design learning according to the student's GA. Formative assessment positively affects learning and student learning competences (Guptan & Rasiah, 2016).

Future researchers can use this test as a requirement to determine students' GA so that the results can be used as a basis for follow-up research as previous research recommended the development of a GA instrument so that future researchers could study research on the relationships among quadrilaterals based on the level of GA (Hodiyanto et al., 2024). Other researchers considering quadrilaterals' GA can also use this instrument. This research can also be adapted by further research in similar research and instrument development. Other researchers can continue this research to level C6 because it was only at levels C1 to C3.

Acknowledgements

We thank Balai Pembiayaan Pendidikan Tinggi (BPPT) and Lembaga Pengelola Dana Pendidikan (LPDP) for providing the Indonesian Education Scholarship 2021.

References:

Adeleke, J. (2012). Predictive Study of Prerequisite Strength, Mathematics Aptitude and Cognitive Achievement in Geometry. *EDULEARN12 Proceedings*, 5759–5765.

Al-Salahat, M. M. S. (2022). The Effects of Learning Strategies on Senior High School Students' Motivation and Learning Outcomes of Overhead Passing in Volleyball. *International Journal of Education in Mathematics, Science and Technology*, *10*(2), 458–476.

Baffoe, E., & Mereku, D. (2011). The van Hiele levels of understanding of students entering senior high school in Ghana. *African Journal of Educational Studies in Mathematics and Sciences*, *8*(1), 51–62.

Balasa, M. M., & Mohammed, T. (2021). Influence of Van Hiele Phase Descriptors on students' Development of Levels of Geometric Understanding and Cognitive Achievement in Geometry: Focus on Senior Public Secondary Schools in Taraba State. *International Journal of Pure and Applied Science*, *21*(9).

Clauser, J. C., & Hambleton, R. K. (2017). Item analysis for classroom assessments in higher education. In *Handbook on measurement, assessment, and evaluation in higher education* (pp. 355–369). Routledge.

Cohen, R. J., Schneider, W. J., & Tobin, R. M. (2022). *Psychological Testing and Assessment An Introduction to Tests and Measurement (Tenth Edition)*. The McGraw-Hill Companies.

Decano, R. S. (2017). Cognitive Development of College Students and their Achievement in Geometry: An Evaluation using Piaget's Theory and Van Hiele's Levels of Thinking. *American Journal of Applied Sciences*, *14*(9), 899–911.

Demirel, S., & Cetin, H. (2023). A Test Development Study on Spatial Visualization for Second-Grade Primary School Students. *International Electronic Journal of Elementary Education*, *16*(1), 89–98.

Ersoy, M., İlhan, O. A., & Sevgi, S. (2019). Analysis of the Relationship between Quadrilaterals Achievement Levels and Van Hiele Geometric Thinking Levels of the Seventh Grade Students. *Higher Education Studies*, *9*(3), 1.

Finch, W. H., & French, B. F. (2019). *Educational and psychological measurement*. Routledge.

Guptan, V., & Rasiah, R. (2016). Competence development in business undergraduates: The role of formative assessment and humour. *The New Educational Review*, *45*, 160–171. https://doi.org/10.15804/tner.2016.45.3.13.

Hodiyanto, Budiarto, M. T., Ekawati, R., Susanti, G., Jeonghyeon, K., & Bonyah, E. (2024). How abstraction of a pre-service teacher in constructing relationships among quadrilaterals. *Journal on Mathematics Education*, *15*(2), 337–360.

Ismail, H. (2020). Investigating Student's Learning Difficulties in Shape and Space Topic: A Case Study. *International Journal of Psychosocial Rehabilitation*, *24*(5), 5315–5321.

Karsenty, R. (2014). Mathematical Ability. In *Encyclopedia of Mathematics Education* (pp. 372–375). Springer Netherlands. https://doi.org/10.1007/978-94-007-4978-8_94.

Kartika, H., Budiarto, M. T., Fuad, Y., Jeonghyeon, K., & Bonyah, E. (2023). Developing and Reporting Psychometric Evidence of Prerequisite Algebra Skills Instrument. *Acta Scientiae*, *25*(4), 133–162.

Lestari, N., & Dwi, J. (2019). The role of prospective mathematics teachers' knowledge of content and students in integrating mathematical literacy. *The New Educational Review*, *57*(3), 151–160. https://doi.org/10.15804/tner.19.57.3.12.

Nasir, N. A. M., Singh, P., Narayanan, G., Habali, A. H. M., & Rasid, N. S. (2022). Development of Mathematical Thinking Test: Content Validity Process. *ESTEEM Journal of Social Sciences and Humanities*, *6*(2), 18–29.

Pandian, V., Awang, M., Ishak, R., & Kok Ming, G. (2023). Validity and Reliability of Organizational Trust Instrument. *International Journal of Academic Research in Progressive Education and Development*, *12*(2). https://doi.org/10.6007/ijarped/v12-i2/16564.

Ramadhan, S., Sumiharsono, R., Mardapi, D., & Prasetyo, Z. K. (2020). The quality of test instruments constructed by teachers in bima regency, Indonesia: Docu-

ment analysis. *International Journal of Instruction*, *13*(2), 507–518. https://doi.org/10.29333/iji.2020.13235a.

Ramirez, T. V. (2016). On Pedagogy of Personality Assessment: Application of Bloom's Taxonomy of Educational Objectives. *Journal of Personality Assessment*, *99*(2), 146––152. https://doi.org/10.1080/00223891.2016.1167059.

Retnawati, H., Arlinwibowo, J., & Sulistyaningsih, E. (2017). The Students' Difficulties In Completing Geometry Items Of National Examination. *International Journal on New Trends in Education and Their Implications*, *8*(4), 1–14.

Scheines, R., & Silver, E. (2014). Discovering Prerequisite Relationships among Knowledge Components. *Proceedings of the 7th International Conference on Educational Data Mining (EDM)*, 355–356.

Schoevers, E. M., Leseman, P. P. M., & Kroesbergen, E. H. (2020). Enriching Mathematics Education with Visual Arts: Effects on Elementary School Students' Ability in Geometry and Visual Arts. *International Journal of Science and Mathematics Education*, *18*(8), 1613–1634. https://doi.org/10.1007/s10763-019-10018-z.

Testa, S., Toscano, A., & Rosato, R. (2018). Distractor Efficiency in an Item Pool for a Statistics Classroom Exam: Assessing Its Relation With Item Cognitive Level Classified According to Bloom's Taxonomy. *Frontiers in Psychology*, *9*, 1585. https://doi.org/10.3389/fpsyg.2018.01585.

Wilson, L. O. (2016). Anderson and Krathwohl–Bloom's taxonomy revised. *Understanding the New Version of Bloom's Taxonomy*.

Xie, F., Zhang, L., Chen, X., & Xin, Z. (2020). Is Spatial Ability Related to Mathematical Ability: A Meta-analysis. *Educational Psychology Review*, *32*(1), 113–155. https://doi.org/10.1007/s10648-019-09496-y.

AUTHORS

HODIYANTO

He is a doctoral student in mathematics education at the Universitas Negeri Surabaya, Indonesia.
E-mail: hodiyanto.21014@mhs.unesa.ac.id
Scopus ID: 57222724552
ORCID: https://orcid.org/0000-0001-9371-1845

MEGA TEGUH BUDIARTO

He is a Professor of Mathematics Education at the Universitas Negeri Surabaya, Indonesia.
E-mail: megatbudiarto@unesa.ac.id
Scopus ID: 57200364500
ORCID: https://orcid.org/0000-0002-9052-3141

ROOSELYNA EKAWATI
She is a Professor of Mathematics Education at the Universitas Negeri Surabaya, Indonesia.
E-mail: rooselynaekawati@unesa.ac.id
Scopus ID: 57210812993
ORCID: https://orcid.org/0000-0002-2646-9064