

## Designing the English Argumentative Essay Writing Test Based on Critical Thinking Skills

\*<sup>1</sup>Moh. Yamin, <sup>2</sup>Slamet Setiawan, <sup>3</sup>Syafi'ul Anam

<sup>1</sup>Universitas Lambung Mangkurat, Indonesia

<sup>2,3</sup>Universitas Negeri Surabaya, Indonesia

**\*Correspondence:**

moh\_yamin@ulm.ac.id

**Submission History:**

Submitted: March 02, 2023

Revised: April 27, 2023

Accepted: April 28, 2023



This article is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

### Abstract

Designing the argumentative essay is the step to producing the writing test instrument used for measuring the student's performance in writing-based critical thinking skills. This study aimed at measuring the validity and reliability of the argumentative essay writing instrument test with critical thinking skills. The procedure to create a credible writing test instrument was conducted by testing the instrument to be valid and reliable. The process of validating is through content validity and expert validation. The process of reliable instrument design was done by checking the score of the students tested in trying out. The result of validating and making the instrument reliable was stated as effective for use to get the data collection. Either the pretest or posttest of the instrument in each component is stated as reliable, and it affirms the instrument's reliability. Notes from the validator and the proof of the statistics indicate the instrument can be used for collecting the data. As a result, the writing test instrument was ready to use for the research. It is suggested that this instrument can be used for other relevant issues and studied more deeply for the next relevant studies.

**Keywords:** Argumentative essay, writing test instrument, critical thinking skills, valid and reliable

### INTRODUCTION

Assessing pupils' writing success is essential to identify their writing performance. The student's performance can be called the result of their long process of writing activities (Karlen, 2017; Toba et al., 2019; Viñas, 2022). It is also the main part that finding out the student's writing ability aims at mapping the students' performance in their writing ability (Asrobi & Prasetyaningrum, 2017; Tavakoli et al., 2018). Because the students' performances are assumed to be different, the instrument test should consider the variety of the student's input. The instrument applied should be tested so that the test does not bias. It is important to say that measuring the students' performance in writing is the way to see how much they

can write and organize their ideas into organized writing. Each paragraph should consist of the logic of writing rationally and systematically based on the data provided.

The argumentative essay writing test needs to involve the characteristics of what to test. The writing tests should be measured to test and determine the student's performance in their writing of argumentative essays (Khunaifi, 2015). Therefore, the characteristics of an argumentative essay need to be inserted into the writing instrument. There are parts that should be highlighted, and those are a thesis, containing an attention grabber, background information, and the writer's position; an argument, containing paragraphs that support and provide evidence for a proposition; and a conclusion, reaffirming the writer's position (Schneer, 2014; Rusfandi, 2015). Besides, critical thinking skills components should also be part of designing the writing test. Critical thinking skills functioned to measure and find out how far the students use high-order thinking skills in responding to the issue, elaborating on the details of the issue, discussing, interpreting, and giving judgment towards the analysis (Khunaifi, 2015; Afshar et al., 2017; Nejmaoui, 2018; Zaini et al., 2018; Osman et al., 2019; Tomak, 2021; Wu, 2021; Malik, 2022).

The indicators of writing tests available are when all supporting components are covered. The writing test is meaningful, while the students' competence is accessed and measured. Based on the content validity, the writing test has fulfilled the requirement of the instrument test by matching the test items with the relevant test specifications and lesson objectives (Widyaningrum & Prabandari, 2013). Klebanov et al. (2019) stated that the validity of the writing instrument is effective in measuring the success of the instrument in practice. By having well-established validity, it was possible to identify occupational skills among students. The validity tested yields increasing trust in providing valid instruments for the prospective replication of research and empirical studies (Suhaini et al., 2021). Besides that, it is important to check the reliability of the writing instrument. It is the effort to determine whether the instrument test is reliable (Kehoe, 2012; Karadere et al., 2013; Bilge Uzun et al., 2019). The result of the test determines the sustainability of the writing test practically. Validity is the most fundamental consideration in instrument development and should be the main consideration in practice. Therefore, validity always refers to whether an instrument measures what it is designed to measure (Ployhart, 2012; Abdullah et al., 2022).

Arranging the writing instrument test to determine the student's performance in their writing achievement needs to be conducted based on the study's goal. It is important to say that the instrument reflects what to predict when it is tested on the samples. Each writing instrument test has its objective. Because writing is a productive skill, it is necessary to design a writing test to measure productive skills in writing performance (Chan, 2013). It is important to design and insert the relevant and basic components dealing with the writing test examined.

The writing instrument should be able to show in detail to make all relevant elements measured. In addition, critical thinking skills are the other components such as inference, evaluating arguments, deduction, recognizing assumptions, and interpretation, Orhan & ÇEVİKER AY, (2022) inserted in the writing test to obtain the students' knowledge in practice in using critical thinking skills in writing the argumentative essay. To arrange the critical thinking skills in the writing instrument test should be well-ordered based on order and priority. Every indicator in the critical thinking skills being displayed in the writing instrument shows the maximal achievement of the student's performance in their writing

skills. Therefore, designing the writing instrument needs accuracy in putting each element in the writing test. It is very important to say that designing a writing test should consider the indicators of components to get scoring.

Validity and reliability are the two aspects that are needed in designing the writing test instrument. The validity of the writing test determines the validity of the writing instrument. Therefore, it is important to say that the validity of the writing instrument is the indicator of how valid the writing instrument will be used in testing the students' performance in their writing achievement (Putka & Sackett, 2010, Orhan & ÇEVİKER AY, 2022, Gülay & Ungan, 2022; Demir et al. 2022). The important thing is that the validity of the writing instrument drives the data collection process so that finding out the students' performance is easier. Testing the reliability of the writing instrument is the other aspect that should be conducted to get a consistent measurement from one time to another time in a different moment (Putka & Sackett, 2010; Heale & Twycross, 2015; Kubai, 2019, Orhan & ÇEVİKER AY, 2022; Maryani et al., 2022; Viñas, 2022). The goal is to ensure that the instrument in writing test is effective in measuring the students' writing achievement performance (Pakirnathan & Kepol, 2018). Heale & Twycross (2015) stated that validity accurately measures all aspects that should be measured from content, construct, and criterion. Based on these considerations, this study aimed to measure the validity and reliability of the argumentative essay writing instrument test. It is expected that the writing instrument measured can drive the student's research to produce a writing product that is valid and reliable and can map the students' skills in writing.

## **METHOD**

The research method used in this study was mixed because the instrument should be tested through statistical and non-statistic tests. The statistic test was for testing the instrument through ANOVA, and the non-statistic test was descriptive through test expert validation of whether the instrument was valid for conducting the research. Therefore, the approach used in this study was quantitative and qualitative. The statistical results aimed to describe the writing test instrument as valid and reliable. Therefore, it was important to state the statistical result in the narration. Based on the qualitative approach, the data were obtained from the questionnaire used to measure the readability and practicality of the instrument from the lecturer teaching the writing class and analyzed through content analysis (Krippendorff, 2004).

The sample of this study was three (3) students from the classes from the population of the writing classes. The students were subject to trying out the writing test instrument. In the data collection, the research started by distributing the instrument to the three students. They were asked to write an argumentative essay based on the issues or topics provided. After writing, they submitted their writing tasks to the researcher to get a scoring based on the scoring rubric provided. Besides that, the writing test instrument was also submitted to the lecturer to get feedback on whether it was effective and available as the writing test instrument. The lecturer gave notes on what to do the next time before the instrument was used to collect the data.

## FINDING AND DISCUSSION

In collecting the data, two writing tests were given to the students. Each writing test has its aim. The first one was to check the students' homogeneity either in the control or experimental group conducted as a pretest. The next one was to find out the student's writing ability after getting treatment as a posttest. The function of the posttest is to test hypotheses. The consideration in constructing the writing test in this study was based on the college's syllabus. The students were required to write four essays based on the topic or issues provided. Each essay has its aims. The first essay was for the pretest, the second and the third were for treatment, and the last essay was posttest. All essays the students did were in the instruction for argumentative essays. The writing test prompt was tested on three students in which the research or study was done. The three students were one male and two female students, although they were possibly not the subject of the research. Principally, they were in the average ability representing all populations of this study. In conducting the study, the students wrote and narrated their opinions dealing with the writing test prompt clearance, and the time allocation was controlled. The steps of designing the writing test instrument were done, starting from developing the blueprint of the tests, writing the tests, validating the tests, revising the writing tests, trying out the tests, analyzing the try-out of the test, and assembling the final form.

As the effort in designing the writing test instrument, the blueprint was done, and it was based on the identification of the syllabus starting from the dimensions including subject matter, level, stimulus, genre, rhetorical task, the pattern of exposition, specification of the audience, role, tone, length, time allotment, transcription mode, scoring criteria, the objective of the test, specification of objectives, and weighting. Dealing with the subject matter in the blueprint, the student's preferences, and familiar topics were considered as an attempt to present a reliable instrument.

After the blueprint was finished being developed, it was continued by writing the writing tests. It was conducted to measure the students' writing ability. The two writing tests were for the pretest and posttest. In this test, the students wrote an argumentative essay based on the issues provided, and they chose. It tried to picture the students' writing abilities. After writing and submitting, their writing products were evaluated and got scores based on the scoring rubric provided.

**Table 1.** Scoring rubric

Aspect	CRITERIA	Grade	Score
<b>INTRODUCTION</b> (Background History/Thesis Statement)	A well-developed introduction with an attention grabber grabs the reader's interest and continues to engage the reader until the thesis statement. The thesis statement in identifying the problem should clearly state the experience or event that will be described as well as the effect on the writer. A conclusion should effectively wrap up and restresses the importance of the thesis.	4	16-20
	The introduction creates interest. The thesis, in identifying the problem, states the position. The conclusion effectively summarizes the topic.	3	11-15
	The introduction adequately explains the background but may lack detail. The thesis in identifying the problem states the topic, but key elements are missing.	2	6-10

	Background details are a random, unclear collection of information. The thesis in identifying the problem is vague and unclear. The conclusion is not effective and does not summarize the main points.	1	1-5
<b>MAIN POINTS</b> (Body Paragraphs)	Well-developed main points/topic sentences relate directly to the thesis. Supporting examples are concrete and detailed. Distinguishing the main idea from subordinate ideas appears in the text. The ability to use the other references is strongly expressed. The analysis is developed with an effective point of view.	4	16-20
	Three or more main points relate to the thesis, but some may lack details. The analysis shows events from the author's point of view but could use more descriptive language. The ability to use the other references appears but is not strongly expressed.	3	11-15
	Three or more main points are present but lack details in describing the event, analysis, evaluation, and synthesis. The ability to use the other references does not work well. Little descriptive language is used.	2	6-10
	Less than three ideas/main points are explained and/or they are poorly developed. The story tells; it does not show.	1	1-5
<b>CRITICAL THINKING SKILLS</b>	The thesis statement starts by identifying strong research problems; differentiating the main and supporting ideas; proposing arguments with analysis and synthesis in the body of the text; paraphrasing references sources to strengthen the arguments; evaluating the evidence; making reflections and generating new insights dealing with the issue in conclusion.	4	23-30
	The thesis statement starts by identifying research problems; differentiating the main and supporting ideas but lack of proposing arguments with analysis and synthesis in the body of the text; paraphrasing references sources to strengthen the arguments; evaluating the evidence; making reflections and generating new insights dealing with the issue in conclusion.	3	16-22
	The thesis statement starts by identifying weak research problems but lack of differentiating the main and supporting ideas but lack of proposing arguments with analysis and synthesis in the body of the text; paraphrasing references sources to strengthen the arguments; evaluating the evidence; making reflections, and generating new insights dealing with the issue in conclusion.	2	8-15
	The story appears poor in identifying research problems; differentiating the main and supporting ideas; proposing arguments with analysis and synthesis in the body of the text; paraphrasing references and sources to strengthen the arguments; evaluating the evidence; making reflections and generating new insights dealing with the issue in conclusion.	1	1-7
<b>ORGANIZATION</b> (Structure and Transitions)	The logical progression of ideas with a clear structure enhances the thesis. Transitions are effective and vary throughout the paragraph, not just in the topic sentences.	4	9-10
	The logical progression of ideas exists. Transitions are present throughout the essay but lack variety.	3	6-8
	The organization is clear. Transitions are present at times, but there is very little variety.	2	3-5
	Writing is not organized. The transitions between ideas are unclear or non-existent.	1	1-2
<b>STYLE</b> (Sentence Flow, Variety, Diction)	Writing is smooth, skillful, and coherent. Sentences are strong and expressive with varied structures. Diction is consistent and words are well chosen.	4	9-10
	Writing is clear and sentences have varied structures; diction is consistent.	3	6-8

	The writing is clear but could use a little more sentence variety to make the writing more interesting.	2	3-5
	Writing is confusing and hard to follow. It contains fragments and/or run-on sentences.	1	1-2
<b>MECHANICS</b> (Spelling, Punctuation, Capitalization)	Punctuation, spelling, and capitalization are all correct. No errors.	4	9-10
	Punctuation, spelling, and capitalization are generally correct with few errors (1-2).	3	6-8
	There are only a few (3-4) errors in punctuation, spelling, and capitalization.	2	3-5
	Distracting errors in punctuation, spelling, and capitalization.	1	1-2

**Grand Total : 100**

Source: adapted from Brown (2004) and Mat Daud (2012)

To produce the credible instrument, it is important to make the writing instrument valid and reliable. It is the core of the instrument that should be maintained to guarantee the trustworthiness of the data. The step that should be done is to prove the theoretical evidence free from problems dealing with the construct and content validity. Seen from the construct validity evidence, the writing ability results were taken from the prompt the students wrote. It got the score with the scoring rubric provided. Whereas viewed from the content validity, it was free from content validity problems in which the results of the writing tests represented all aspects being measured.

Driving empirical evidence of being free problem related to criterion-related validity was the other effort that was conducted. It aimed to make a match between the scores in this study and the actual performance taken from Advanced Writing. The writing tests the students did was the component to test students' writing ability, ensuring high validity, and integrating all elements of writing. In order to make sure the reliability of the writing test instrument, this effort was also done. Commonly, the inter-rater reliability problem came when there were two raters' scores in producing the writing test results inconsistent. It happened because of being lack of attention to scoring criteria, inexperience, inattention, or even preconceived biased. It is important to make the raters have the same perception about the scoring rubric. Besides that, ANOVA Hoyt in SPSS 20.0 program was used to find out the inter-rater reliability of writing ability. Here are the results of the consistency test of students' activity assessment conducted by two raters using Hoyt's ANOVA consistency test or inter-rater reliability on writing ability.

**Table 2.** Inter-Rater Reliability Testing Results

	<b>MS Error</b>	<b>MS Rater</b>	<b>Reliability</b>	<b>Note</b>
<b>Pretest</b>				
Introduction	2.167	20.167	0.893	Reliable
Main Points	1.167	10.667	0.891	Reliable
Critical Thinking Skills	2.667	10.667	0.750	Reliable
Organization	0.167	0.667	0.750	Reliable
Style	1.500	13.500	0.889	Reliable
Mechanics	0.167	0.667	0.750	Reliable
Total Score	12.667	66.667	0.810	Reliable
<b>Posttest</b>				

Introduction	0.500	1.500	0.667	Reliable
Main Points	0.167	4.167	0.960	Reliable
Critical Thinking Skills	4.167	20.167	0.793	Reliable
Organization	0.167	2.667	0.938	Reliable
Style	0.667	8.167	0.918	Reliable
Mechanics	1.167	4.167	0.720	Reliable
Total Score	2.167	204.167	0.989	Reliable

The consistency results based on the writing ability assessment from the pretest and posttest of the writing prompt contained six aspects covering introduction, main points, critical thinking skills, organization, style, and mechanics. The result is a reliability value of more than 0.600 (Reliability > 0.600), so the assessment between the two raters was stated as “consistent.” It indicates that there was no inter-rater reliability problem. Therefore, either the pretest or posttest instrument was stated as reliable to use for collecting the data. To ensure the validation of the writing test instrument, the process of validating the instrument either for pretest or posttest was conducted and done by the lecturer teaching the subject. The indicators used in validating are dealing with the test appropriateness towards students’ level, the essay length, the objective of the tests, the instructions of the test, and scoring rubrics. The validation result means that all are in line with the students’ selected topics, and the rubric responded to the argumentative essay characteristics (Gülay & Urgan, 2022).

After the process of validating the instrument was finished, it was followed by revising the instrument based on the responses given by the expert. The test of the instrument students was also conducted. The result of the writing prompt test stated the readability of the writing test instruction, familiarity with the text genre, time allocation, and student’s knowledge of the topics. After this process was over, assembling the writing test instrument into final form as the instrument test for data collection was constructed and the instrument was ready for use.

Arranging the writing test instrument to answer the research problems proposed should be made as keen as possible, as careful as possible, as valid as possible, and as reliable as possible. The writing test instrument constructed based on the indicators prepared determined the data collection and analysis path. Therefore, it is important to consider the importance of validity and reliability in making the writing test instrument. Kural (2018) says that considering validity and reliability in the assessment of writing is the effort to produce a writing test instrument valid and reliable. Park (2005) said that the issue of establishing the reliability of scoring among pools of raters should be conducted as the step to embody the credible instrument for use in theory and practice. Besides that, in a writing test, this issue becomes the greatest concern because of depending on human interpretation in rating students’ compositions. The quality of the writing test instrument tested through validity and reliability is promising so that the data collection and analysis run well.

Dobrić (2019), in line with Kural and Park, also states that validity relies on the degree of what to measure. It is the quality of the inferences that should be made. Therefore, validation is the process in which we gather and evaluate evidence to support the appropriateness, meaningfulness, and usefulness of the inferences and decisions that should be made based on the measurement scores provided. Dealing with the writing test instrument made to measure the students’ performance in writing the argumentative essay

through critical thinking skills, validating the instrument was made in the content of validity (Dobrić, 2019) because it is the main part of producing the credible instrument for use. Content validity deals with the domain specifications underlying the test. It is important to say that the writing test instrument should consider the indicators of argumentative essays and critical thinking skills as the materials the students should write. The students' performance in their writing products is seen in how they write the argumentative essay with critical thinking skills (Schneer, 2014; Rusfandi, 2015; Pei et al., 2017; Widyastuti, 2018; Nejmaoui, 2018).

To make the writing test instrument credible, the instrument should be judged reliable. To judge reliability, it is important to make the score result consistent with measurement over time or stability of measurement over a variety of conditions. This technique to estimate reliability is with a measure of association, the correlation coefficient often termed the reliability coefficient. The reliability coefficient is the correlation between two or more variables (here tests, items, or raters) measuring the same thing (Drost, 2011; Zohrabi, 2013; Heale & Twycross, 2015; Kubai, 2019). It is also stated by Putka & Sackett (2010) that reliability refers to the precision of a particular measurement. Since reliability is defined as the precision of a measurement instrument, the reliability of an instrument is evaluated by determining the variance of the measured variable that is due to true differences between the observational units and the variance that occurred due to measurement error.

It is important to say that designing the writing test instrument should be seen from effectiveness and efficiency. This condition drives the practicality of the writing test instrument for data collection. The attendance of validity and reliability test to the writing test instrument ensures that the test is valid and reliable to use (Putka & Sackett, 2010; Zohrabi, 2013; Noble, 2019). Therefore, it is necessary to state that the result of the validity and reliability test determines if the instrument is available to use or not (Viñas, 2022). The writing test instrument tested in this study is the effort to produce an instrument credible and ready for use. In addition, as seen from the lecture's responses dealing with the instrument, the writing instrument is available based on the goal to use (Cope, 2010; Connor et al., 2011; Belotto, 2018). It is effective because the writing instrument represents the students' skills in writing critical thinking writing of argumentative essay. Therefore, the existence of the writing instrument seen either from a quantitative or qualitative aspect can measure the students' performance in writing argumentative essay based on critical thinking skills. In a word, the writing instrument is ready to use and give comprehensive and holistical to portrait the students' ability to write their argumentative essay with critical thinking skills.

## **CONCLUSION**

The writing test instrument for argumentative essay with critical thinking skills was viewed as the step to sharpen and open the students' knowledge dealing with the issues happening. By having such an instrument, it was an effort to produce the writing test credible and accountable for data collection. The writing test with critical thinking skills was meant to dig the students' performance in better writing skills in the argumentative essay because the power of the argumentative essay was based on the ability to give interpretation, analysis, synthesis, evaluation, and generating new perspectives (Mat Daud, 2012) to the

issue proposed. By writing the argumentative essay framework with critical thinking (Sabu & Vernandes, 2019), it moves the students to be able to think critically. As an effort to achieve this goal, testing the writing instrument should be done to detect whether the instrument for use in representing the indicators achieved is seen from the characteristics of argumentative essay and critical thinking skills for those learning in the English Department. The result showed validity and reliability, either seen from a quantitative or qualitative approach for testing the instrument test. It is suggested to use data collection dealing with academic writing and other genres of essays relevant to developing the student's critical thinking.

## REFERENCES

- Abdullah, M. N. L. Y., Omar, A. F., Ping, T. A., & Chun, T. C. (2022). Validity and Reliability of the Postgraduate Self-Determined Learning Questionnaire (PSLQ). *Asian Journal of University Education*, 18(1), 91–105. <https://doi.org/10.24191/ajue.v18i1.17175>
- Afshar, H. S., Movassagh, H., & Arbabi, H. R. (2017). The interrelationship among critical thinking, writing an argumentative essay in an L2 and their subskills. *Language Learning Journal*, 45(4), 419–433. <https://doi.org/10.1080/09571736.2017.1320420>
- Asrobi, M., & Prasetyaningrum, A. (2017). Trait Based Assessment on Teaching Writing Skill for EFL Learners. *English Language Teaching*, 10(11), 199. <https://doi.org/10.5539/elt.v10n11p199>
- Beigman Klebanov, B., Ramineni, C., Kaufer, D., Yeoh, P., & Ishizaki, S. (2019). Advancing the validity argument for standardized writing tests using quantitative rhetorical analysis. *Language Testing*, 36(1), 125–144. <https://doi.org/10.1177/0265532217740752>
- Belotto, M. J. (2018). Data analysis methods for qualitative research: Managing the challenges of coding, interrater reliability, and thematic analysis. *Qualitative Report*, 23(11), 2622–2633.
- Bilge Uzun, N., Alici, D., & Aktas, M. (2019). Reliability of the analytic rubric and checklist for the assessment of story writing skills: G and decision study in generalizability theory. *European Journal of Educational Research*, 8(1), 169–180. <https://doi.org/10.12973/eujer.8.1.169>
- Brown, H. D. (2004). *Language Assesment: Principles and Classroom Practices*. Pearson Education, Inc.
- Chan, S. H. C. (2013). *Establishing The Validity of Reading-Into-Writing Test Tasks for the UK Academic Context*.
- Connor, U., Belotto, M. J., Bauch, G., Berrou, C., Declercq, D., Graell I Amat, A., Ould-Cheikh-Mouhamedou, Y., Saouter, Y., Sayir, J., Tavares, M. B. S., Walker, D., Dobson, K., & G.Burgess, A. B. and R. (2011). Reflections on qualitative data analysis. *Analyzing Qualitative Data*, 1(1), 247. <https://doi.org/10.1002/9780470975220.ch1>
- Cope, M. (2010). Interpreting and Communicating Qualitative. In *Coding qualitative data* (Issue January).
- Demir Çelebi, Ç., & Yüksel, M. (2022). Successful Aging Scale: Validity and Reliability Study. *International Journal of Psychology and Educational Studies*, 9(1), 79–90. <https://doi.org/10.52380/ijpes.2022.9.1.521>
- Dobrić, N. (2019). Reliability, validity, and writing assessment: A timeline. *ELOPE: English Language Overseas Perspectives and Enquiries*, 15(2), 9–24. <https://doi.org/10.4312/ELOPE.15.2.9-24>

- Drost, E. A. (2011). Validity and Reliability in Social Science Research. *Education Research and Perspectives*, 38(1), 105–123.
- Gülay, E., & Urgan, S. (2022). Development of academic writing block scale (Awbs): A validity and reliability study. *Participatory Educational Research*, 9(2), 178–198. <https://doi.org/10.17275/per.22.35.9.2>
- Heale, R., & Twycross, A. (2015). Validity and reliability in quantitative studies. *Evidence-Based Nursing*, 18(3), 66–67. <https://doi.org/10.1136/eb-2015-102129>
- Karadere, M. E., Yavuz, K. F., Kayran, İ. E., & Şafak, Y. (2013). Reasoning with Inductive Argument Test: A Study of Validity and Reliability. *Journal of Cognitive Behavioral Psychotherapy and Research (JCBPR)*, 2, 156–161.
- Karlen, Y. (2017). The development of a new instrument to assess metacognitive strategy knowledge about academic writing and its relation to self-regulated writing and writing performance. *Journal of Writing Research*, 9(1), 61–86. <https://doi.org/10.17239/jowr-2017.09.01.03>
- Kehoe, J. F. (2012). What to Make of Content Validity Evidence for Cognitive Tests? Comments on Schmidt (2012). *International Journal of Selection and Assessment*, 20(1), 14–18. <https://doi.org/10.1111/j.1468-2389.2012.00574.x>
- Khunaifi, A. R. (2015). The effects of teaching critical thinking on students' argumentative essay. *Journal on English as a Foreign Language*, 5(1), 45. <https://doi.org/10.23971/jefl.v5i1.91>
- Krippendorff, K. (2004). Content Analysis: An Introduction to Its Methodology. In *Organizational Research Methods* (2nd ed.). SAGE Publications, Inc. <https://login.proxy.libraries.rutgers.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=buh&AN=48779086&site=ehost-live>
- Kubai, E. (2019). *Reliability and Validity of Research Instruments*.
- Kural, F. (2018). Does indirect writing assessment have any relevance to direct writing assessment? Focus on validity and reliability. *Journal of Language and Linguistic Studies*, 14(4), 342–351. [www.jlls.org](http://www.jlls.org)
- Malik, M. M. (2022). *What Is Critical Thinking?* (Issue January). <https://doi.org/10.13140/RG.2.2.35191.96165>
- Maryani, I., Prasetyo, Z. K., Wilujeng, I., & Purwanti, S. (2022). Higher-order Thinking Test of Science for College Students Using Multidimensional Item Response Theory Analysis. *Pegem Egitim ve Ogretim Dergisi*, 12(1), 292–300. <https://doi.org/10.47750/pegegog.12.01.30>
- Mat Daud, Nor Shidrah Binti. (2012). *Developing critical thinking skills in tertiary academic writing through the use of an instructional rubric for peer evaluation* (Issue July). [https://ir.canterbury.ac.nz/bitstream/handle/10092/6866/thesis\\_fulltext.pdf?sequence=1](https://ir.canterbury.ac.nz/bitstream/handle/10092/6866/thesis_fulltext.pdf?sequence=1)
- Mazlili Suhaini, Ahmad, A., & Bohari, N. M. (2021). Assessments on Vocational Knowledge and Skills: A Content Validity Analysis. *European Journal of Educational Research*, 10(1), 641–655. <https://eric.ed.gov/?id=EJ1250404>
- Nejmaoui, N. (2018). Improving EFL Learners' Critical Thinking Skills in Argumentative Writing. *English Language Teaching*, 12(1), 98. <https://doi.org/10.5539/elt.v12n1p98>
- Noble, H. W. (2019). *Modern versus Traditional Teaching Methodologies: an Experimental Design*.

- Orhan, A., & ÇEVİKER AY, Ş. (2022). Developing the Critical Thinking Skill Test for High School Students: A Validity and Reliability Study. *International Journal of Psychology and Educational Studies*, 9(1), 130–142. <https://doi.org/10.52380/ijpes.2022.9.1.561>
- Osman, S., Mohammad, S., Abu, M. S., Mokhtar, M., Bunyamin, M. A. H., Abdullah, A. H., Jambari, H., & Ahmad, J. (2019). Math-related critical thinking theory in civil engineering design. *Pertanika Journal of Social Sciences and Humanities*, 27(2), 899–913.
- Pakirathan, P. G., & Kepol, N. (2018). Perceived teachers' self-disclosure, writing performance and gender of Malaysian ESL Undergraduates. *Pertanika Journal of Social Sciences and Humanities*, 26(4), 2195–2210.
- Park, T. (2005). An investigation of an ESL placement test of writing using Many-Facet Rasch measurement. *Teachers College, Columbia University Working Papers in TESOL & Applied Linguistics*, 4(1), 1–21. <http://journals.tc-library.org/ojs/index.php/tesol/article/viewArticle/41>
- Pei, Z., Zheng, C., Zhang, M., & Liu, F. (2017). Critical Thinking and Argumentative Writing: Inspecting the Association among EFL Learners in China. *English Language Teaching*, 10(10), 31. <https://doi.org/10.5539/elt.v10n10p31>
- Ployhart, R. E. (2012). The Content Validity of Cognitively Oriented Tests: Commentary on Schmidt (2012). *International Journal of Selection and Assessment*, 20(1), 19–23. <https://doi.org/10.1111/j.1468-2389.2012.00575.x>
- Putka, D. J., & Sackett, P. R. (2010a). Reliability and validity. *Handbook of Employee Selection*, 9–49. <https://doi.org/10.4324/9780203809808>
- Putka, D. J., & Sackett, P. R. (2010b). Reliability and validity. *Handbook of Employee Selection*, 11, 9–49. <https://doi.org/10.4324/9780203809808>
- Rusfandi. (2015). Argument-Counterargument Structure in Indonesian EFL Learners English Argumentative Essays: A Dialogic Concept of Writing. *RELC Journal*, 46(2), 181–197. <https://doi.org/10.1177/0033688215587607>
- Sabu, V., & Vernandes, B. (2019). An Analysis of Student's Critical Thinking Skills in Writing Argumentative Essay. *Vivid: Journal of Language and Literature*, 8(2).
- Schneer, D. (2014). Rethinking the Argumentative Essay. *TESOL Journal*, 5(4), 619–653. <https://doi.org/10.1002/tesj.123>
- Tavakoli, E., Amirian, S. M. R., Burner, T., Davoudi, M., & Ghaniabadi, S. (2018). Operationalization of Formative Assessment in Writing: An Intuitive Approach to the Development of an Instrument. *Applied Research on English Language*, 7(3), 319–344.
- Toba, R., Noor, W. N., & Sanu, L. O. (2019). The Current Issues of Indonesian EFL Students' Writing Skills: Ability, Problem, and Reason in Writing Comparison and Contrast Essay. *Dinamika Ilmu*, 19(1), 57–73. <https://doi.org/10.21093/di.v19i1.1506>
- Tomak, B. (2021). Designing an instructional unit on “argumentative essay” for prep year students: An action research in a state university in Turkey. *African Educational Research Journal*, 9(1), 1–19. <https://doi.org/10.30918/aerj.91.20.202>
- Viñas, L. F. (2022). Testing the Reliability of two Rubrics Used in Official English Certificates for the Assessment of Writing. *Revista Alicantina de Estudios Ingleses*, 36, 85–109. <https://doi.org/10.14198/RAEI.2022.36.05>
- Widyaningrum, F. A. D., & Prabandari, C. S. (2013). Content Validity and Authenticity of the 2012 English Test in the Senior High School National Examination. *Language and Language Teaching Journal*, 16(01), 23–29. <https://doi.org/10.24071/llt.2013.160103>

- Widyastuti, S. (2018). Fostering critical thinking skills through argumentative writing. *Cakrawala Pendidikan*, 37(2), 182–189. <https://doi.org/10.21831/cp.v37i2.20157>
- Wu, Y. (2021). Critical Thinking in Argumentative Essays. *Business Prospects*, 2(2), 60–65. <https://doi.org/10.52288/bp.27089851.2021.12.10>
- Zaini, M. H., Baharum, N. D., & Ahmad Sidiki, A. F. S. (2018). Incorporating critical thinking: Teaching strategies in an english language programme. *Pertanika Journal of Social Sciences and Humanities*, 26(4), 2379–2393.
- Zohrabi, M. (2013). Mixed method research: Instruments, validity, reliability and reporting findings. *Theory and Practice in Language Studies*, 3(2), 254–262. <https://doi.org/10.4304/tpls.3.2.254-262>