

## **Design, Construction, Grading of Essay Questions 1.0 for Teachers**

**Dr. Patricia E. Allanson**

Liberty University  
United States of America

**Dr. Charles E. Notar**

Jacksonville State University  
United States of America

### **Abstract**

---

*The two types of essays and the construction rules, key terms, and how to grade guidance is provided for classroom teachers. The information will provide a reliable essay question that will assure a students' ability to demonstrate their comprehension/cognition in the content area being tested.*

---

**Key words:** (Construction/development/design/grading scoring) Restricted response, Short answer, Extended response, Open ended (test items/questions); Holistic/global/analytic scoring.

### **Introduction**

Students take an array of assessments throughout their school career to determine grades, promotion, graduation, and learning progress (Salend, 2009). Essay type assessment questions have been around long before the surge of high-stake multiple-choice standardized tests, and are a major part of formal education. Prior to the origin of the standardized test era (1900-1930), assessments were handled through either an oral examination or in an essay format. The Greek philosopher Socrates (469-399 BCE) tested his students through conversation while 7<sup>th</sup> century Imperial China government job applicants submitted written essays addressing Confucian philosophy. In today's world, written essays are seen not only in academia, but in magazines, newspapers, and employment applications where applicants are asked to describe prior job experience. Essays also play a role in the college admission process providing for a "story behind the transcript" (Hoover, 2013) depicting character traits of the candidate in which GPA and test results cannot provide.

Like other type assessment items such as multiple-choice questions, essays are used as a method to evaluate student performance in the academic world. Essay test items are the easiest to construct, however, they are often more difficult to score. Due to this factor, many educators have veered away from using essays as a way to assess student learning. Students also assert their dislike when faced with essay type questions and cite complaints such as "You don't give us enough space to write our response", "The questions are like you're trying to trick us", "The directions were confusing" (Salend, 2011). Teachers who want to assess upper levels of Bloom's Taxonomy often use the essay test format where students are asked to select, organize, analyze, synthesize, or evaluate information (Kelly, 2019), however, creating valid and reliable tests that measure higher level thinking is challenging. Salend (2011) asserts that poorly designed tests can have a negative affect on academic growth, therefore there is a need for teachers who may not be aware of how to write effective essay questions that initiate higher levels of learning (Linton, et al., 2014). The objective of this article is to provide the basics of training in essay development and scoring.

**Uses of essays**

Derived from a Latin word *exagium*, essay roughly translates to presenting one’s case. In the academic world, an essay is defined as "A test item which requires a response composed by the examinee, usually in the form of one or more sentences, of a nature that no single response or pattern of responses can be listed as correct, and the accuracy and quality of which can be judged subjectively only by one skilled or informed in the subject." (Stalnaker, 1951, as cited by Reiner, et al., 2002, p. 6). The central purpose of the essay type test item is to measure students’ ability to select; organize; integrate; relate; and evaluate related ideas as a whole, and demonstrate complete understanding of complex learning objectives by applying facts and principles or interpreting cause-effect relationships. Essay type items are not used to test recall of facts. Figure 1 provides a visual of when to use the various types of test questions in regard to Bloom’s taxonomy. Stalnaker (1951) proposed that essay questions should align with four criteria:

1. Requires written composition rather than selected response.
2. The written response must be more than one sentence.
3. Allows for different, original or pattern of responses, and
4. Subjective scoring that judges the quality of responses.

An essay, in a sense, is a riddle that students can solve using prior knowledge, making connections, and drawing conclusions. This is an invaluable skill for any student to have. Those who can master it have a better chance of being successful academically. To that end the authors consider the essay as a determinant of A/B grades.

**Advantages of essays**

There are several advantages of using essays to measure higher level cognitive learning. As mentioned, essay items directly assess complex higher-level objectives and take considerably less time to construct and administer than other item types. They are adaptable to a wide variety of subjects; provide methods for students to organize and present ideas in a logical manner; and challenges students to create responses rather than choose from a list of predetermine responses thus eliminating the problem of guessing (Reiner, et al., 2002; Clay, 2001; Greifeneder, Scheibehenne, Kleber, 2010).

Research has also concluded that when students are aware that they will be completing essay type questions, they seem to employ better study habits and techniques (Gupta, Jain, & D’Souza, 2016). Students are also allowed freedom of thought and produce original responses which shows all levels of comprehensiveness or accuracy.

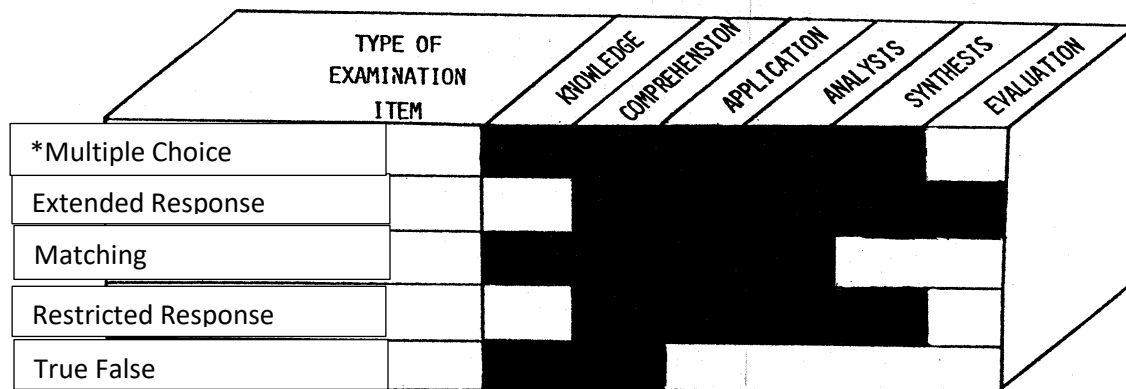


Figure 1: Types of exam questions and Bloom’s taxonomy (Notar, Beard, & Akpan, 2018).

**Disadvantages/Limitations of essays**

Since essay type items evaluate more complex cognitive skills at a deeper level, they are limited to a small sampling of content or learning objectives. Constructing responses takes longer and is influenced by several extraneous factors such as the students writing skills, fluency, handwriting, speed, etc. If the question is worded ambiguously, the student may become confused as to the direct implication of the question and their response may not provide a true sample of their understanding.

Another disadvantage of essay type items is that they are difficult and time consuming to grade in a reliable manner (Reynolds, et al., 2006). Most often, essays are subjectively graded by individual teachers (teacher made tests), or by two or more professionals/raters (standardized tests). Individuals scoring tests must be wary of their current mood, bias, and potential halo effects (knowing a student very well and having a preconceived positive or negative perspective, Malouff, et al., 2014) which reduces equity and fairness.

### **Avoiding Common Misconceptions**

Many educators believe that essay questions automatically assess higher-order thinking skills, however, this will depend on how the question is written and even in the ways that responses are graded. In order to assess at a higher level, questions must be written so that they require more from students rather than simple recall. Another misconception is that essay questions eliminates guessing answers. Bluffing, also referred as “vague generalities, padding, name dropping, etc.” (Reiner et. al, 2002, p. 13) is just another form of guessing thus the nature changes but does not eliminate it.

### **Essay Key Terms**

To determine when essay questions should be used, it is important to review the action verbs associated with the learning objectives or intended learning outcomes. There are any number of verbs that can be used when writing an essay question. The list below provides the more common verbs used across Bloom’s, et al. (1956) taxonomy:

List: simplest form of essay, maybe considered a completion test item. It requires recall of factual material.

Arrange: Requires recall, but also a given order of recalled facts that need to be specified.

Select: List items that belong in a category. It requires simple evaluation to pick the most important, least valuable, or something that fits under a certain grouping such as presidents

Describe: Give the important characteristics of an object, process, or phenomenon. Usually requires a paragraph, not a single sentence or definition.

Discuss: Must provide full elaboration of the subject tracing development, arguments for and against, and perhaps relationships to other facts or ideas.

Define: To place in a category and distinguish from everything else. This is difficult, most students describe or discuss.

Illustrate: To illustrate is to give an example of a definition, law, principle, etc.

Explain: Places special emphasis upon cause and effect relationships.

Compare: May or may not state a specific basis for comparison, but will call for the advantages and disadvantages of two ideas, their similarities and differences, etc.

Summarize: To state in as brief form as possible the essentials of a unit of material. Another term is review.

Outline: Is related to summarize, but usually implies organization into main topics and subheadings, whereas summarizing may be done in a paragraph.

Interpret: An extension of define when meaning of a rather profound or obscure quotation is desired. May call for practical application of a general principle.

Criticize: To evaluate the correctness or adequacy of an idea, usually giving suggestions for improving it or reasons why it should be abandoned, is the way examinee usually criticizes.

Formulate new problems or procedures: Originality and ability to think creatively with the tools and materials of a given subject matter field makes this generally the most difficult of all types of essays.

### **Extended and Restricted Response Essays**

There are overall two categories of test items: Selection test items such as multiple-choice questions where the respondent must select the correct answer from a supplied list; and supply type items where the respondent is responsible for supplying the answer. Essay questions fall in the supply test items category and include two different types: Extended response (open ended) and Restricted response (short answer). Restricted response items limit student responses based on the wording of the question whereas extended responses place no limits on student response where they are “free to use their own judgement and integrate outside information to help support their opinion” (Kelly, 2019, p. 1).

Restricted response, also referred to as short answer questions, are organized in a manner that specifies or limits student responses. They typically assess learning objectives at the lower levels of Bloom’s taxonomy and require students to list, define, or describe (Reynolds, et al., 2006).

The restricted response question usually limits both the content and the response. The content is usually restricted by the scope of the topic to be discussed such as in basing questions on specific problem thus limiting the nature of the response.

Extended response items pave the way for more organization, creativity and depth of knowledge. They can be used to assess higher-levels of Bloom’s taxonomy such as in describing application of a set of principles, explaining cause and effect relationships, formulating conclusions, or to produce, organize, and express ideas (see Figure 2). According to Reynolds, et al. (2006), extended response type items additionally can be used as a way to assess grammar and writing skills as these types of items require coherent writing skills.

Figure 2 provides guidance on when to use key terms with which type of essay.

<u>Key term</u>	<u>Type Essay</u>
List	Restrictive
Arrange	Restrictive
Select	Restrictive
Describe	Both
Discuss	Extended
Define	Restrictive
Illustrate	Restrictive (possible both)
Explain	Both
Compare	Both
Summarize	Restrictive (possible both)
Outline	Restrictive
Interpret	Extended
Criticize	Extended
Formulate new problems or procedures	Extended

Figure 2: Key terms and types of essays (Notar, 2003).

### Essay Construction

The following guidelines for writing essay questions are an amalgamation of a number of websites and various authors. The literature is old and scarce on rules which lead the authors to use the internet where a number of colleges and universities provide guidance. Regardless of the age of publication dates the rules are still valid today and will be in the future (see Construction references).

As with all types of assessments, constructing essay items should always start with the behavioral objectives (Mager, 1975; Gagne, & Briggs, 1979) and Table of Specifications (Notar, et al., 2004). The objectives and table will ensure test creation will cover the main topics, concepts, and skills teachers taught during the time preceding the test. Test items should be written for a specific objective or focus, and restricted in use as appropriate or when no other item type can be used. Essay items also need to cover appropriate amounts of material (Salend, 2009), and are void of vague terms such as for example, frequently, usually (Brookhart & Nitko, 2008; Elliott et al., 2010). Directions on how students will need to complete the essay response should concisely and clearly state expectations including the length, purpose, time limits, and if points will be deducted for incorrect spelling, poor grammar usage, or illegible penmanship (Salend, 2011). The language used in the directions should also be aligned with the same language used during instruction. Kelly (2019) also suggests writing questions so that they are more interesting or novel to students to help in motivation.

Piontek (2008) provides a simple list of six guidelines for developing essay items as compiled from previous research:

1. Restrict the use of essay questions to educational outcomes that are difficult to measure using other formats.
2. Construct the item to elicit skills and knowledge in the educational outcomes.
3. Write the item so that students clearly understand the specific task.
4. Indicate the amount of time and effort students should spend on each essay item.

5. Avoid giving students options as to which essay questions they will answer.
6. Consider using several narrowly focused items rather than one broad item. (pp.6-7)

Other factors to consider in constructing essay questions include adjusting linguistics, readability and legibility to make sure it is appropriate for students. Salend (2009, 2011) suggests eliminating unnecessary words; reducing the length of sentences; using familiar language referenced during instruction, and avoiding the use of double negatives, ambiguous terms, abbreviations, contractions, acronyms, quotations, and parentheses. Additionally, she recommends choosing visual designs that increase legibility such as consistent and familiar type size and fonts. Research has shown that carefully constructed essay questions have the power to initiate problem solving, encourage original perspectives, and deepen complex knowledge structures (Bierer, et al., 2009). While constructing essay questions, keep in mind two important issues - an appropriate grading system and the criteria that will be used to evaluate the responses (Bacha, 2001).

### **Grading**

Once construction is completed and responses submitted, it is now time to embark on the arduous task of grading. While students spend minutes writing their response, instructors can spend hours grading which can be a serious problem (Carlson, 2013). Regardless of which type of essay grading you chose there are a number of rules you need to consider and possibly implement. The following guidelines for grading essay questions are an amalgamation of a number of current websites and authors (see grading references). The authors definitely agree with Reiner, et al. (2002) that a knowledgeable colleague should critically review the essay question, the model answer, and the intended learning outcome for alignment. This is good advice for any test.

Since an essay is a performance based test, there are several components to consider including the initial task itself constructed by the writer, submitted by the examinee, and then graded by the rater(s), using a rating/grading system (Hamp-Lyons, 2003). The first step in grading is selecting a grading system which may be predetermined during test item construction. Grading essays falls into two types: *holistic/global and analytic grading*. Nakamura (2004) provides a simple table (Table 1) that compares holistic and analytic grading.

Holistic grading is mainly used for “impressionistic evaluation” (Bacha, 2001, p. 374) and scored using letter grades, percentages, or ordinal scales corresponding to a set of predetermined criteria. Hughes (1989) defines holistic grading as the process of assigning a single score for a writing sample based on an overall impression of the quality of it. For holistic grading, the rater will read for an overall impression, for both the content as well as the quality of the writing, and then translate his or her impressions into a score or letter grade. A final grade or score is established as the average of two or more raters. Many researchers have argued that holistic grading focuses on too much on what the writer does well rather than on areas of weakness, and therefore does not provide for diagnostic information. Holistic grading serves its purpose for scoring on a large-scale basis, and identifying overall writing proficiency, however, analytic scoring would be best suited for providing feedback on specific aspects of writing skills, progress in these skills, and for promotion purposes.

Hughes (1989) defines analytic scoring as a method of grading that requires a separate score for each of a number of aspects of a task or subcategories, such as grammatical accuracy, vocabulary, idiomatic expression, organization, relevance, coherence. The grader breaks each question down to its different parts and identifies the major points covered by the students answer as indicated on predetermined criteria. The grader will then assign points to each section depending on the value they feel each deserves. This will ensure crucial elements of the answer are identified and scored more or less separately.

Table 1. A comparison of holistic/global and analytic scales in terms of six qualities of test usefulness. (adapted from Weigle 2002, p.121)

Quality	Holistic Scale	Analytic Scale
Reliability	lower than analytic but still acceptable	higher than holistic
Construct Validity	holistic scale assumes that all relevant aspects of writing ability develop at the same rate and can thus be captured in a single score; holistic scores correlate with superficial aspects such as length and handwriting	analytic scales more appropriate for L2 writers as different aspects of writing ability develop at different rates
Practicality	relatively fast and easy	time-consuming; expensive
Impact	single score may mask an uneven writing profile and may be misleading for placement	more scales provide useful diagnostic information for placement and/or instruction; more useful for rater training
Authenticity	White (1995) argues that reading holistically is a more natural process than reading analytically	raters may read holistically and adjust analytic scores to match holistic impression
Interactiveness*	n/a	n/a

A rubric is often recommended in the literature for evaluating using the analytic scoring method. Rubrics provide the overall expectations and criteria (arranged in levels) for the test item, and the potential points that can be earned based on each level. Brown and Bailey (1984) introduced an analytic scoring scale that specified five major categories: organization, logical development of ideas, grammar, punctuation/spelling/mechanics, and style and quality of expression; and five different levels in each category ranging from unacceptable to excellent.

Essay tests (n.d.) provides a simple way to determine which grading process to use. If you are evaluating the essay as a whole and rating it against other students you would use holistic grading, whereas with analytic grading you will list specific criteria to be met and award points based on their inclusion. Regardless of the grading method Haluska (2012) there are five qualities that should be considered:

1. **Unity:** Does every part of the **essay** reinforce the thesis statement? Do any words, sentences, or paragraphs drift off the opinion that the writer is trying to prove?
2. **Support:** Does each paragraph contain clear explanation and actual evidence from real life - anecdotal, statistical, etc.?
3. **Organization:** Does each point/paragraph come exactly where it should to lead the reader along an effective train of thought? Does every interior paragraph begin with a transition that explains why that point needs to come where it does? ("Next," or "Then we have," or "Lastly" don't do that, since an expository **essay** is seldom chronological. Phrases like "More important," "Even more surprising," or "As a result," give reasons for paragraph placement, and are better transitions.) However, we should caution students that transitions are ideas, not words.
4. **Style:** Is the wording vivid, clear, and above all, succinct?
5. **Mechanics:** Are grammar, spelling, and punctuation correct? Don't proofread. Notice patterns or repeated errors (in content or form). Make comments in pencil on the students' answers. Teacher comments make essay tests a good learning experience (p. 30).

Once a grading method is determined a best practice is to compose the perfect answer to check for adequacy against the grading criteria and to identify what constitutes excellent and unacceptable responses. Students should always be informed of the type of grading method used and provided with a copy of the grading criteria or rubric. During the grading process, many authors suggest using a systematic process including scoring all responses to one question before proceeding to the next question (reducing the halo effect); grading papers anonymously using student IDs or last four numbers of social security; and grading randomly to increase accuracy of the scoring (Gronlund & Linn, 1990; Wiggins, 1998; Worthen, et al., 1993; McMillan, 2001; Piontek, 2008). It is also recommended to recheck papers periodically when grading a large number to ensure scores are not shifting appreciably. Finally, Tierney (2013) suggests being cognizant of what you may see as sheer drudgery and tedium of grading essays.

Clay (2001) recommends providing brief comments and corrections so that students have a basis for understanding their mistakes and provide for a good learning experience. The comments should encourage or affirm good ideas, but firm in suggesting areas of improvement. The comments should be viewed as a teaching tool used to address strengths and weaknesses rather as a way to “pick apart” students work. Final comments should guide students in the inquiry process. If a pattern of common mistakes occurs, it is recommended to review these as a whole class rather than writing comments on each paper

As discussed previously, a limitation of essay tests is that they lack in reliability due to subjective grading practices. Therefore, it is important to try and be as reliable as possible when scoring your essay items. Kelly (2019) provides a few tips to help improve reliability in grading:

1. Prepare the essay rubric in advance.
2. Avoid looking at names.
3. Score one item at a time.
4. Avoid interruptions when scoring a specific question.
5. If an important decision like an award or scholarship is based on the score for the essay, obtain two or more independent readers.
6. Beware of negative influences that can affect essay scoring. These include handwriting and writing style bias, the length of the response, and the inclusion of irrelevant material.
7. Review papers that are on the borderline a second time before assigning a final grade”, (pp. 2- 3).

Low reliability and high cost of human essay grading have initiated an interest in automated essay scoring (AES) and are used in many high-stakes testing. Although AES produces reliable scores similar to human scores, they are limited in evaluation procedures that assess higher-order thinking skills (Attali, et al., 2013). The authors contend that requiring a student to write an essay requires the classroom teacher to score the essay, not a machine. There is no better way to be in touch with your students outside of one on one contact than reading a student’s thoughts.

## **Conclusion**

In reviewing the literature you will see that it says that essay tests are one of the easiest tests to construct. Following the above guidelines/rules for the construction and grading of essay test items provides a reliable and valid method to measure student achievement, however, this is only if students know how to take them. Since many college freshmen come to campus ill prepared to take essay tests, instructors must prepare students in such endeavors. This can be accomplished by simply posting Essay Key Words in a prominent place in your classroom. Also, Kapinus (2014) provides one of the best guides for teaching students how to answer extended response test items which can be found at: <http://textproject.org/library/text-matters/preparing-students-in-writing-responses-to-open-ended-questions/>

## References Body of Article

- Bierer, S. B., Taylor, C. A., & Dannefer, E. F. (2009). Evaluation of essay questions used to assess medical students' application and integration of basic and clinical science knowledge. *Teaching and Learning in Medicine, 21*(4), 344-350.
- Bloom, B. S. (Ed.). Engelhart, M. D., Furst, E. J., Hill, W. H., Krathwohl, D. R. (1956). *Taxonomy of Educational Objectives, Handbook I: The Cognitive Domain*. New York: David McKay, Co. Inc.
- Brookhart, S. M., & Nitko, A.J. (2008). *Assessment and grading in classrooms*. Columbus, OH: Merrill/Pearson Education.
- Cashin, W. E. (1987). *Improving essay tests*. Idea Paper, No. 17. Manhattan, KS: Center for Faculty Evaluation and Development, Kansas State University.
- Clements, T. (2013). Tutoring announces a three-pronged approach to acing the SAT writing test. *PR Newswire US*. 11/21/2013.
- Clay, B. (2001). *Is this a trick question? A short guide to writing effective test questions*. Kansas Curriculum Center. file:///C:/Users/Charles%20Notar/Desktop/Documents/Downloads/Handout-Module6(1).pdf
- ED400fall09 (2002). *Essay - restricted response*. (n.d.). Retrieved from <https://ed400fall09.wikispaces.com/Essay+++Restricted+Response>
- Elliot, S. M., Kettler, R. J., Beddow, P. A., Kurz, A., Compton, E., McGrath, d., Bruen, C., Hinton, K., Palmer, P., Rodriguez, M. C., Bolt, D., & Roach, A. T. (2010). Effects of using modified items to students with persistent academic difficulties. *Exceptional Children, 76*, 475-495.
- Essay Items Hope You Can Write* (2005). (n.d.). Retrieved from [http://www.sagepub.com/upm-data/6429\\_Chapter\\_7\\_\\_Salkind\\_\(Tests\)\\_I\\_Proof\\_3.pdf](http://www.sagepub.com/upm-data/6429_Chapter_7__Salkind_(Tests)_I_Proof_3.pdf)
- Gagne, R. M & Briggs, L. (1979). *Principles of instructional design*. Holt, Rinehart and Winston.
- Greifeneder, R., Scheibehenne, B., & Kleber, N. (2010). Less may be more when choosing is difficult: Choice complexity and too much choice. *Acta Psychologica, 133*(1), 45-50. doi:10.1016/j.actpsy.2009.08.005
- Guidelines for test design and construction*. (n.d.). Retrieved from <http://homepages.rpi.edu/~verwyc/chap2tm.htm>
- Gronlund, N. E., & Linn, R. L. (1990). *Measurement and evaluation in teaching*. (6th ed.). New York: Macmillan Publishing Company.
- Gupta, C., Jain, A., & D'Souza, A. (2016). Essay versus multiple-choice: A perspective from the undergraduate student point of view with its implications for examination. *Gazi Medical Journal., 27*(1), 8–10. <https://doi.org/10.12996/gmj.2016.03>
- Heller, S. & Stacy, J. (2013). Building reading, writing and analysis in the AP U.S. history classroom. *Social Education, 77*(5), 256-259.
- Holistic vs. analytic scoring of writing*. (n.d.). Retrieved from [http://personal.uncc.edu/medomoto/4200/writing/holistic\\_writing.htm](http://personal.uncc.edu/medomoto/4200/writing/holistic_writing.htm)
- Hoover, E. (2013). For the common application, new essay prompts and more words. *Chronicle of Higher Education, 59*(23), A11-A11.
- Hoover, E. (2013). Two, three essays? More can mean less. *The New York Times*. Retrieved from <https://www.nytimes.com/2013/04/14/education/edlife/extra-essay-requirements-on-college-applications-can-discourage-candidates.html>
- Kapinus, B. (2014). *Preparing students in writing responses to open-ended questions*. Retrieved from <http://textproject.org/library/text-matters/preparing-students-in-writing-responses-to-open-ended-questions/>
- Kelly, M. (2019). *Essay tests: Creating and scoring essay tests*. Retrieved from <https://www.thoughtco.com/creating-scoring-essay-tests-8439>
- Linton, D. L., Pangle, W. M., Wyatt, K. H., Powell, K. N., & Sherwood, R. E. (2014). Identifying key features of effective active learning: The effects of writing and peer discussion. *CBE - Life Sciences Education, 13*(3), 469-477.
- Mager, R. (1975). *Preparing instructional objectives* (2nd Edition). Belmont, CA: Lake Publishing Co.
- Malouff, J. M., Stein, S.J., Bothma, L.N., Coulter, K., & Emmerton, A. J. (2014). Preventing halo bias in grading the work of university students. *Cogent Psychology, 1*(1). Doi: 10.1080/23311908.2014.988937
- Notar, C. E., Beard, L., & Akpan, J. (2018). Introduction to test item classifications. *International Journal of Social Science and Business, 3*(3), 6-10.



- Notar, C. E., Zuelke, D. C., Wilson, J. D., & Yunker, B. D. (2004). The table of specifications: Insuring accountability in teacher made tests. *Journal of Instructional Psychology*, 31(2), 115-129.
- Reiner, C. M., Bothell, T. W., Sudweeks, R. R., & Wood, B. (2002). *Preparing effective essay questions: A self-directed workbook for educators*. Stillwater, Oklahoma: New Forums Press.
- Reynolds, C. R., Livingston, R. B., & Willson V. L. (2006). "Chapter 9: The development and use of constructed-response items." *Measurement and Assessment in Education*. Boston: Pearson/Allyn & Bacon, 222-44.
- Salend, S. J. (2009). *Classroom testing and assessment for ALL students: Beyond standardization*. Thousand Oaks, CA: Corwin.
- Salend, S. J. (2011). Creating Student-Friendly Tests. *Educational Leadership*, 69(3), 52-58.
- Study Guides and Strategies. (n.d.). *Constructing essay exams*. Retrieved from <http://www.studygs.net/teaching/tsttak4a.htm>
- Terry, R. M. (1989). Teaching and evaluating writing as a communicative skill. *Foreign Language Annals*, 22(1), 43-54.
- Tierney, J. (2013). *Why teachers secretly hate grading papers*. Retrieved from <https://www.theatlantic.com/national/archive/2013/01/why-teachers-secretly-hate-grading-papers/266931/>
- Vanderbilt University Center for Teaching. (n.d.). *Grading student work*. Retrieved from <https://cft.vanderbilt.edu/guides-sub-pages/grading-student-work/>
- Verner, S. (n.d.). *3 Easy steps to grading student*. Retrieved from <http://busyteacher.org/8867-3-easy-steps-to-grading-student-essays.html>
- Walker, C, & Schmidt, E. (2004). *Smart tests: Teacher-made tests that help students learn*. Ontario, CN: Pembroke.

### References for Rules of Construction of Essay Questions

- Alam, S. (2012). *Measurement, Assessment, and Evaluation in Education*. Compiled by Dr JJ, FSG, UiTM, Retrieved from <http://drjj.uitm.edu.my/DRJJ/OBE%20FSG%20Dec07/OBEJan2010/DrJJ-Measure-assess-evaluate-ADPRIMA-n-more-17052012.pdf>
- Bacha, N. (2001). *Writing evaluation: What can analytic versus holistic essay scoring tell us?* Retrieved from [https://doi.org/10.1016/S0346-251X\(01\)00025-2](https://doi.org/10.1016/S0346-251X(01)00025-2)
- Bierer, S. B., Taylor, C. A., & Dannefer, E. F. (2009). Evaluation of essay questions used to assess medical students' application and integration of basic and clinical science knowledge. *Teaching and Learning in Medicine*, 21(4), 344-350.
- Brookhart, S. M., & Nitko, A.J. (2008). *Assessment and grading in classrooms*. Columbus, OH: Merrill/Pearson Education.
- Carlson, A. (2013). *Writing across the curriculum: The pros and cons of holistic grading*. Retrieved from <http://campusnorth.com/holistic-grading/>
- Dawson, D. (2006). *Short answer and short essay*. (n.d.). Retrieved from [http://www.uwo.ca/tsc/resources/pdf/short\\_answer.pdf](http://www.uwo.ca/tsc/resources/pdf/short_answer.pdf)
- Elliot, S. M., Kettler, R. J., Beddow, P. A., Kurz, A., Compton, E., McGrath, D., Bruen, C., Hinton, K., Palmer, P., Rodriguez, M. C., Bolt, D., & Roach, A. T. (2010). Effects of using modified items to students with persistent academic difficulties. *Exceptional Children*, 76, 475-495.
- Essay test (2011). Retrieved from <http://educationalevaluationhema.blogspot.com/2011/03/essay-test.html>
- Guidelines for the construction of essay test. (n.d.). Retrieved from <http://www.greencoffeeforslimming.com/it/component/k2/itemlist/user/71333>
- Help Teaching (n.d.). *How to write good test questions*. Retrieved from [http://www.helpteaching.com/about/how\\_to\\_write\\_good\\_test\\_questions/](http://www.helpteaching.com/about/how_to_write_good_test_questions/)
- How to make essay tests. (n.d.). Retrieved from [http://dante.udallas.edu/DallasDiocese/Assessment/traditional/essay/essays/essayitems\\_files/v3\\_document.htm](http://dante.udallas.edu/DallasDiocese/Assessment/traditional/essay/essays/essayitems_files/v3_document.htm)
- Kelly, M. (2019). *Essay tests: Creating and scoring essay tests*. Retrieved from <https://www.thoughtco.com/creating-scoring-essay-tests-8439>
- Livingston, R. B., & Willson, V. (2006). *Measurement and Assessment in Education*. Boston: Pearson, Allyn and Bacon, 221.
- Maxwell, A. (2010). Assessment strategies for a history exam, or, why short-answer questions are better than in-class essays. *History Teacher*, 43(2), 233-245.

- McKeachie, W. J., & Svinicki, M. D. (2006). Assessing, testing, and evaluating: Grading is not the most important function. In McKeachie's *Teaching tips: Strategies, research, and theory for college and university teachers* (12th ed., pp. 74-86). Boston: Houghton Mifflin Company.
- Measuring complex achievement with essay questions.* (n.d.). Retrieved from <https://www.scribd.com/document/127171147/CONSTRUCTING-ESSAY-QUESTIONS>
- Meador, D. (2016). *An extended response item can enhance student learning.* Retrieved from <https://www.thoughtco.com/what-is-an-extended-response-item-3194254>
- Notar, C. E. (2003). EFD 632 Assessment in Education. Jacksonville State University.
- Piontek, M. (2008). Best practices for designing and grading exams. *CRLT Occasional Paper No. 24*. Ann Arbor, MI. Center for Research on Learning and Teaching.
- Reiner, C. M., Bothell, T. W., Sudweeks, R. R., & Wood, B. (2002). *Preparing effective essay questions: A self-directed workbook for educators.* Stillwater, Oklahoma: New Forums Press.
- Reynolds, C. R., & McMillan, J. H. (2001). *Classroom assessment: Principles and practice for effective instruction.* Boston: Allyn and Bacon.
- Salend, S. J. (2009). *Classroom testing and assessment for ALL students: Beyond standardization.* Thousand Oaks, CA: Corwin.
- Salend, S. J. (2011). Creating Student-Friendly Tests. *Educational Leadership*, 69(3), 52-58.
- Stanford University. (n.d.). *Tips for grading.* Retrieved from [http://web.stanford.edu/dept/CTL/handouts/PDF/grading\\_papers.pdf](http://web.stanford.edu/dept/CTL/handouts/PDF/grading_papers.pdf)
- Study Guides and Strategies. (n.d.). *Constructing essay exams.* Retrieved from <http://www.studygs.net/teaching/tsttak4a.htm>
- Svinicki, M. D. (1999). Evaluating and grading students. In *Teachers and students: A sourcebook for UT- Austin faculty* (pp. 1-14). Austin, TX: Center for Teaching Effectiveness, University of Texas at Austin.
- University of Michigan (n.d.). *Best practices for designing and grading exams.* Adapted from M.E. Piontek (2008). Retrieved from [http://www.crlt.umich.edu/P8\\_0](http://www.crlt.umich.edu/P8_0)
- University of Washington (2017). *Constructing tests.* Retrieved from <http://www.washington.edu/teaching/constructing-tests/#essayquestions>
- Wiggins, G. P. (1998). *Educative assessment: Designing assessments to inform and improve student performance.* San Francisco: Jossey-Bass Publishers.
- Worthen, B. R., Borg, W. R., & White, K. R. (1993). *Measurement and evaluation in the schools.* New York: Longman.

### References for Rules of Grading of Essay Questions

- Attali, Y., Lewis, W., & Steier, M. (2013). Scoring with the computer: Alternative procedures for improving the reliability of holistic essay scoring. *Language Testing*, 30(1), 125-141.
- Bacha, N. (2001). *Writing evaluation: What can analytic versus holistic essay scoring tell us?* Retrieved from [https://doi.org/10.1016/S0346-251X\(01\)00025-2](https://doi.org/10.1016/S0346-251X(01)00025-2)
- Brown, J. D. & Bailey, K. M. (1984). A categorical instrument for scoring second language writing skills. *Language Learning*, 34, 21-24.
- Carlson, A. (2013). *Writing across the curriculum: The pros and cons of holistic grading.* Retrieved from <http://campusnorth.com/holistic-grading/>
- Clay, B. (2001). *Is this a trick question? A short guide to writing effective test questions.* Retrieved from <http://www.k-state.edu/ksde/alp/resources/Handout-Module6.pdf>
- Essay tests: Creating and scoring essay tests.* (n.d.). Retrieved from <https://www.thoughtco.com/creating-scoring-essay-tests-8439>
- Gronlund, N. E., & Linn, R. L. (1990). *Measurement and evaluation in teaching* (6th ed.). New York: Macmillan Publishing Company.
- Guidelines for grading an essay.* (n.d.) retrieved from [http://core.ecu.edu/hist/zipfk/guidelines\\_for\\_grading\\_an\\_essay.htm](http://core.ecu.edu/hist/zipfk/guidelines_for_grading_an_essay.htm)
- Haluska, J. (2012). The formula essay reconsidered. *Education Digest*, 7(4), 25-30.
- Hamp-Lyons, L. (2003). Writing teachers as assessors of writing. In Kroll, B. (Ed.), *Exploring the dynamics of second language writing* (pp.162-189). New York: Cambridge University Press.
- Hughes, A. (1989.) *Testing for language teachers.* Cambridge: Cambridge University Press.

- Hunter, D. M., Jones, R. M., & Randhawa, B. S. (2014). *Preparing students in writing responses to open-ended questions*. Retrieved from <http://www.textproject.org/library/text-matters/common-core-state-standards-and-assessment/preparing-students-in-writing-responses-to-open-ended-questions/>
- Kayapinar, U. (2014). Measuring essay assessment: Intra-rater and inter-rater reliability. *Eurasian Journal of Educational Research*, 57, 113-135.
- Kelly, M. (2017). *Tips to cut writing assignment grading time*. Retrieved from <https://www.thoughtco.com/cut-writing-assignment-grading-time-7854>
- Kelly, M. (2019). *Essay tests: Creating and scoring essay tests*. Retrieved from <https://www.thoughtco.com/creating-scoring-essay-tests-8439>
- McMillan, J. H. (2001). *Classroom assessment: Principles and practice for effective instruction*. Boston: Allyn and Bacon.
- Nakamura, Y. (2004). A comparison of holistic and analytic scoring methods in the assessment of writing. <http://hosted.jalt.org/pansig/2004/HTML/Nakamura.htm>
- Piontek, M. (2008). Best practices for designing and grading exams. *CRLT Occasional Paper No. 24*. Ann Arbor, MI. Center for Research on Learning and Teaching.
- Reiner, C. M., Bothell, T. W., Sudweeks, R. R., & Wood, B. (2002). *Preparing effective essay questions: A self-directed workbook for educators*. Stillwater, Oklahoma: New Forums Press.
- Tierney, J. (2013). *Why teachers secretly hate grading papers*. Retrieved from <https://www.theatlantic.com/national/archive/2013/01/why-teachers-secretly-hate-grading-papers/266931/>
- University of California Berkeley. (2017). *Grading essays*. Retrieved from <http://gsi.berkeley.edu/gsi-guide-contents/student-writing-intro/grading/>
- Weigle, S. C. (2002). *Assessing writing*. Cambridge: Cambridge University Press.
- Wiggins, G. P. (1998). *Educative assessment: Designing assessments to inform and improve student performance*. San Francisco: Jossey-Bass Publishers.
- Worthen, B. R., Borg, W. R., & White, K. R. (1993). *Measurement and evaluation in the schools*. New York: Longman.
- Wollack, J. A. (2003). *Helpful tips for creating reliable and valid classroom tests: Writing and scoring essay and short-answer questions*. Retrieved from <https://testing.wisc.edu/LL11-03.pdf>
- Wiseman, C. S. (2012). A comparison of the performance of analytic vs. holistic scoring rubrics to assess L2 Writing. *Iranian Journal of Language Testing*, 2(1), 59-92.