



P-ISSN 2355-2794  
E-ISSN 2461-0275

## Teachers' Made Writing Assessment Instrument: Pedagogical Insight from Process to Product

Khusnul Khotimah<sup>1,2</sup>  
Deisyi Anna Batunan<sup>\*1,3</sup>  
Nur Mukminatien<sup>1</sup>

<sup>1</sup>Department of English Language Education, Faculty of Letter, Universitas Negeri  
Malang, Malang 65145, INDONESIA

<sup>2</sup>Department of English Education, Faculty of Teacher Training and Education,  
Universitas Mataram, Mataram 83115, INDONESIA

<sup>3</sup>Department of English Education, Faculty of Language and Arts, Universitas Negeri  
Manado, Tondano 95618, INDONESIA

### Abstract

*Despite the urgency of assessment instruments in teaching foreign language writing, evidence regarding how teachers develop assessment instruments and maintain their quality, particularly for comparison and contrast essay writing based on a genre-based approach to teachers' classroom practices, is scantily reported. To address this empirical need, this article showcases how a set of teacher-made assessment instruments for a comparison and contrast essay based on a genre-based approach in a writing essay course at a public university in East Java was carefully and deliberately developed. To attain the research objective, this research enacted design-based research (DBR) inspired by Cavallaro and Sembiante (2021). In addition to content validity, face validity and inter-rater reliability were ensured by involving ten students, three fellow writing teachers, and two trained raters. This assessment instrument development project produces a set of valid formative assessment instruments, comprising a writing prompt and scoring rubric. The inter-rater reliability shows a high degree of positive correlation with a reliability coefficient of 0.862 from a correlational analysis based on Pearson correlation coefficient. This in-practice assessment instrument*

---

\* Corresponding author, email: [deisyibatunan@unima.ac.id](mailto:deisyibatunan@unima.ac.id)

**Citation in APA style:** Khotimah, K., Batunan, D. A., & Mukminatien, N. (2024). Teachers' made writing assessment instrument: Pedagogical insight from process to product. *Studies in English Language and Education*, 11(1), 95-115.

Received July 6, 2022; Revised March 27, 2023; Accepted December 9, 2023; Published Online January 31, 2024

<https://doi.org/10.24815/siele.v11i1.26774>

*development can be used as a reference, contributing to the teachers' assessment literacy development. It also serves as a valuable model for future researchers to develop and expand nested classroom assessment scholarship. A well-structured validity and reliability process of assessment instrument development will practically enhance the teaching practices in the essay writing course in particular and EFL teaching in general.*

**Keywords:** Classroom assessment, comparison and contrast essay, teacher-made assessment instrument, teachers' assessment literacy, writing assessment.

## 1. INTRODUCTION

The scholarship towards establishing quality education through the teaching process cannot be separated from the discussion on effective assessment. In the context of EFL writing instruction, effective assessment is also crucial for raising students' writing achievement. In its development, the practice of assessment is evolving from being previously dominated by a summative orientation as Assessment of Learning (AoL) to initiatives that emphasize more on the use of it to promote students' learning (Assessment for Learning, or AfL) and on the use of it to develop students' metacognitive skills (Assessment as Learning, or AaL). AoL, underpinned by behaviorist theory, emphasizes the use of objective evidence like scores to measure students' writing performance (Shepard, 2000). AfL, grounded in a social constructivist framework, puts the spotlight on classroom assessment to serve as a baseline to inform and facilitate students' learning in EFL writing classes (Fitriyah et al., 2022). This approach creates opportunities for students to acquire knowledge and fosters their agency in overcoming associated challenges (Fitriyah et al., 2023).

Furthermore, AaL involves students in self and peer assessment to oversee their learning advancement and cultivate their self-regulation skills (Alderson et al., 2017). To date, a huge number of assessment scholars (e.g., Fitriyah et al., 2023; Lee, 2017; Leong et al., 2018; Popham, 2009; Wu et al., 2021) argue that when the classroom tends to use assessment for learning and assessment as learning rather than the assessment of learning, students will learn better, which could address teachers' expectations. On the same note, Stiggins (2002) asserts, "If we wish to maximize students' achievement in American schools, we must pay far greater attention to the improvement of classroom assessment" (p. 1). These perspectives highlight the importance of classroom assessment in successful EFL writing teaching-learning, suggesting more attention be paid to the teachers' assessment literacy in developing a particular assessment instrument.

Through a systematic literature review, Glušac and Milić (2022) conclude that teacher-made assessment instruments have a positive impact on students' learning, including students' writing performance. This is attributed to the fact that these instruments typically provide information on the test's purpose, procedures, and tasks, aspects that standardized tests may lack. However, this sort of instrument usually possesses some flaws and is considered problematic, showing a lack of teachers' design knowledge and skill (DiDonato-Barnes et al., 2013; Glušac & Milić, 2022), and

is frequently not validated (Ugwu & Mkpuma, 2019). This postulation reiterates the urgent need to enhance teachers' assessment literacy, particularly by preparing them to be more knowledgeable and skilled in developing valid and reliable self-made assessment instruments. Relatedly, a significant volume of literature has been extensively devoted to scrutinizing teacher-made assessment instruments. The scholarship mainly focuses on providing evidence of teacher-made assessment's validity and reliability (e.g., Effendi & Mayuni, 2022; Ing et al., 2015; Minda, 2018; Setiabudi et al., 2019), content knowledge, cognitive evidence, and authentic evidence (Pesiwarissa et al., 2021). In the context of writing, the text types were also limited to argumentative text (Mukminatien et al., 2021), descriptive text (Santi et al., 2019), and narrative and expository text (Jeong, 2017). The perusal chronicling the step-by-step process of how a teacher-made writing assessment instrument was developed and validated remains underreported (Glušac & Milić, 2022; Schedl & Malloy, 2014), especially for comparison and contrast essay writing based on a genre-based approach. The explicit picture of the development of an assessment instrument would give a practical contribution to the development of teachers' assessment literacy to self-design and develop a valid and reliable assessment for their effective writing teaching practices. Thus, this article aims to respond to this empirical void by picturing the development assessment instrument with two guiding research questions:

1. How is the assessment instrument developed?
2. How is the assessment instrument's quality maintained?

## 2. LITERATURE REVIEW

### 2.1 Teacher-Made Assessment Instrument: Teacher's Role and the Key Principles

Teaching is an overall process of instructional activity that covers planning the instruction, developing and delivering materials, facilitating students' learning, and assessing students' performance and achievement. Therefore, teachers' roles in the classroom are multifaceted and, ideally, should not be reduced to one particular domain. Teachers should play a part as curriculum evaluators and assessors by constructing a test as an instrument of assessment to gain information about the students' knowledge, ability, and skills as an outcome of the instructional activities (Brown, 2004; Lee, 2017).

Assessment literacy was first introduced by Stiggins in 1991, making reference to the apprehension and prowess of enacting the knowledge of assessment in a varied procedure of learners' attainment (Xu & Brown, 2016). Twelve years later, Inbar-Lourie (2008) introduced the language assessment literacy term, referring to a multifaceted tenet (Pill & Harding 2013) that consists of knowledge, skills, and principles of language assessment (Davies, 2008). Despite a wealth of body literature showing studies related to the attempt to design teacher-made test instruments, yet, some researchers (e.g., DiDonato-Barnes et al., 2013; Glušac & Milić, 2022; Stiggins, 1991) pinpoint that teachers are still incapable of carrying out a vigorous test and do not have adequate knowledge and skills to design assessment instruments. In a similar vein, some empirical studies also sound out unsatisfactory conclusions, pointing out some challenges faced by teachers, such as a lack of knowledge on language

assessment (Razavipour et al., 2011), discrepancies between teachers' knowledge and practice (Jannati, 2015), and the propensity to apply traditional ways of assessment (Tsagari & Vogt, 2017). Scholars (e.g., DeLuca et al., 2016; Janatifar & Marandi, 2018; Tsagari & Vogt, 2017) suggest that more training on assessment should be offered.

Correspondingly, to bring a quality classroom assessment, teachers should adhere to some key principles. Brown (2004) points out that a good classroom test should meet these criteria: validity, reliability, practicality, and authenticity. Validity is understood as the degree to which a test instrument can measure what is intended to be measured (Blumberg et al., 2005; Field, 2005; Robson, 2011). Brown (2004) and Pallant (2011) suggest that validity has four types: content validity, face validity, construct validity, and criterion-related validity. The test instrument has content validity if it measures the behavior, skills, and knowledge that are supposed to be measured (Brown, 2004; Creswell, 2005). Face validity seems akin to content validity, but it wants to see if the test instrument is clear and free from multi-interpretation from the point of view of non-experts such as the test takers (Oluwatayo, 2012). Construct validity is intended to see how the constructed test instrument aligns with the existing theories. Criterion validity is demonstrated to see 'to what extent the test's criterion has been met (Brown, 2004). Furthermore, reliability is defined as the stability or consistency of the scores that students achieve each time they use the test, regardless of who the evaluator/assessor is and how many times the test is taken (Brown, 2004; Livingston, 2018; O'Neill & Murphy, 2017). Next, practicality refers to the test that should be easy to run and should not be costly under reasonable time restrictions with a clear scoring scheme (Brown & Abeywickrama, 2019). Lastly, authenticity could be defined as how closely the test corresponds to real-life situations (Brown, 2004). The test should be interesting, meaningful, and contextualized for the students.

## **2.2 Comparison and Contrast Writing in Genre-Based Approach**

The approaches in teaching writing have been dynamic from time to time. To date, there have been four teaching writing approaches: product, process, genre, and process genre-based approach (Badger & White, 2000). The product approach as the name suggests highly emphasizes the product of writing that focuses on grammar structure, vocabulary, and cohesive markers in the whole of the writing. As a result, to achieve this accuracy, students need to undergo these processes of writing: familiarization (students need to get exposed to the attributes of the texts), controlled writing, and guided writing (students are guided to achieve their independence to write assisted with table consisting needed phrases and sentences and pictures) and free writing (students can produce the whole text on their own). The process approach focuses on writing development and views the writing process as a cyclical process of writing starting from prewriting, composing/drafting, revising, to editing. In pre-writing, students need to do brainstorming to generate as many ideas about the topic as possible and organize those ideas in an outline. During the composing/drafting stage, students start composing the writing based on the outline as their first draft. Students then need to revise and edit the draft. As a cyclical process, editing is not the end of the writing process, but students can revise their writing at any point to produce more drafts and have the last piece for publication. Furthermore, genre could be understood both as an act of grouping things into categories by spotting their attributes

and as a platform for people to interrelate under a shared accord (Suharyadi, 2021). It is grounded in systemic functional theory and later developed by its adherents, including J.R. Martin (Firkins et al., 2007).

Badger and White (2000) point out that the genre approach shares some principles in common with the product approach; for instance, they both focus on linguistics aspects and the importance of providing students with a model of texts, serving as language input that students could take as a model (Badger & White, 2000; Hyland, 2007). Differing from other writing approaches, this approach emphasizes the significance of social context: a product of writing is embedded in the purpose, audience, and structure of the text (Hyland, 2007).

A genre-based approach in writing classrooms has been realized in several designs such as the curriculum cycle (Callaghan & Knapp, 1989), a wheel model of a teaching and learning cycle (Hammond et al., 1992), and the teaching and learning cycle (TLC) (Caplan & Farling, 2017; Hermansson et al., 2019). These frameworks accord with necessitating students to work hand in hand and individually to produce a piece of writing and emphasize exposing students to exemplary texts along with their functions, structure, and target audience. The process genre approach is introduced to overcome the weaknesses of previous writing approaches as this approach deals with the writing development and genre of the text. In the process genre approach, learners need to pay attention to the context, audience, and purpose of the text, and genre (social function, generic structure, and language feature of one text). Then, they will undergo a series of cyclical processes namely planning, drafting, and publishing.

Relatedly, an essay is a group of paragraphs that discusses one single idea and has an introductory paragraph, developmental paragraphs, and a concluding paragraph that discusses causes and effects, reasons, examples, processes, classifications, or points of comparison and contrast (Smalley et al., 2001). A comparison and contrast essay is defined as a type of essay that discusses both similarities and differences; the similarities between two entities and the divergences that distinguish them (Englert et al., 1991). In other words, the function of this text is to showcase how two things, persons, and places are the same and to demonstrate how two things, persons, and places are different. In terms of its generic structure, like other essays, it has introductory, developmental/body, and concluding paragraphs. In the introductory paragraph, it has an opening statement, connecting sentences, and a thesis statement. In body paragraphs, each body paragraph has a topic sentence, supporting sentences, and a concluding sentence. Moreover, Oshima and Hogue (1998) explain that the idea in a comparison and contrast essay could be developed in the block and point-by-point organization. In block organization, the writer should write or organize all the differences in one place (block/paragraph) and move to all the similarities in another block/paragraph.

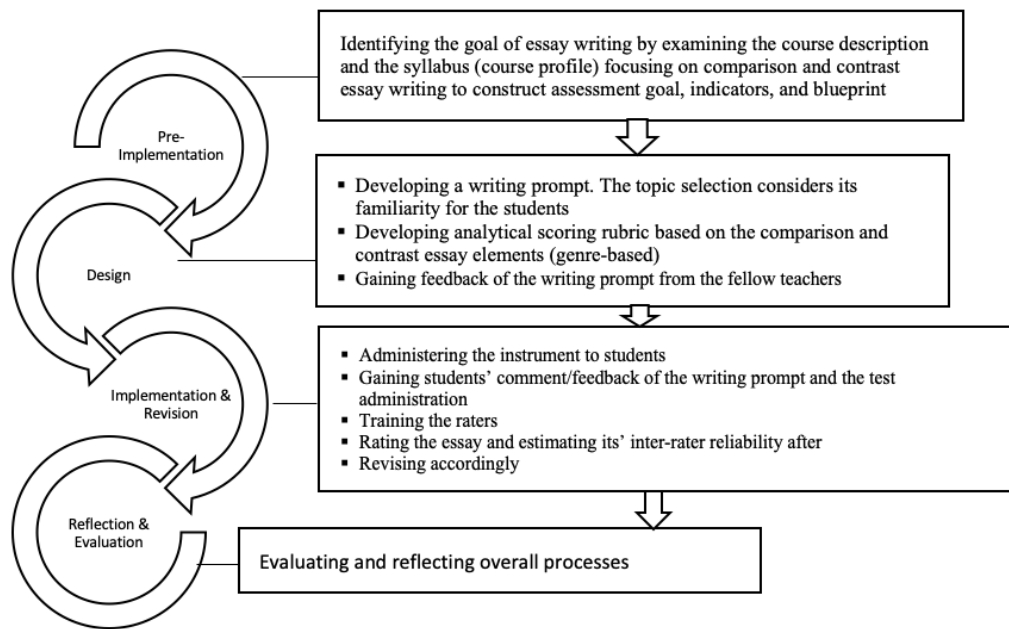
On the other hand, the ideas are organized based on the point of comparison in the point-by-point organization. In other words, in one section, under one point of comparison, two subjects are being compared/contrasted. Finally, a concluding paragraph summarizes the ideas presented in the introductory and body paragraphs. Sometimes, it echoes the thesis statement using different sentences or expressions. As comparison and contrast writing aims to highlight the differences, similarities, or both similarities and differences, this kind of writing requires the writer to use comparative degree and transitional signals that compare and contrast. The transitional signals are mostly used in comparison and contrast writing, for example, “however, in contrast,

in comparison/on the other hand, on the contrary, while/whereas, differ from, compared to, compare with, and different from/dissimilar” for contrast and “likewise, similar, as just as, not only... but also, like/just like/similar to, similar to, the same as, and likewise” for comparison (Oshima & Hogue, 2007, p. 143).

Five components of writing are assessed in the comparison and contrast essay namely content, organization, vocabulary, language use, and mechanics (Hughes, 2008). Regarding content, it refers to students' ability to discuss the given topic in their writing. For example, if the students are asked to write a comparison and contrast between two courses that they have taken, the content of the essay should address these two courses. Students should not write anything other than the given topic. As for organization, the students should demonstrate their abilities to meet the comparison and contrast essay development. Apart from having complete introductory, body/developmental, and concluding paragraphs, students' comparison and contrast essays should perform a point-by-point or block-organization pattern. In terms of vocabulary, students' essays should include a wide range of vocabulary use with collocation. In language use, students should be able for example to use tenses, subject and verb agreement, and transitional signals correctly. Finally, mechanics refers to students' ability to use punctuation, capitalization, and spelling appropriately.

### **3. METHOD**

The primary goals of this study were to explore the development and quality control of a classroom assessment instrument for writing a comparison and contrast essay for second-year EFL undergraduate students at a public university in East Java, Indonesia. To attain these research objectives, this research enacted design-based research (DBR), inspired by Cavallaro and Sembiente (2021). DBR is set to study particular instructional aspects more comprehensively while designing an instructional product. This research design is suitable for this scholarly work since it is congruent with the aims of this study. It possesses four main procedures, i.e., 1) pre-implementation; 2) design; 3) implementation and revision; and 4) reflection and evaluation. The pre-implementation aims to achieve familiarity with the researched environments' practices, norms, and behaviors. In Phase 2, the assessment instrument for comparison and contrast essay is developed. Feedback from fellow teachers is gathered. Later, in phase 3, the products are implemented and revised accordingly. Lastly, the overall processes and products are evaluated and reflected. The details of how this research design is integrated into the assessment instrument development are depicted in Figure 1.



**Figure 1.** Assessment instrument development steps in DBR phases, inspired by Cavallaro and Sembiante (2021).

Following the first phase of pre-implementation, the second phase of design, and the third phase of implementation and revision, we involved three fellow writing teachers and 10 students to give feedback on the writing prompt. The three fellow writing teachers were asked to analyze and give written comments on the developed assessment instrument, about its language (if the instruction and the language used are clear and not ambiguous), content (if the test content is relevant to the instructional objective (s), assessment goal(s), and the test indicators), and format (if the test time allotment is reasonably sufficient, the format is well organized and complete, containing: test identity and instruction). The 10 students were asked to write a comparison or contrast essay about two different places: two tourist attractions, a traditional market and supermarket, and two cities. Further, after doing the test, the students were asked to complete a survey and write a short reflection about their experience in completing the test using the developed writing test prompt. Additionally, two raters were recruited and trained to rate the students' essays. The process of involving those students, teachers, and raters was based on relevance and accessibility. The students were those who were participating in the essay writing course from a different cohort with a similar level of language proficiency (intermediate to pre-advanced level) to the targeted students for the assessment instrument's enactment. They were involved because one of the authors had access to gain their consent to participate. The three fellow teachers and two raters were recruited based on their expertise in teaching writing, particularly essay writing courses with more than seven years of teaching experience.

Enmeshed in the DBR approach, the overall processes of the assessment instrument development, including the participating individuals, were used as a source of data collection. In this manner, self-observational data (the authors as the assessment instrument developers), 10 students' comparison and contrast essays, students and three fellow teachers' feedback towards the instrument, and the two

raters' scoring results were gathered and collectively analyzed, and interpreted to address the research questions. We used data from the self-observation processes of the four DBR phases to investigate how the assessment instrument was developed. Further, we utilized the data from the students' essays and written feedback after the try-out, the fellow teachers' written feedback towards the assessment instrument, and the raters' scoring results of students' essays to determine how the quality of the assessment instrument was maintained. The data from each data source was first analyzed separately using qualitative or quantitative analysis, as appropriate.

In the quantitative analysis, to estimate the assessment instrument inter-rater reliability, the Pearson correlation coefficient formula was employed to determine the scoring consistency of the two sets of scores produced by the two raters. The result of the correlation analysis was interpreted based on the common use of the interpretation table depicted in Table 1.

**Table 1.** Size of correlation interpretation.

| Size of correlation       | Interpretation                            |
|---------------------------|-------------------------------------------|
| .90 to 1.00 (-.90-1.00)   | Very high positive (negative) correlation |
| .70 to .90 (-.70 to -.90) | High positive (negative) correlation      |
| .50 to .70 (-.50 to -.70) | Moderate positive (negative) correlation  |
| .30 to .50 (-.30 to -.50) | Low positive (negative) correlation       |
| .00 to .30 (-.00 to -.30) | Negligible correlation                    |

For the qualitative analysis, we aimed to identify the common patterns from the overall fellow teachers' and students' written feedback towards the teacher-made assessment instrument. To achieve this, we carefully transcribed the overall written feedbacks, repeatedly read them, coded them, and interpreted them accordingly to address the intended research objectives. Some simple calculations from the brief survey were also performed to determine the inclusive patterns. In the last phase of reflection and evaluation, the results of qualitative and quantitative analyses were combined to make a relevant interpretation of the overall results. Some key points to ponder and implications were also drawn accordingly.

## **4. FINDINGS**

### **4.1 The Development of the Assessment Instrument**

The very first phase of writing prompt and scoring rubric development for a comparison and contrast essay was identifying the course description of the essay writing course, delineated as follows:

- (1) This course develops students' ability to write expository essays using different methods of development: logical division of ideas, chronological order, comparison and contrast, cause and effect, and classification.

This course description was generated by a team of writing teachers by considering the contextual instructional design and the supporting theories (Brown, 1995; Nation & Macalister, 2010). The course description indicates that the essay writing course aims to develop students' writing abilities for expository essays,

focusing on the method to develop logical division of ideas, chronological order, comparison and contrast, cause and effect, and classification. It signifies that a comparison and contrast essay is not the sole essay taught in the course. The development of this essay's assessment instrument is, therefore, part of the wider instructional goals and is not designed to represent the whole course assessment.

The next step was examining the alignment between this course description (course profile) and the developed instructional objectives. This process served as a baseline to generate the test goal and indicators. This deliberate and careful analysis aimed to ensure the content validity of the assessment instrument that every teacher could do either independently or in a team of essay-writing teachers. From this process, the assessment goal and indicators were formed.

#### *4.1.1 The goal of the writing assessment*

The assessment aims to measure the students' ability to write a comparison and contrast essay about two different places: two tourist attractions, a traditional market and supermarket, and two cities, effectively, accurately, and appropriately by employing the features of a comparison and contrast structure.

#### *4.1.2 Indicators of the writing assessment*

This assessment aims to assess students' ability to:

- (1) employ the generic structure of a comparison and contrast essay: Making a) an introductory paragraph, b) body paragraphs using block or point-by-point structure, and c) a concluding paragraph correctly,
- (2) choose appropriate language use: vocabulary (adjectives, phrases of comparison and contrast) and grammatical structure (comparative degree, tenses) correctly,
- (3) use English sentences and transactional signals correctly and effectively,
- (4) use correct mechanics (spelling, punctuation, capitalization).

#### *4.1.3 Developing the scoring rubric*

The scoring rubric was developed by considering the assessment's goal and indicators representing the course objectives and covering features of a comparison and contrast essay. The element of the scoring rubric was formed in such a way, explicating the genre-based approach as how the comparison and contrast essay is taught in the instructions. Since the purpose of the scoring rubric is to properly guide the raters or teachers in assessing students' writing, its development should be anchored in a systematic, careful, and deliberate endeavor. In this respect, it aimed to provide valid and reliable evidence of students' writing achievement.

The instructional objectives reflected in the assessment goal and indicators would serve as the performance indicators that would be enacted as the guideline to provide scores of students' writing. Practically, the indicators were used to generate specific criteria for determining the proper score for a particular writing quality. The details of the indicators and the nested criteria are shown in Table 2.

**Table 2.** The blueprint of comparison and contrast essay assessment instrument.

| <b>Assessment goal:</b><br>The assessment aims to measure the students' ability to write a comparison or contrast essay about two different places: two tourist attractions, a traditional market and supermarket, and two cities effectively, accurately, and appropriately by employing the features of a comparison and contrast structure correctly. |                                                                                                                                                                                                                                                                                                                                                                 |             |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------|
| Essay Component<br>(The components were formulated on the basis of the instructional objectives represented in the assessment goal)                                                                                                                                                                                                                      | Criteria (indicators) for qualified comparison and contrast essay                                                                                                                                                                                                                                                                                               | Weight in % |
| Introductory Paragraph                                                                                                                                                                                                                                                                                                                                   | <ul style="list-style-type: none"> <li>- write a complete structure (opening statement, connecting sentence, and thesis statement)</li> <li>- introduce the point of comparison and contrast</li> <li>- attract the reader's attention (hook)</li> </ul>                                                                                                        | 16          |
| Body Paragraphs                                                                                                                                                                                                                                                                                                                                          | <ul style="list-style-type: none"> <li>- contain a clearly focused topic sentence, supporting sentences, and a concluding sentence that supports the thesis statement</li> <li>- use clear, specific, and sufficient details (supporting sentences) in supporting and developing the topic sentence for a thorough comparison or contrast.</li> </ul>           | 28          |
| Concluding Paragraph                                                                                                                                                                                                                                                                                                                                     | Unify/ tie the overall essay and the points of comparison and contrast effectively                                                                                                                                                                                                                                                                              | 16          |
| Language Use (Vocabulary and Grammar)                                                                                                                                                                                                                                                                                                                    | <ul style="list-style-type: none"> <li>- use appropriate and a wide range of technical terms and registers</li> <li>- use correct grammatical structures: tenses, comparison degree, and other required grammatical structures.</li> </ul>                                                                                                                      | 20          |
| Sentences and Transition                                                                                                                                                                                                                                                                                                                                 | <ul style="list-style-type: none"> <li>- write smooth transitions between ideas or paragraphs</li> <li>- employ comparison and contrast transition words to show a relationship between ideas correctly</li> <li>- use effective sentences</li> <li>- employ a variety of sentence structures (simple, complex, compound, complex compound sentence)</li> </ul> | 12          |
| Mechanics                                                                                                                                                                                                                                                                                                                                                | <ul style="list-style-type: none"> <li>- use punctuation correctly</li> <li>- use spelling correctly</li> <li>- use capital letters correctly</li> </ul>                                                                                                                                                                                                        | 8           |

The elements of the blueprint represented in Table 2 were used to develop the scoring rubric. Anchored in GBA, the assessment criteria were ensured to cover the generic structure of a comparison and contrast essay that is distinctive to other essay types. The common features of writing performance indicators were either embedded within the comparison and contrast essay features (e.g., content and organization) or made as explicit criteria (e.g., language use, mechanics). The rating scale ranged from 1 to 100, whereby each performance indicator has a different weight in the overall score. The weight of indicators was formulated based on the undertaking portion of each indicator: 60% for the content and organization of overall paragraphs (16% for the introductory paragraph, 28% for body paragraphs, and 16% for the concluding paragraph), 20% for language use, 12% for sentences and transition, and 8 % for mechanics. The minimum score was set at 25 since there was no zero in the scoring guideline. This decision was made to appreciate the students' effort in writing the essay, as every one of them did write it. Empirically speaking, the decision to give different weights to each indicator was inspired by the work of [Bloxham et al. \(2016\)](#),

who underscore the effectiveness of collaborative marking processes in aligning standards and the inherent complexity of assessment decisions.

This research suggests that teachers or educators should prioritize transparency and actively assist students in understanding the nuanced nature of these judgments. Attempting to address the challenge of balancing standardization and complexity in designing assessment criteria, adjusting the weights of different assessment criteria could be an effective approach for improving the accuracy and consistency of writing assessments. Further, [Ghanbari and Barati \(2020\)](#) found that putting more weight on the content indicator might increase the reliability of the writing assessment instrument. In a nutshell, the developed scoring rubric encompasses the necessary features of comparison and contrast essays and common writing conventions in the performance criteria.

#### 4.1.4 Developing the writing prompt

The writing prompt development was based on the selected relevant topics available in the students' textbook and from authentic open-access sources. It was closely related to the materials being taught in the classroom. The writing prompt is provided in Figure 2.

| Comparison and Contrast Essay Writing Prompt<br>(90 minutes) |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
|--------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Direction</b>                                             | <p>1. Suppose that you had a 2-week vacation last month. You visited many places such as tourist attractions, markets, and cities. Having back from your vacation, you were emailed by your friend to recommend the best place to visit. Therefore, write down an essay comparing and contrasting two places to help your friend decide where to visit. You may choose one of the following topics:</p> <ul style="list-style-type: none"> <li>Two tourism attractions</li> <li>Two markets</li> <li>Two cities</li> </ul> <p>Be sure to include an engaging introduction, a clear thesis, well-defined and detailed paragraphs, and a fitting conclusion that ties everything together. You may use point-by-point or block organization.</p> <p>2. Your essay will be evaluated by the following aspects: generic structure of a comparison and contrast essay, language use (vocabulary and grammatical structure), sentence construction and transactional signals and mechanics (spelling, punctuation, and capitalization)</p> <p>3. You have 90 minutes to write your 400-500 words essay without looking up the dictionary.</p> |

**Figure 2.** Comparison and contrast essay writing prompt.

#### 4.1.5 Administering the instrument

Before assessing the students' writing, the instrument was reviewed by three writing fellow teachers who were not involved in the instrument development process. Some revisions were made as a result of this process. The modified assessment tool was subsequently administered to a separate group of undergraduate students within the English department who had similar language level to the original group of students under consideration. This tryout aimed to evaluate the instrument in terms of its appropriateness, clarity, and adequacy of the time allotment. In so doing, the tryout process was deliberately observed. After the tryout, students were also asked to fill out a brief survey and invited to reflect.

#### *4.1.6 Rating the essay and estimating the inter-rater reliability*

After the process of writing the assessment, 10 essays were collected. These essays were then blind-rated by two trained raters, wherein they worked independently following particular procedures. The rating process was done after sufficient training, during which the raters were trained to score properly based on the scoring rubric. First, they were invited to see the details of each performance indicator. Then, a demonstration to assess one essay was conducted. Lastly, they trained themselves to score the students' essays based on the prescribed scoring rubric. After ensuring that the two raters had consistent scores, they started scoring the students' essays.

With the scores of students' essays, the inter-rater reliability was estimated by employing a correlational analysis to see the consistency of the two sets of scores produced by the two raters. The result of this calculation shows a high positive correlation between the two sets of scores.

## **4.2 Maintaining the Quality of the Assessment Instrument**

To ensure the quality of the assessment instrument, the test developers employed content and face validity and inter-rater reliability. The content validity of this developed teacher-made assessment instrument had been achieved since this assessment instrument for a comparison and contrast essay was congruent with the course objectives and essay writing course syllabus from the institution. The test developers perused the existing catalog and the writing course syllabus to ensure the developed assessment instrument met the stated indicators and measured the intended students' comparison and contrast essay writing skills, knowledge, and abilities.

In terms of face validity, we asked three fellow writing teachers' opinions on the writing assessment instrument sheet in terms of the language (if the instruction and the language used are clear and not ambiguous), content (if the test content is relevant to the instructional objective(s), the assessment goal(s), and the test indicators), and format (if the test time allotment is reasonably sufficient and the format is well organized and complete, containing the test identity and instruction). We requested that they put a checkmark on the given test instrument aspects to evaluate whether they are 'yes', 'no', or 'not sure'. Overall, they provided positive feedback by which they mostly ticked 'yes' for all elements. Apart from placing a tick, they gave some notes for test improvement. Table 3 summarizes writing fellow teachers' feedback for the developed writing assessment instrument.

**Table 3.** Fellow writing teachers' feedback.

| <b>Fellow Writing Teachers (FWT)</b> | <b>Good points highlighted (+)</b>                                                                                                                                                                    | <b>Points needing improvement (-)</b>                                                                                                                        |
|--------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>FWT 1</b>                         | The assessment instrument is well developed with reasonable time allotment and clear instruction. I think it covers features of comparison and contrast essays. The template is helpful for students. | My suggestion is to put a brief explanation about important indicators to assess students' writing. This will help students on how they should write better. |

Table 3 continued...

|              |                                                                                                                   |                                                                                                                                                                                                                                                                                                                                             |
|--------------|-------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>FWT 2</b> | This writing assessment instrument is well-developed. It has strengths in terms of language, content, and format. | The indicators should refer to the course objective which focuses on the comparison and contrast essay writing and its characteristics. Also, the writing prompt needs to refer to the indicators in detail.                                                                                                                                |
| <b>FWT 3</b> | In my opinion, the test instrument is well-structured and quite complete.                                         | If I may give suggestions, it would be better if the number of words students have to write is included and the time for the test is reduced because half an hour is too long for a writing test. In addition, students may also be reminded about the accuracy of their works and whether they are allowed to look up dictionaries or not. |

In general, these three fellow writing teachers acknowledged that the assessment instrument was well-developed in terms of language, content, and format. They reached a consensus in terms of the offered suggestion to enhance its quality: to be explicit about the indicators of the assessment instrument. Based on the feedback from fellow writing teachers, we revised and improved our developed writing assessment instrument. Related to the given suggestion about the time for completing the assessment, when it came to the piloting phase, we found out that the given time (90 minutes) was sufficient for the students to complete their writing.

Following this step, we administered a pilot study (trying it out) to ten students to see the readability (clear and free from ambiguity), the adequacy of time allotment, and the test relevance (with the learned material). In the students' opinion sheet on the writing assessment instrument, these participating students were given the aspects of the assessment instrument (language, content, and format) and asked to provide a tick on the three options ('yes', 'no', 'not sure') under each aspect of the assessment instrument. All these participating students agreed that the assessment instrument was clear and unambiguous, the time allotment was adequate, and the test was relevant to the learned materials. Table 4 shows the participants' opinions.

**Table 4.** Students' participants' opinions about the assessment instrument.

| Statements                                                                                                                               | Students' responses |      |           |      |           |      |
|------------------------------------------------------------------------------------------------------------------------------------------|---------------------|------|-----------|------|-----------|------|
|                                                                                                                                          | Yes                 |      | No        |      | Not Sure  |      |
|                                                                                                                                          | In number           | In % | In number | In % | In number | In % |
| <b>Language</b><br>(The instruction and writing prompt are clear and free from ambiguity.)                                               | 10                  | 100  | 0         | 0    | 0         | 0    |
| <b>Content</b><br>(The test content is relevant to the learned materials.)                                                               | 10                  | 100  | 0         | 0    | 0         | 0    |
| <b>Format and test time</b><br>(The format is appropriate and practical to follow.<br>The test time allotment is reasonably sufficient.) | 10                  | 100  | 0         | 0    | 0         | 0    |

The students also shared their opinions about the assessment. Most of them felt nervous about being assessed, yet, after listening to the explanation about the aim of

the assessment, which would not have academic consequences, they felt much better. One of the students voiced her reflections on the assessment process:

- (2) My first impression about the test was that I was a little nervous and I thought the test would be difficult. But when I saw the test and tried to do it, it turned out that the test was easy to understand. The explanation of the test was also easy to understand. The adequate time given also helped me to take this test calmly. This test was helpful to test my own writing skills. (S1)

To see the consistency of the scoring (reliability of the test), we employed inter-rater reliability by asking two experienced writing teachers to mark students' works. Table 5 shows the results from two raters, Rater 1 and Rater 2.

**Table 5.** Results from Rater 1 and Rater 2.

| Students' code | Scores from Rater 1 |    |    |    |    |   |       | Scores from Rater 2 |    |    |    |    |   |       |
|----------------|---------------------|----|----|----|----|---|-------|---------------------|----|----|----|----|---|-------|
|                | 1                   | 2  | 3  | 4  | 5  | 6 | Total | 1                   | 2  | 3  | 4  | 5  | 6 | Total |
| S1             | 8                   | 14 | 12 | 10 | 9  | 4 | 57    | 8                   | 14 | 8  | 10 | 9  | 4 | 53    |
| S2             | 4                   | 4  | 2  | 10 | 6  | 4 | 53    | 8                   | 14 | 8  | 10 | 3  | 4 | 47    |
| S3             | 8                   | 14 | 12 | 15 | 9  | 6 | 64    | 8                   | 14 | 8  | 10 | 6  | 6 | 52    |
| S4             | 12                  | 28 | 12 | 20 | 9  | 6 | 87    | 8                   | 14 | 12 | 15 | 6  | 6 | 61    |
| S5             | 8                   | 28 | 16 | 15 | 9  | 8 | 84    | 8                   | 21 | 12 | 15 | 6  | 6 | 68    |
| S6             | 16                  | 28 | 12 | 15 | 9  | 8 | 88    | 12                  | 21 | 12 | 15 | 6  | 6 | 72    |
| S7             | 8                   | 21 | 12 | 15 | 9  | 6 | 71    | 8                   | 28 | 12 | 15 | 9  | 6 | 78    |
| S8             | 12                  | 28 | 16 | 20 | 12 | 8 | 96    | 12                  | 28 | 16 | 20 | 12 | 6 | 94    |
| S9             | 12                  | 28 | 16 | 16 | 12 | 8 | 92    | 8                   | 28 | 16 | 15 | 9  | 6 | 82    |
| S10            | 12                  | 7  | 12 | 10 | 6  | 4 | 51    | 4                   | 7  | 8  | 10 | 6  | 2 | 37    |

The table of scores from the two raters, Table 5, shows that students performed satisfactorily on grammar and vocabulary aspects, but these students' participants underperformed on sentence construction, transition, and mechanics. Some students' writings lacked comparison and contrast phrases and transition signals, making their writings more descriptive.

After obtaining a set of scoring results from each Rater who rated the test independently, these two sets of scores were analyzed using a Pearson correlation coefficient. Based on the table of correlation interpretation (see Table 1), the results of the statistical analysis indicate a high degree of positive correlation between Rater 1 and Rater 2 in assessing the same set of essays using the developed scoring rubric. Table 6 shows the result.

**Table 6.** Pearson correlation coefficient result.

|         |                     | Rater 1             | Rater 2 |
|---------|---------------------|---------------------|---------|
| Rater 1 | Pearson Correlation | 1                   | .862**  |
|         | Rater 1             | Pearson Correlation | 1       |
|         | N                   | 10                  | 10      |
| Rater 2 | Pearson Correlation | .862**              | 1       |
|         | Sig. (2-tailed)     | .001                |         |
|         | N                   | 10                  | 10      |

\*\* Correlation is significant at the 0.01 level (2-tailed).

## 5. DISCUSSION

The findings of this DBR article have shown how a teacher-made assessment instrument was developed to adhere to the characteristics of a good classroom assessment instrument of practicality, reliability, validity, and authenticity as suggested by [Brown \(2004\)](#). The systematic development process from the pre-implementation to the reflection and evaluation ([Cavallaro & Sembiente, 2021](#)), involving fellow writing teachers, students, and trained raters, discloses a strong endeavor to produce a quality instrument. Practicality, reliability, and validity were maintained throughout the process of instrument development (content and face validity), observation of the assessment administration, feedback from fellow teachers and students, and the process of estimating inter-rater reliability. In short, the fellow writing teachers' and students' feedback has helped the test developers gradually enhance the quality of the developed assessment instrument. Coupled with the result of inter-rater reliability, which shows a high degree of positive correlation with a reliability coefficient of 0.862, this instrument is ready to be employed in real classroom settings. In congruence with [Hyland \(2007\)](#), the teacher-made assessment instrument was developed based on a genre-based approach that places a significant emphasis on the social context. As a result, this instrument is inherently authentic. Likewise, the test developers effectively incorporated the principles of effective assessment instruments, including weighing each component of students' writing indicators, as suggested by [Ghanbari and Barati \(2020\)](#). They also included the component of writing assessment for a comparison and contrast essay, as postulated by [Hughes \(2008\)](#), as well as comparison and contrast transitional signals, as advised by [Oshima and Hogue \(2007\)](#). Additionally, they incorporated block and point-by-point organization, as demonstrated by [Oshima and Hogue \(1998\)](#). This finding enunciates that designing quality assessment instruments is doable for teachers who hope to have the capacity to be qualified assessment designers and curriculum evaluators ([Brown, 2004](#); [Lee, 2017](#)) as a reflection of assessment literacy mastery ([Davies, 2008](#)).

Furthermore, the developed genre-based comparison and contrast essay writing assessment instrument represents the concern of equally translating the principles of the genre-based approach both in the teaching-learning process and its assessment. The coherence between the teaching approach and the employed assessment principle is fundamental to generating students' better learning quality. In other words, designing the right and appropriate assessment instrument for the delivered teaching to advocate students' optimum learning is the essence of the assessment literacy that every teacher should possess ([Stiggins, 1991, 2002](#)). The detailed and explicit explanation of the way this writing assessment instrument developed is likely to be an alternative reference to address the empirical dilemma, such as teachers' inability to construct a rigorous assessment instrument ([Stiggins, 1991](#)), numerous challenges for teachers to develop the instrument ([Razavipour et al., 2011](#)), discrepancies between teachers' knowledge and practice ([Jannati, 2015](#)), and propensity to apply traditional methods of assessment ([Tsagari & Vogt, 2017](#)). This written source of knowledge could be a complementary measure for the practical training on assessment literacy suggested by [DeLuca et al. \(2016\)](#), [Tsagari and Vogt \(2017\)](#), and [Janatifar and Marandi \(2018\)](#).

When receiving feedback from fellow writing teachers, suggesting explicit indicators in the assessment instrument means that teachers should consistently

connect the learned concept to its implementation in writing. In other words, teachers should be able to show how conceptual ideas are put into written form to tie all learned and practiced materials together. According to the in-depth examination, it is noteworthy that students tend to underperform in two areas of basic writing skills, namely sentence transitions, and mechanics, despite having presumably learned them in their previous writing courses. Furthermore, it was observed that some students' essays lacked comparison and contrast phrases as well as transition signals, resulting in overly descriptive writing. This highlights the importance for writing teachers to review and emphasize these aspects in daily instruction, along with essay development materials.

## **6. CONCLUSION**

Given the increasing attention to the teacher's role as an assessment instrument designer, coupled with the inadequate documentation of related references in the literature, this article sheds light on the careful and deliberate development of a set of teacher-made assessment instruments for a comparison and contrast essay. These instruments were based on a genre-based approach in a writing course at a public university in East Java and were designed to maintain their quality. This assessment instrument development project produced a set of formative assessment instruments comprising a writing prompt and scoring rubric. The findings highlight some fundamental procedures to enact in the instrument development to generate a practical, reliable, valid, and authentic teacher-made instrument. Based on the observation of the writing assessment administration to students, feedback from fellow teachers and students, and the process of estimating inter-rater reliability, the developed instrument was valid and reliable. The statistical analysis, employing a Pearson correlation coefficient, indicates that the correlation coefficient between Rater 1 and Rater 2 was 0.862 ( $p = 0.001$ ), which was lower than 0.01. It implies that the inter-rater reliability is in high positive correlation showing a high degree of agreement between trained Rater 1 and Rater 2 in assessing and judging the same essays based on the same developed scoring rubric.

This in-practice assessment instrument development can be used as a reference, contributing to the teachers' assessment literacy development, particularly in essay writing courses for undergraduate students in English language teaching. It also serves as a valuable model for teachers who would like to do similar assessment development and for future researchers to develop and expand the nested classroom assessment scholarship. A well-structured validity and reliability process of assessment instrument development will practically enhance the teaching practices in the essay writing course in particular and EFL teaching in general. We suggest that future research is directed to investigate the empirical evidence of employing this developed instrument, enacting other relevant research designs such as case studies or experimental research. Similar design-based research highlighting the process of developing assessment instruments for other language skills is also preferable to enrich the existing literature on teachers' assessment literacy development.

## REFERENCES

- Alderson, J. C., Brunfaut, T., & Harding, L. (2017). Bridging assessment and learning: A view from second and foreign language assessment. *Assessment in Education: Principles, Policy & Practice*, 24(3), 379-387. <https://doi.org/10.1080/0969594x.2017.1331201>
- Badger, R., & White, G. (2000). A process genre approach to teaching writing. *ELT Journal*, 54(2), 153-160. <http://doi.org/10.1093/elt/54.2.153>
- Bloxham, S., den-Outer, B., Hudson, J., & Price, M. (2016). Let's stop the pretence of consistent marking: Exploring the multiple limitations of assessment criteria. *Assessment & Evaluation in Higher Education*, 41(3), 466-481. <http://dx.doi.org/10.1080/02602938.2015.1024607>
- Blumberg, B., Cooper, D. R., & Schindler, P. S. (2005). *Business research methods*. McGraw-Hill Education.
- Brown, H. D. (2004). *Language assessment: Principles and classroom practices*. Pearson Education.
- Brown, H. D., & Abeywickrama, P. (2019). *Language assessment: Principles and classroom practice* (3<sup>rd</sup> ed.). Pearson Education Inc.
- Brown, J. D. (1995). *The element of a language curriculum*. Heinle & Heinle.
- Callaghan, M., & Knapp, P. (1989). *The discussion genre*. Metropolitan East Disadvantaged School Program.
- Caplan, N. A., & Farling, M. (2017). A dozen heads are better than one: Collaborative writing in genre-based pedagogy. *TESOL Journal*, 8(3), 564-581. <https://doi.org/10.1002/tesj.287>
- Cavallaro, C. J., & Sembiante, S. F. (2021). Facilitating culturally sustaining, functional literacy practices in a middle school ESOL reading program: A design-based research study. *Language and Education*, 35(2), 160-179. <https://doi.org/10.1080/09500782.2020.1775244>
- Creswell, J. W. (2005). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*. Pearson Merrill Prentice Hall.
- Davies, A. (2008). Textbook trends in teaching language testing. *Language Testing*, 25(3), 327-347. <https://doi.org/10.1177/0265532208090156>
- DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016). Approaches to classroom assessment inventory: A new instrument to support teacher assessment literacy. *Educational Assessment*, 21(4), 248-266. <https://doi.org/10.1080/10627197.2016.1236677>
- DiDonato-Barnes, N., Fives, H., & Krause, E. S. (2013). Using a table of specifications to improve teacher-constructed traditional tests: An experimental design. *Assessment in Education: Principles, Policy & Practice*, 21(1), 90-108. <https://doi.org/10.1080/0969594X.2013.808173>
- Effendi, T., & Mayuni, I. (2022). Examining teacher-made English test in a language school. *LADU: Journal of Languages and Education*, 2(2), 67-76. <https://doi.org/10.56724/ladu.v2i2.109>
- Englert, C. S., Raphael, T. E., Anderson, L. M., Anthony, H. M., & Stevens, D. D. (1991). Making strategies and self-talk visible: Writing instruction in regular and special education classrooms. *American Educational Research Journal*, 28(2), 337-372. <https://doi.org/10.3102/00028312028002337>
- Field, A. P. (2005). *Discovering statistics using SPSS*. Sage Publications Inc.

- Firkins, A., Forey, G., & Sengupta, S. (2007). Teaching writing to low proficiency EFL students. *ELT Journal*, 61(4), 341-352. <http://dx.doi.org/10.1093/elt/ccm052>
- Fitriyah, I., Basthomi, Y., Khotimah, K., & Gozali, I. (2022). Implementation of assessment for learning in online EFL writing class: A case of novice undergraduate teachers. *LEARN Journal: Language Education and Acquisition Research Network*, 15(2), 129-159.
- Fitriyah, I., Gozali, I., Widiati, U., El Khoiri, N., & Singh, A. K. J. (2023). EFL writing teachers' practices and values of assessment for and as learning in a constrained context. *Computer-Assisted Language Learning Electronic Journal*, 24(2), 104-128.
- Ghanbari, N., & Barati, H. (2020). Development and validation of a rating scale for Iranian EFL academic writing assessment: A mixed-methods study. *Language Testing in Asia*, 10, Article 17. <https://doi.org/10.1186/s40468-020-00112-3>
- Glušac, T., & Milić, M. (2022). Quality of written instructions in teacher-made tests of English as a foreign language. *English Teaching & Learning*, 46(1), 39-57. <https://doi.org/10.1007/S42321-021-00079-1>
- Hammond, J., Burns, A., Joyce, H., Brosnan, D., & Gerot, L. (1992). *English for special purposes: A handbook for teachers of adult literacy*. Macquarie University Press.
- Hermansson, C., Jonsson, B., Levlin, M., Lindhé, A., Lundgren, B., & Norlund Shaswar, A. (2019). The (non)effect of joint construction in a genre-based approach to teaching writing. *The Journal of Educational Research*, 112(4), 483-494. <https://doi.org/10.1080/00220671.2018.1563038>
- Hughes, A. (2008). *Testing for language learners*. Cambridge University.
- Hyland, K. (2007). Genre pedagogy: Language, literacy, and L2 writing instruction. *Journal of Second Language Writing*, 16(3), 148-164. <https://doi.org/10.1016/j.jslw.2007.07.005>
- Inbar-Lourie, O. (2008). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing*, 25(3), 385-402. <https://doi.org/10.1177/0265532208090158>
- Ing, L. M., Musah, M. B., Al-Hudawi, S. H. V., Tahir, L. M., & Kamil, N. M. (2015). Validity of teacher made assessment: A table of specification approach. *Asian Social Science*, 11(5), 193-200. <http://dx.doi.org/10.5539/ass.v11n5p193>
- Janatifar, M., & Marandi, S. S. (2018). Iranian EFL teachers' language assessment literacy (LAL) under an assessing lens. *Applied Research on the English Language*, 7(3), 361-382. <https://doi.org/10.22108/are.2018.108818.1221>
- Jannati, S. (2015). ELT teachers' language assessment literacy: Perceptions and practices. *The International Journal of Research in Teacher Education*, 6(2), 26-37.
- Jeong, H. (2017). Narrative and expository genre effects on students, raters, and performance criteria. *Assessing writing*, 31, 113-125. <https://doi.org/10.1016/j.asw.2016.08.006>
- Lee, I. (2017). *Classroom writing assessment and feedback in L2 school contexts*. Springer Singapore.
- Leong, W. S., Ismail, H., Costa, J. S., & Tan, H. B. (2018). Assessment for learning research in East Asian countries. *Studies in Educational Evaluation*, 59, 270-277. <https://doi.org/10.1016/j.stueduc.2018.09.005>

- Livingston, S. A. (2018). *Test reliability-basic concepts* [Research Memorandum ETS RM-18-01]. Educational Testing Service.
- Minda, M. H. (2018). Content validity of EFL teacher-made assessment: The case of communicative English skills course at Ambo University. *East African Journal of Social Sciences and Humanities*, 3(1), 41-62. <https://doi.org/10.20372/eajssh.v3i1.470>
- Mukminatien, N., Yaniafari, R. P., El Khoiri, N., & Ariyanto, M. S. A. (2021). The rater reliability of an argumentative writing assessment using genre-based scoring rubric. In N. Hayati, Suharyadi, S. Andreani, & U. P. Astuti (Eds.), *The changing face of ELT: A festschrift for Prof. Ali Saukah and Prof. M. Adnan Latief* (pp. 175-191). Universitas Negeri Malang.
- Nation, I. S. P., & Macalister, J. (2010). *Language curriculum design*. Routledge.
- Oluwatayo, J. (2012). Validity and reliability issues in educational research. *Journal of Educational and Social Research* 2(2), 391-400.
- O'Neill, P., & Murphy, S. (2017). Postsecondary writing assessment. In C. Secolsky & D. B. Denison (Eds.), *Handbook on measurement, assessment, and evaluation in higher education* (pp. 586-604). Routledge.
- Oshima, A., & Hogue, A. (1998). *Writing academic English*. Longman.
- Oshima, A., & Hogue, A. (2007). *Introduction to academic writing*. Pearson Education.
- Pallant, J. (2011). *A step-by-step guide to data analysis using the SPSS program: Survival manual*. Open University Press.
- Pesiwarissa, R., Anaktototy, K., & Maruanaya, H. J. (2021). An analysis of EFL teacher-made tests. *Matai International Journal of Language Education*, 2(1), 1-13. <https://doi.org/10.30598/matail.v2i1.5485>
- Pill, J., & Harding, L. (2013). Defining the language assessment literacy gap: Evidence from a parliamentary inquiry. *Language Testing*, 30(3), 381-402. <https://doi.org/10.1177/0265532213480337>
- Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory Into Practice*, 48(1), 4-11. <https://doi.org/10.1080/00405840802577536>
- Razavipour, K., Riazi, A., & Rashidi, N. (2011). On the interaction of test washback and teacher assessment literacy: The case of Iranian EFL secondary school teachers. *English Language Teaching*, 4(1), 156-161. <https://doi.org/10.5539/elt.v4n1p156>
- Robson, C. (2011). *Real world research: A resource for users of social research methods in applied settings* (2<sup>nd</sup> ed.). John Wiley and Sons Ltd.
- Santi, E., Setiawan, M. H., & Khumaedi, M. (2019). Developing assessment instrument on writing on descriptive paragraph based on picture of junior high school students. *Journal of Research and Educational Research Evaluation*, 8(2), 116-122. <https://doi.org/10.15294/jere.v8i2.36086>
- Schedl, M., & Malloy, J. (2014). Writing items and tasks. In A. J. Kunnan (Ed.), *The companion to language assessment* (pp. 796-813). Wiley. <https://doi.org/10.1002/9781118411360.wbcla025>
- Setiabudi, A., Mulyadi, M., & Puspita, H. (2019). An analysis of validity and reliability of a teacher-made test. *Journal of English Education and Teaching*, 3(4), 522-532. <https://doi.org/10.33369/jeet.3.4.522-532>

- Shepard, L. A. (2000, February). *The role of classroom assessment in teaching and learning* [CSE Technical Report 517]. National Center Research on Evaluation, Standards, and Student Testing University of California.
- Smalley, R. L., Ruetten, M. K., & Kozyrev, J. R. (2001). *Refining composition skills. Rhetoric and grammar*. Heinle & Heinle.
- Stiggins, R. (1991). Assessment literacy. *Phi Delta Kappan*, 72(7), 534-539. <https://www.jstor.org/stable/20404455>
- Stiggins, R. (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan Magazine*, 83(10), 758-765. <https://doi.org/10.1177/003172170208301010>
- Suharyadi, S. (2021). Genre and text type. In N. Hayati, Suharyadi, S. Andreani, & U. P. Astuti (Eds.), *The changing face of ELT: A festschrift for Prof. Ali Saukah and Prof. M. Adnan Latief* (pp. 135-147). Universitas Negeri Malang.
- Tsagari, D., & Vogt, K. (2017). Assessment literacy of foreign language teachers around Europe: Research, challenges, and prospects. *Language Testing and Assessment*, 6(1), 41-63. <https://doi.org/10.58379/UHIX9883>
- Ugwu, N. G., & Mkpuma, S. O. (2019). Ensuring quality in education: Validity in teacher-made language tests in secondary schools in Ebonyi state. *American Journal of Educational Research*, 7(7), 518-523.
- Wu, X. M., Zhang, L. J., & Dixon, H. R. (2021). Implementing assessment for learning (AfL) in Chinese university EFL classes: Teachers' values and practices. *System*, 101, 102589. <https://doi.org/10.1016/j.system.2021.102589>
- Xu, Y., & Brown, G. T. L. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, 58, 149-162. <https://doi.org/10.1016/j.tate.2016.05.010>

## APPENDIX

### Comparison and Contrast Essay Writing Prompt

|         |               |
|---------|---------------|
| Name :  | _____         |
| Course: | Essay Writing |
| Date :  | _____         |
| Time :  | 90 minutes    |

|        |
|--------|
| Score: |
|--------|

#### Direction

Suppose that you had a 2-week vacation last month. You visited many places such as tourist attractions, markets, and cities. Having back from your vacation, you were emailed by your friend to recommend the best place to visit. Therefore, write down an essay comparing and contrasting two places to help your friend decide where to visit. You may choose one of the following topics:

- Two tourism attractions
- Two markets
- Two cities

Be sure to include an engaging introduction, a clear thesis, well-defined and detailed paragraphs, and a fitting conclusion that ties everything together. You may use point-by-point or block organization.

Your essay will be evaluated by the following aspects: generic structure of a comparison and contrast essay, language use (vocabulary and grammatical structure), sentence construction and transactional signals, and mechanics (spelling, punctuation, and capitalization)

You have 90 minutes to write your 400-500 words essay without looking up the dictionary. Use the following template.

|                                                                |       |
|----------------------------------------------------------------|-------|
|                                                                | Title |
|                                                                |       |
|                                                                |       |
|                                                                |       |
|                                                                |       |
|                                                                |       |
| Introductory paragraph                                         |       |
|                                                                |       |
|                                                                |       |
|                                                                |       |
|                                                                |       |
|                                                                |       |
| Body paragraph 1                                               |       |
|                                                                |       |
|                                                                |       |
|                                                                |       |
|                                                                |       |
|                                                                |       |
| Body paragraph 2                                               |       |
|                                                                |       |
|                                                                |       |
|                                                                |       |
|                                                                |       |
|                                                                |       |
| Body paragraph 3 (Not necessary if you use Block Organization) |       |
|                                                                |       |
|                                                                |       |
|                                                                |       |
|                                                                |       |
| Concluding paragraph                                           |       |