# Developing Rating Scale Descriptors for Assessing the Stages of Writing Process: The Constructs Underlying Students' Writing Performances

**Parviz Maftoon** [*]
*Professor of TEFL, Islamic Azad University, Science and Research Branch, Tehran, Iran*
**&**
**Kourosh Akef** [†]
*PhD student, Islamic Azad University, Science and Research Branch,*
*English Language Department, Islamic Azad university of Central Tehran, Iran*

**Abstract**
The purpose of the present study was to develop appropriate scoring scales for each of the defined stages of the writing process, and also to determine to what extent these scoring scales can reliably and validly assess the performances of EFL learners in an academic writing task. Two hundred and two students' writing samples were collected after a step-by-step process oriented essay writing instruction. Four stages of writing process – generating ideas (brainstorming), outlining (structuring), drafting, and editing – were operationally defined. Each collected writing sample included student writers' scripts produced in each stage of the writing process. Through a detailed analysis of the collected writing samples by three raters, the features which highlighted the strong or weak points in the student writers' samples were identified, and then the student writers' scripts were categorized into four levels of performance. Then, descriptive statements were made for each identified feature to represent the specified level of performance. These descriptive statements, or descriptors, formed rating scales for each stage of the writing process. Finally, four rating sub-scales, namely brainstorming, outlining, drafting, and editing were designed for the corresponding stages of the writing process. Subsequently, the designed rating scales were used by the three raters to rate the 202 collected writing samples.

[*] *E-mail address*: p_maftoon@yahoo.com
[†] *E-mail address*: kourosh.akef@gmail.com
*Address*: Islamic Azad University, Science and Research Branch, Pounak Square, Tehran, Iran.

[ Downloaded from ndea10.khu.ac.ir on 2026-11-06 ]

The scores thus obtained were put to statistical analyses. The high inter-rater reliability estimate (0.895) indicated that the rating scales could produce consistent results. The Analysis of Variance (ANOVA) indicated that there was no significant difference among the ratings created by the three raters. Factor analysis suggested that at least three constructs, –language knowledge, planning ability, and idea creation ability – could possibly underlie the variables measured by the rating scale.

**Keywords:** Writing assessment; Rating scales; Brainstorming; Outlining; Drafting; Editing

**Introduction**

Among the four major language skills, creating a coherent and extended piece of writing has always been considered the most difficult task to do in a language. Writing is a skill that even most native speakers of a language can hardly master. Foreign language learners, especially those who want to continue their education in academic environments, usually find writing a highly difficult and challenging task (Nunan, 1999; Richards & Renandya, 2002; Weigle, 2002; Hyland, 2003).

Over the years, different approaches have been adopted for teaching and assessing writing. According to Raimes (1991), traditionally, in audio-lingual approach, writing was viewed as transcribed speech. The primary focus of this approach was on formal accuracy (Silva & Matsuda, 2002). The learners' writing skill was assessed mainly through discrete-point tests of vocabulary, grammar, and sentence patterns, as well as through tests of controlled compositions. Therefore, the main focus of this approach was on the students' final written products.

Later, particularly after mid 1970s, understanding the need of the language learner for producing longer pieces of written language led scholars to realize that there was more to writing than constructing well-formed grammatical sentences. This realization led to the development of the paragraph-pattern approach (Raimes, 1991, 2002), which emphasized the importance of organization at extrasentential levels. The major concern of this approach was the logical construction and arrangement of discourse forms, especially to create different forms of essays. This was also a product-oriented approach in which learners were required to focus their attention on forms or final products (Silva & Matsuda, 2002). The assessment in

[ Downloaded from ndea10.khu.ac.ir on 2026-11-06 ]

this approach was based on how well learners would be able to create error-free final products.

However, these product-oriented approaches were not consistent with the new emerging ideas of discourse analysis after mid 1980s which emphasized the non-linear generation of thought and its expression in the process of communication (Nunan, 1999). According to process approach to writing, writing is a recursive, explanatory, and generative process. It focuses on the writer and the process or the strategies involved in writing. In the classroom, the objective of the process approach is to help the learner develop practical strategies for getting started, drafting, revising, and editing (Zamel, 1983; Raimes, 1985; Silva & Mutsuda 2002).

White and Arndt (1991) see a process-focused approach to writing as an enabling approach. They believe that the goal of this approach is "to nurture the skills with which writers work out their own solutions to the problems they set themselves, with which they shape their raw material into a coherent message" (p.5). They suggest that producing a text involves six recursive (nonlinear) procedures of generating ideas, focusing, structuring, drafting, reviewing, and evaluating. Seow (2002) also maintains that the writing process can be broadly seen as comprising four main stages: planning, drafting, revising, and editing.

Unfortunately, in writing classrooms, the process approach has not been applied effectively specially in academic environments, although many instructors believe that they have adopted some of its features in their teaching methodology. Specially, when it comes to assessment of students' writing ability, the criteria for success has mostly been either the instructors' professional but subjective intuitions, or at best the grammatical accuracy of the students' final products.

However, when assessing language skills are concerned, an important issue that should be taken into account is that writing assessment has always been considered a kind of performance assessment, and performance assessment focuses on the evaluation of learners in the *process* of performing the assigned tasks. The term performance assessment is used to refer to any assessment procedure that involves either the observation of behavior in the real word, or a replication of a real-life activity –that is, a performance ability being assessed, or the evaluation of that performance by some raters (Weigle, 2002). McNamara (1996) also maintains that

[ Downloaded from ndea10.khu.ac.ir on 2026-11-06 ]

"the *performance process* (the nature of the performance task) is considered to be the distinctive feature of performance assessment" (p.9). However, writing assessment procedures in academic contexts are a long way off from the pure form of performance assessment.

The main issue in the field of language testing is to embrace the notion of performance assessment as a means of achieving a close link between the test situation and authentic language use (Lynch & McNamara, 1998). Many educators have come to recognize that performance assessments are an important means of gaining a dynamic picture of learners' academic and linguistic development (Bachman, 1990, 1991; Gipps, 1994; Genesee & Upshur, 1996; Brown & Hudson, 1998; Chapell & Brindly, 2002).

Performance assessment is particularly useful with English Foreign Language (EFL) learners because it takes into account strategies that learners use in order to show what they can already do with the language they are learning. In foreign language environments, especially in writing classes, the students are usually penalized for their errors and for the qualities they have not yet achieved. In performance assessment, unlike traditional testing (in product oriented approaches), learners are evaluated on what they can put together and produce rather than on what they are able to recall and reproduce. In other words, in performance assessment, the actual performances of relevant tasks are required of the test takers, rather than the more abstract demonstration of knowledge achievement (McNamara, 1996). According to Bachman (2000), this type of assessment has been referred to by other scholars as alternative (Herman et al., 1992) or authentic (Newman et al., 1998; Terwilliger, 1997, 1998; Wiggins, 1989, 1993, cited in Bachman, 2000) assessment, whose goal is to "gather evidence about how learners are approaching, processing, and completing real-life tasks in a particular domain" (Huerta-Macias, 1995, p.9).

A true performance-based assessment is distinguished from the traditional measurements in terms of two factors: (1) a *performance process* of the examinees which is observed, and (2) *an agreed judging process* (e.g., a rating scale) by which the performance process is judged (McNamara, 1996). In other words, in the performance-based assessment the *candidate's performance* is rated or judged according to a *scale*.

[ Downloaded from ndea10.khu.ac.ir on 2026-11-06 ]

Thus, in writing assessment which is a type of performance assessment (Weigle, 2002), the rating scale is an important element. A score in a writing assessment is the outcome of an interaction that involves not merely the test taker and the test, but the test taker, the task, the written text, the rater(s) and the rating scale (McNamara, 1996). McNamara also notes that the scale that is used in assessing performance tasks, such as writing tasks, represents, implicitly or explicitly, the theoretical basis upon which the test is founded; that is, it embodies the test or the scale developers' notion of what skills or abilities are being measured by the test. Therefore, form the above mentioned discussions it can be concluded that for assessing the performance of students at every stage of the process, an appropriate rating scale would be needed to measure the performance of the students at that stage.

Weigle (2002) mentions three main types of rating scales: primary trait scales, holistic scales, and analytic scales. In primary trait scoring, the rating scale is defined with respect to the specific writing assignment, and the students' scripts are judged according to the degree of success with which the student writers have carried out the assignment. However, in a typical holistic scoring, each script is read and judged against a rating scale, or scoring descriptor, that outlines the scoring criteria. Yet, in analytic scoring, scripts are rated concerning several aspects of the written task rather than assigning a single score to the scripts. Depending on the purpose of the assessment, scripts might be rated on such features as content, organization, cohesion, register, vocabulary, grammar, or mechanics. Analytic scoring schemes thus provide more detailed information about a student writer's performance in different aspects of writing. It is for this reason that many scholars prefer analytic scoring over holistic scorings (Bachman & Palmer, 1996; North & Schneider, 1998; Weigle, 2002).

Bachman and Palmer (1996) also maintain that in situations where the use of language is tested in tasks that involve extended production responses, the quality of the response is judged through rating scales in terms of levels of ability required for completing those test tasks. They argue that developing rating scales should be based on two principles. First, the operational definitions in the scales should be based on theoretical definitions of the construct. Second, the scale levels should tap specified levels in different areas of language ability, in which the lowest level in the rating scale would be defined as no evidence of the ability and the highest level as evidence of mastery of the ability. Bachman and Palmer (1996) further mention

[ Downloaded from ndea10.khu.ac.ir on 2026-11-06 ]

two practical advantages of using analytic scales: First, these scales provide a profile of the areas of language ability that are rated. Second, analytic scales tend to reflect what raters actually do when rating samples of language.

Regarding the above-mentioned issues, this study aims at designing an appropriate model for the assessment of EFL learners' writing performances at the tertiary level. The purpose of the present study is, in fact, to develop rating scale descriptors for assessing writing performance of EFL learners at the operationally defined stages of the writing process, and also to determine whether the suggested rating scale descriptors could reliably and validly assess the performance of student writers at each stage of the process.

**This study**

The present study was carried out in two distinct phases: a qualitative phase and a quantitative phase. The qualitative phase was needed to identify a number of distinctive features in the student writers' scripts created in each stage of the writing process which were determinant in classifying those scripts into the appropriate performance levels. These features form the variables of this study which served as the input for the statistical analysis in the second phase, or the quantitative phase of the study.

The aim of the qualitative phase of this study was to analyze students' scripts at each stage to see if there were any features in each stage that characterize the students' performances in that stage and could be used as a basis for the design of a scoring scale for that very stage. In other words, the purpose is to see if it is possible to classify the students' scripts in each stage into different levels of performance such as VERY GOOD, GOOD, FAIR, and POOR based on the characteristics, or features (e.g., weaknesses and strengths in performing the given writing tasks) observed in each script. The classification of the scripts into different levels of performance was done through a close and thorough analysis of students' scripts, as well as through consultation with experienced writing instructors.

In the second phase, or the quantitative phase of the study, statistical procedures including inter-rater and intra-rater reliability estimations, the Analysis of Variance (ANOVA) were consulted in order to check the reliability of assessments resulted from the application of the rating scales.

[ Downloaded from ndea10.khu.ac.ir on 2026-11-06 ]

In order to check the validity and to identify how many constructs underlie the variables identified in the qualitative phase of the study for the stages of the writing process, explanatory factor analysis (EFA) was conducted. The aim here was to represent the set of variables observed and identified in the qualitative phase of this study in terms of a smaller number of hypothetical variables or constructs. In other words, the purpose was to see to what domains of language or cognitive abilities the identified variables in this study belonged.

**Research Questions**

The research questions for the study were as follows:

1. Is there any distinguishing feature in the students' scripts at each stage of the writing process?
2. Is there any statistically significant difference among the ratings made by the three raters for the evaluation of the student writers' scripts?
3. What underlying constructs are measured by the variables assessed through the application of the rating scale?

**Method**

**Participants**

The participants in this study consisted of university students, and three raters.

University students (student writers)

The subjects participating in this study were university students who were studying English translation at the College of Foreign Languages, Islamic Azad University, Central Tehran Branch and Karaj Branch. The samples were collected during three subsequent semesters from September 2004 to June 2006. The subjects were taking the Essay Writing course which is usually presented in the fourth or fifth semester of their education. Totally 450 samples were collected from 210 students.

Since sex of the subjects was not a relevant variable, there was no control for sex variable applied. Because of the unbalanced percentages of male and female students studying the English language at Islamic Azad Universities, in the group of subjects participating in this study, females outnumbered males.

Raters

In order to see how the designed rating scales could function in evaluating the student's written performances, the students' scripts were scored by three raters,

[ Downloaded from ndea10.khu.ac.ir on 2026-11-06 ]

including the researcher, based on the designed rating scales. Two raters who shared almost similar backgrounds in terms of qualifications and teaching experience as the researcher were selected. The raters had about three years experience teaching advanced writing and essay writing courses in universities. They were also the Ph.D. holders and they were all members of academy of Islamic Azad University, Karaj and South-Tehran Branch. Three forty-five-minute training sessions were held. In these sessions, the raters were briefed about the purpose of the study and the designed rating scales.

**Instruments**

The materials for the Phase One of this study consisted of a number of essay writing prompts, as well as a set of instructions which guided the students how to perform in assigned writing tasks. After collecting appropriate data, the goal was to design an instrument for rating the students' sample scripts in each stage of the writing process. Here, a detailed explanation of the materials used in this study is presented.

Essay writing prompt

In order to elicit the required writing samples of the subjects in different stages of the writing process, a number of writing tasks were designed (Appendix B). Each writing task consisted of a single prompt plus a set of instructions. Each task required the subjects to write a five-paragraph essay. The participants were instructed to produce separate scripts for each stage of the writing process –i.e., a script for brainstorming (generating-ideas) stage, a script for structuring stage, at least two scripts for the drafting stage, and finally one script for the editing stage. For classroom tasks, 90 minutes were allocated for each complete task.

Suggested rating scales

After the data were collected, the main task was to design a set of rating scales which could assess the performance of the students in each stage of the writing process. First, a general holistic scale was designed to help the researcher categorize the students' scripts into four levels of performance – namely VERY GOOD, GOOD, FAIR, and POOR. The detailed explanations of the way these categorizations were defined and the resulting holistic scale will be presented later in "Procedures" section under the "designing rating scales" heading. The students' scripts were categorized according to the consideration that how effectively their

[ Downloaded from ndea10.khu.ac.ir on 2026-11-06 ]

performances could address the requirements of a given task at every stage of the writing process.

Second, based on the features identified in student's scripts in each level of performance, separate rating scales were designed for each stage of the writing process. Therefore, four rating scales were designed for the stages brainstorming, outlining, drafting, and editing. Each rating scale included five variables which represented the identified features in that very stage. These variables will be explained in "designing rating scale" section with greater details.

**Procedures**

A careful systematic procedure was adopted in this study. First, in order to collect suitable data for the purpose of this study, a careful systematic step-by-step teaching procedure was required to enable the subjects participating in this study to produce appropriate output in each stage of the writing process. Then, after the suitable data were collected from the trained student writers, other step-by step systematic procedures were taken for designing scales, rater training, and scoring the subjects' scripts. Here, a detailed description of these procedures is presented.

**Teaching procedures**

To collect appropriate sample scripts at each stage of the writing process, a specific process-product approach was adopted. The subjects were taking the essay-writing course in the fourth or fifth semester of their academic studies. The assigned textbook for this course is normally "*The Practical Writer with Readings*" by Bailey and Powell (1989). This is mostly a product-oriented textbook which mainly deals with product-related issues, such as the format of a five-paragraph essay, topic sentences and supporters, coherence and unity, reminders and transitions. For the purpose of this study, in addition to this textbook, another textbook entitled "*Process Writing*" by White and Arndt (1991) was also chosen which clearly and comprehensively present a step-by-step procedure for teaching writing skills through a process-oriented approach.

The strategy adopted for teaching in this study was to devote one session to process writing and one session to product writing alternatively. For example, in the process session, different techniques of generating ideas and finding a topic were presented, and the student writers practiced this process in groups, individually, as a class activity on board, and as homework assignments. Then, the

[ Downloaded from ndea10.khu.ac.ir on 2026-11-06 ]

student writers practiced and learned how to find and form their main idea out of a random list of phrases and sentences they had created. In the following product session, the student writers were taught about topic sentences, sentences that can form more appropriate topic sentences, and the main features of a good topic sentence. Necessary feedbacks were presented in each session during class activities, and during correcting the students' homework assignments. These feedbacks were given both on form and on the contents of student's scripts (Williams, 2003).

The students were required to submit their homework scripts each session. These scripts were corrected and then returned to the students again so that they could observe their problems and errors, and they could incorporate the information they got from these feedback comments on their future drafts. At certain points, when it seemed that the students had learned how to adequately perform in each stage, their scripts were collected for future use as part of the data of the present study.

After the teaching syllabus was completely covered, the students were asked to write a complete five paragraph essay on a selected prompt as a homework assignment. They were asked to write down their writing performances in each stage of writing on a separate piece of paper. As a result, each student handed in at least four separate paper sheets, each of which represented his or her performance in the stages of writing process, namely generating ideas (brainstorming), structuring (outlining), drafting, and editing.

In three final sessions, the students were required to write five-paragraph essays on assigned prompts in the class and submit their scripts at the end of each session. Again like the homework assignment, the students were asked to hand in their writing scripts in each stage of the writing process on separate paper sheets. These scripts were also collected to serve as data in this study.

**Data collection**

The writing samples were collected during three subsequent semesters from September 2004 to June 2006. The students were instructed to provide writing samples in each stage of the writing process.

[ Downloaded from ndea10.khu.ac.ir on 2026-11-06 ]

Not all of the samples were found appropriate for the purpose of this study. Some samples did not have scripts for all of the stages of the writing process. From among the pool of samples, the researcher selected only those which form four-stage script sets. Therefore, the data collected from each subject consist of samples of his or her performance on the defined stages of generating ideas, organizing, drafting, and editing. Totally, 202 four-stage script sets could be selected from among the pool of the collected sample scripts.

The data for this study included students' single-stage scripts representing their performance in every single stage of the writing process, as well as their four-stage script sets which they created while writing a five-paragraph essay on a single prompt. The single-stage scripts were produced by the students as class activities or homework assignments for every stage after they had received instructions about how to perform successfully in that stage. These samples were collected throughout the semesters.

The four-stage script sets were made up of collections of scripts which students created in different stages of writing (i.e. generating ideas, structuring, drafting, and editing) while responding to a single prompt for writing a five-paragraph essay. These samples were collected in the last three sessions of each semester both as homework assignments and as live class performance.

**Designing rating scales**

After the required data were collected, rating scales for each stage of the writing process were designed following a thorough observation and analysis of the collected scripts. The procedure for designing the rating scales consisted of operationally defining the stages of the writing process, categorizing students' scripts according to their level of performance, describing the features observed in the scripts at each level of performance, assigning cut-off scores to each level of performance, and, finally, revising the statements describing the features of performance observed in each level.

At first, the collected scripts were carefully and closely observed to locate features which could highlight strong or weak points in student's scripts, and which could form the bases for categorizing the samples into different levels of performance. Two essay-writing instructors assisted the researcher in identifying these features.

[ Downloaded from ndea10.khu.ac.ir on 2026-11-06 ]

Secondly, based on the identified features, the scripts produced by the students in each stage were categories into four levels of performance rated holistically as: VERY GOOD, GOOD, FAIR, and POOR. VERY GOOD is defined as the level of performance in which the students' performance effectively addressed the requirement of a given stage of writing and little or no weakness points were observed. GOOD refers to a level of performance in which the students were successful in accomplishing the task but at same time minor, negligible weaknesses could be observed. FAIR refers to the level in which students' performances address the task of that stage but contain noticeable, important weaknesses. The performance in this level is not adequate enough to help the students perform successfully in later stages (usually revision is needed). POOR refers to the level of performance in which serious weaknesses are observed and the students' performance cannot address the task in a given stage. Generally, the performance at this level is not acceptable at all. Table 1 summarizes this information.

**Table 1**
A holistic scale of students' levels of performance on stages of writing process

| Levels of Performance | Description |
|---|---|
| **VERY GOOD** | Students have effectively accomplished the requirement of a given stage of the writing process. Few or no weak points are observed. |
| **GOOD** | Students have successfully accomplished the task of a given stage of the writing process, but there are minor, unimportant weak points. |
| **FAIR** | Students have, to some extent, addressed the task of a given stage of the writing process, but noticeable, important weaknesses can be observed in their performance. The quality of the performance is not adequate enough to be used in later stages of the writing process. Revision is needed. |
| **POOR** | Students have not been able to accomplish the task. There are serious weak points in the performance. Generally, the performance is not acceptable. |

Based on the features of strengths and weaknesses identified in the student's scripts in each stage of the writing process, the students' scripts in each stage were categorized into the above-mentioned levels. Therefore, for each stage, four

[ Downloaded from ndea10.khu.ac.ir on 2026-11-06 ]

performance levels (categories) were identified. The scripts placed in one level shared similar performance features holistically described by the level descriptor.

The next step was to operationally define each stage of the writing process. These stages were defined theoretically according to the available literature, as well as empirically, through careful analysis of the data. Here a brief explanation of these procedures is presented.

Based on the White and Arndt's (1991) model, for practical writing stages were defined for each of which the students could have output scripts:

**Generating ideas (*brainstorming*):** It can be operationally defined as one of the (prewriting) stages of the writing process in which student writers draw upon their long term memory in order to discover a topic, identify the purpose, and generate ideas for a piece of writing (White & Arndt, 1991). In this stage, students are required to focus their attention to the topic and jot down every idea that comes to their mind about the topic.

**Structuring (*outlining*):** It can be operationally defined as a stage of the writing process in which student writers involve in grouping the generated ideas together and deciding upon how to sequence these ideas (White & Arndt, 1991). In this stage student writers were required to arrange their random ideas first in a form of a spidergram, and then sequence these grouping of ideas in form of an appropriate outline.

**Drafting (*writing/revising*):** It can be operationally defined as the stage of the writing process in which student writers start producing stretches of written texts (Seow, 2002). In this stage students were required to start writing in order to create several drafts of their essay. In this study, the students were asked to go through the write-revise-rewrite cycle at least once as suggested by White and Arndt (1991).

**Editing (*reviewing/evaluating/proof reading*):** It can be operationally defined as one of the stages of the writing process in which students are engaged in tidying up their text as they prepare the final draft to be evaluated by the teacher. In this stage the students edited their own or their peers' works for grammatical well-

[ Downloaded from ndea10.khu.ac.ir on 2026-11-06 ]

formedness, appropriate use of vocabulary, correct spelling and punctuation, and appropriate supportive materials, such as examples, quotations, and the like.

After that, the researcher, assisted by two essay-writing instructors, identified and described the salient features of the scripts placed at each level (category) of performance as clearly and comprehensively as possible. These descriptions were written as statements highlighting and summarizing the most outstanding features observed at each level of performance in which the student writers showed distinctive strengths or weaknesses. These statements formed the scale descriptors for each level. The set of descriptors describing the levels of performance for each stage comprised the analytic rating scales for that given stage. Therefore, based on the original holistic scale, analytic rating scales were designed for every defined stage of the writing process.

Then, cut-off scores were assigned to each level descriptor. Numbers from 4 to 1 in descending order were assigned to levels VERY GOOD to POOR, respectively. Level 0 was also assigned to cases where there was no observable performance in a given stage of the writing process. The final form of these rating scales for each stage of writing process is presented Table 2 below.

**Rater training and scoring procedure**

Because of the unavoidable variability that exists among different raters, attempts were made to reduce the variability of raters' judgment and also to increase the raters' levels of agreement with each other. As it was mentioned earlier, three forty-five minute rater training sessions were held in order to brief the raters about the purpose of the study and create a consensus among the raters.

A series of carefully selected scripts illustrating salient features of students' different levels of performance were rated by the researcher using think-aloud ratings. Then the raters were asked to rate another series of selected scripts independently in the training session and then to discuss the results. In order to ascertain that an acceptable level of agreement existed among the raters, fifty writing samples were rated by the three raters using the TOEFL Writing Scoring Guide (2007). This was done to see how much agreement can be achieved at the onset between raters using a standard writing scoring scale. Based on these ratings, the inter-rater reliability was estimated. When the inter-reliability estimate reached

[ Downloaded from ndea10.khu.ac.ir on 2026-11-06 ]

a satisfying level ($r_{tt}$=0.92), the raters started rating the students' scripts based on the designed rating scale.

**Data Analysis**

The aim of the second phase of the study was to find statistical support for the findings of the qualitative phase. The crux of the matter was to determine to what extent the rating scale can function appropriately for the assessment of the writing performances of Iranian EFL students in every stage of the writing process, and how far the rating scale can provide a diagnostic tool for essay writing instructors to uncover the students' weaknesses and strengths in each stage of the writing process.

After the ratings were done, statistical techniques were used to provide supports for the validity and the reliability of the assessment made by the suggested rating scale. These statistical techniques included the inter-reliability and intra-reliability estimates, as well as factor analysis.

**Results**

As it was mentioned before, the data in this study consisted of 202 students' sets of writing samples. Each set included separate scripts representative of the students' performances in each operationally defined stage of writing process, namely, generating ideas (brainstorming), outlining, drafting, and editing. The answer to the first question of this study included extracting the features from the students' writing samples and designing rating scales based on these features. The rating scales designed in this study include four sub-scales for each of the above-mentioned stages of the writing process. Each sub-scale includes five components with scale descriptors which describe the students' quality of performance on five-operationally defined components according to four levels of performance namely, VERY GOOD, GOOD, FAIR, and POOR. Table 2 briefly summarizes the information about the designed rating scale and its sub-scales as well as the levels of performance.

[ Downloaded from ndea10.khu.ac.ir on 2026-11-06 ]

**Table 2**
The designed rating scale and its components

| **Stages of writing process** | **Components** | **Levels of performance** | | | | |
|---|---|---|---|---|---|---|
| Generating Ideas (Brainstorming) | Number of ideas: the number of guesses or new ideas that the student could create for the specific prompt | 0 | 1 | 2 | 3 | 4 |
| | Development of ideas: the extent to which the students were able to express their ideas in forms well developed concepts | 0 | 1 | 2 | 3 | 4 |
| | Aspects: the extent to which students can address the positive or negative aspects of the ideas given in the writing prompt | 0 | 1 | 2 | 3 | 4 |
| | Diversity of ideas: the extent to which the student can create ideas which are facts, opinions, examples, etc | 0 | 1 | 2 | 3 | 4 |
| | Usefulness: the extent to which the ideas were useful in creating an outline | 0 | 1 | 2 | 3 | 4 |
| Outlining | Content of outline: the extent to which the outline is detailed | 0 | 1 | 2 | 3 | 4 |
| | Relevance: the extent to which the outline present relevant and logical grouping of the ideas. | 0 | 1 | 2 | 3 | 4 |
| | Use of subordinate ideas: The extent to which the outline has subordinate ideas to support coordinate ideas | 0 | 1 | 2 | 3 | 4 |
| | Application of the ideas created in brainstorming: the extent to which the outline reflected the application of ideas in the brainstorming stage | 0 | 1 | 2 | 3 | 4 |
| | Effective use of outline in drafting stage: the extent to which student writers had used the outline in creating their draft. | 0 | 1 | 2 | 3 | 4 |
| Drafting | Writing fluency: the ability of students to produce a continuous piece of writing without causing difficulty or breakdown of communication | 0 | 1 | 2 | 3 | 4 |
| | Having a clear central idea; the extent to which the draft presents a central idea to be communicated to the audience | 0 | 1 | 2 | 3 | 4 |
| | Relevance (unity): the extent to which the draft provides relevant information about the central idea of the prompt through supporting sentences | 0 | 1 | 2 | 3 | 4 |
| | Coherence: the extent to which the meaning of sentences are linked logically by using the cohesive devices throughout the text | 0 | 1 | 2 | 3 | 4 |
| | Organization: the extent to which the draft can attract readers' attention, presents blueprints, has body paragraphs with topic sentences and supporters, and has a conclusion which is linked to the introductory paragraph and the thesis statement | 0 | 1 | 2 | 3 | 4 |
| Editing | Grammatical accuracy: The extent to which the final draft contains structures with few noticeable grammatical errors or word order problems | 0 | 1 | 2 | 3 | 4 |
| | Appropriate use of vocabulary: The extent to which the final draft features the precise and effective use of vocabulary | 0 | 1 | 2 | 3 | 4 |
| | Organization/ coherence revision: the extent to which the student has been able to revise the organization and coherence problems in his/her draft | 0 | 1 | 2 | 3 | 4 |
| | Relevance/ adequacy of information: the extent to which the student has been able to improve the relevance of information in his/her final draft | 0 | 1 | 2 | 3 | 4 |
| | Mechanics of writing (spelling and punctuation): The extent to which the draft appears accurate in spelling and punctuation | 0 | 1 | 2 | 3 | 4 |

[ Downloaded from ndea10.khu.ac.ir on 2026-11-06 ]

**Inter-rater and intra-rater reliability estimations**

The second question of this study dealt with the extent to which the suggested rating scale can produce consistent results. Using the designed rating scale, the students' scripts were rated by three raters who were Ph.D. holders with at least three years of experience in teaching essay-writing courses at the tertiary level. First, an attempt was made to establish an acceptable level of inter-rater reliability between the raters using the TOEFL Writing Scoring Guide (2007) before they actually started rating the scripts using the suggested rating scales ($r_{2/3}$ = 0.82, $r_{1/2}$ = 0.79, $r_{1/3}$ = 080). After the ratings of fifty scripts, the initial inter-rater reliability ($r_{tt}$) was estimated for the 3 raters ($r_{tt}$=0.92). According to Henning (1987), when more than two raters are involved, the average of all correlation coefficients should be calculated, and then this average should be adjusted by means of the Spearman-Brown Prophecy Formula to make the final reliability estimate reflect the number of participating raters.

The inter-rater reliability was estimated again (0.895) after the raters rated the students' writing samples using the suggested rating scales. Table 3 summarizes the results.

**Table 3**

Correlation coefficient between the raters and the inter-rater reliability estimates

| **Raters** | **1-2** | **1-3** | **2-3** | **Inter-rater reliability** |
|---|---|---|---|---|
| r (TOEFL) | 0.820 | 0.792 | 0.810 | 0.920 |
| r (rating scales) | 0.721 | 0.745 | 0.756 | 0.895 |

Analysis of Variance (ANOVA) was also used to check if there is any significant difference among the ratings produced by the three raters. Table 4 shows the results of this analysis.

**Table 4**

The Analysis of Variance for the rating produced by the raters

| | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| Between Groups | 48.803 | 2 | 24.402 | 2.649 | .074 |
| Within Groups | 354.067 | 147 | 9.211 | | |
| Total | 402. 870 | 149 | | | |

[ Downloaded from ndea10.khu.ac.ir on 2026-11-06 ]

As Table 4 indicates the observed F value (2.649) is below the critical F value (3.07) at 0.05 level of significance. Therefore, it can be concluded that there was no significant difference among the ratings produced by the raters.

In order to estimate the intra-rater reliability, 30 samples were randomly selected from the pool of 202 samples. Using the designed rating scales, these 30 samples were rated again by Rater 1 (the first researcher) without any reference to the previous ratings. Calculating the correlation coefficients between the previous and the second ratings of these 30 samples indicated intra-rater reliability estimations for Rater 1. Table 5 summarizes the results.

**Table 5**
Intra-rater reliability estimation for Rater 1

| Rater 1 | Brainstorming | Outlining | Drafting | Editing | Total score |
|---|---|---|---|---|---|
| R (intra-rater) | 0.931 | 0.958 | 0.981 | 0.974 | 0.987 |

After the students' scripts were rated using the rating scales suggested in this study, results were entered into the SPSS program data matrix in order to perform the statistical analyses. Then, the data for these variables were used to perform factor analysis in order to find statistical supports for the validity of the results produced by these rating scales.

**Factor analysis**

One of the purposes of this study was to see how many constructs underlie the variables identified for the stages of the writing process. In other words, the purpose was to see to what domains of language or cognitive abilities the identified variables in this study belonged. The 20 variables on which the students' writing performances were rated were entered into the Exploratory Factor Analysis (EFA) using SPSS software. Table 6 displays the variables entered into the factor analysis.

[ Downloaded from ndea10.khu.ac.ir on 2026-11-06 ]

**Table 6**
Variables entered into factor analysis study

| | Mean | SD | Analysis N |
|---|---|---|---|
| B- Number | 2. 85 | .932 | 202 |
| B- Development | 2.55 | 1.087 | 202 |
| B- Aspect | 2.45 | .941 | 202 |
| B- Diversity | 1.79 | .939 | 202 |
| B- Usefulness | 2.30 | 1.047 | 202 |
| O- Detail | 2.27 | .972 | 202 |
| O- Relevance | 2.42 | .944 | 202 |
| O- Subordination | 1.70 | .977 | 202 |
| O- Application | 2.23 | 1.047 | 202 |
| O- Effectiveness | 2.61 | 1.069 | 202 |
| D- Fluency | 2.90 | .852 | 202 |
| D- Central idea | 3.09 | .884 | 202 |
| D- Relevance | 2.65 | .839 | 202 |
| D- Coherence | 2.74 | .761 | 202 |
| D- Organization | 2.83 | .898 | 202 |
| E- Grammar | 2.71 | .703 | 202 |
| E- Vocabulary | 2.58 | .861 | 202 |
| E- Organization | 3.03 | .896 | 202 |
| E- Relevance | 2.77 | .856 | 202 |
| E- Mechanics | 2.29 | .962 | 202 |

After several trial runs of the SPSS program, Principal Axis Factoring method was used for factor extraction, and Oblimin method with Kaiser Normalization was chosen for rotating factor loadings. Cudeck (2000) suggests that the Direct Oblimin with parameter zero is the best method in a variety of circumstances on both algebraic, as well as practical grounds.

Factor analysis showed that three factors can be extracted from the variables entered into the analysis. Table 7 shows the factor loadings for each variable on the factors after oblique rotation. Each value represents the partial (direct) correlation between the item and the rotated factor. The values where sorted by size for the ease of interpretation.

[ Downloaded from ndea10.khu.ac.ir on 2026-11-06 ]

**Table 7**
Pattern matrix for oblique rotation [a]

| | Factor | | |
|---|---|---|---|
| | 1 | 2 | 3 |
| E- Vocabulary | .877 | | |
| E- Grammar | .862 | | |
| D- Coherence | .849 | | |
| E- Relevance | .816 | | |
| E- Mechanics | .816 | | |
| E- Organization | .783 | | |
| D- Fluency | .765 | | |
| D- Relevance | .743 | | |
| D- Organization | .665 | | |
| D- Central idea | .478 | | |
| O- Application | | .911 | |
| O- Effectiveness | | .880 | |
| O- Detail | | .877 | |
| B- Usefulness | | .801 | |
| O- Relevance | | .672 | |
| O- Subordination | | .613 | |
| B- Development | | | .785 |
| B- Diversity | | | .675 |
| B- Aspect | | | .673 |
| B- Number | | | .580 |

Extraction Method: Principal Axis Factoring
Rotation Method: Oblimin with Kaiser Nominalization
a. Rotation converged in 5 iterations

All editing and drafting variables have the highest loadings on Factor 1. All of the outlining variables plus one brainstorming variable, namely Effect, were highly loaded on Factor 2, and the rest of brainstorming variables were highly loaded on Factor 3. The interesting point to note is the order the variables were arranged in Factor 1. Almost all editing variables were listed one after another with only Coherence which was a drafting variable came after grammar. The rest of the drafting variables followed one another successively after the editing variables.

These factors can represent the constructs underlying the 20 variables which were measured using the rating scales suggested in this study. Now the issue at

[ Downloaded from ndea10.khu.ac.ir on 2026-11-06 ]

stake is to determine the nature and role of these factors in the overall assessment of the students' writing ability. Table 8 summarizes the information presented in Table 7, and lists the 20 variables under the relevant extracted factors.

**Table 8**
The arrangement of variables based on factor loadings

| **Factors** | | |
|---|---|---|
| **1**<br>**Language ability** | **2**<br>**Planning Ability** | **3**<br>**Idea Creation Ability** |
| E-vocabulary | O-Application | B-Development |
| E-Grammar | O-Effectiveness | B-Diversity |
| D-Coherence | O-Detail | B-Aspect |
| E-Relevance | B-Usefulness | B-Number |
| E-Mechanics | O-Relevance | |
| E-Organization | O-Subordination | |
| D-Fluency | | |
| D-Relevance | | |
| D-Organization | | |
| D-Central idea | | |
| | | |

B = Brainstorming; O = Outlining; D = Drafting; E = Editing

The highest loading variables on each factor, to some extent, can reveal the nature of that factor. Since Vocabulary and Grammar were highly loaded on the first factor, this factor (construct) can be labeled Language Knowledge.

The highest loading variables on Factor 2 were Application and Effectiveness. The Application variable in outlining tried to measure the extent to which the student could apply the ideas they had generated in the brainstorming stage in creating their outlines. Similarly, the Effectiveness variable tried to measure the extent to which the students could use their outlines effectively in generating their first drafts. These two variables reflect the strategies the students use in order to organize their outlines and their drafts. Thus, the second extracted factor is labeled Planning Ability.

As it can be seen in Table 8, the variable Usefulness of brainstorming was also loaded on the second factor. This is because the Usefulness variable tried to

[ Downloaded from ndea10.khu.ac.ir on 2026-11-06 ]

measure how useful the generated ideas were in the brainstorming stage, and how well these ideas could be used in creating outlines. Again, this variable addressed planning strategies. Hence, there was no wonder why it was loaded among the other variables on the second factor.

Development was the highest loading variable on Factor 3. This variable tried to measure how well the students could develop the ideas they had generated. This indicates that Factor 3 possibly dealt with the power of thinking and generating ideas. As a result, the third factor was labeled Idea Creation Ability. Interestingly, all other brainstorming variables were loaded on Factor 3.

Therefore, the results of factor analysis conducted in this study showed that the rating scale suggested in this study could be considered a valid tool for assessing the students writing ability. After applying necessary rearrangement, the suggested rating scale can be used to effectively measure the underlying constructs of writing ability. The final modified form of the rating scales is presented in Appendix A.

**Discussion**

The statistical analyses indicated that using the suggested rating scale could guarantee the reliability and the validity of assessing writing performances of student writers. The inter-rater and intra-rater reliability estimates revealed that the rating scale could help raters to make more consistent ratings.

The results also indicated that the variables defined in this study could measure at least three underlying constructs of Language Knowledge, Planning Ability, and Idea Creation Ability. Drafting and editing sub-scales measured the student writers' knowledge of language. Outlining sub-scale assessed the student writers' planning abilities, and brainstorming sub-scale rated the students based on their idea creation abilities. Consequently, the statistical analyses provided evidence in favor of the construct validity of the assessments made by the suggested rating scale.

The suggested rating scale developed as such may face some criticisms. Scoring students' writing scripts has always been a very time-consuming task in itself. Using analytic scales for rating scripts makes this problem even worse. It requires raters to make more than one decision for every script. In the case of this study, the samples consisted of at least four scripts produced by each student for each stage of

[ Downloaded from ndea10.khu.ac.ir on 2026-11-06 ]

the writing process. This would remarkably increase the amount of time needed to score students' scripts. Therefore, especially when the number of students is rather high, using these rating sub-scales may not seem much practical. Because of this very problem, such detailed rating sub-scales cannot be used for large-scale writing assessments.

However, putting the large-scale assessments aside, the use of such detailed rating sub-scales for the evaluation of the students' performances in different stages of the writing process can be well justified in language classrooms and academic environments. First, it should be taken into consideration that the present rating scale with its sub-scales is by no means intended for the assessment of students' writing proficiency. Rather, it is merely geared to the evaluation of students' achievements in essay writing classes. In addition, these rating sub-scales are mainly designed for formative assessments to evaluate and monitor students' progress during a course of study. These sub-scales can provide a performance profile for every student, showing his or her weaknesses or strengths at each stage of the writing process. Based on these profiles, teachers can modify their teaching methods and materials so as to make them more effective and appropriate for the students' needs, and capabilities. In addition to teaching and testing, the suggested rating scale can provide helpful insights for a researcher who likes to delve into the nature of the writing process, the strategies learners use in composing a piece of writing, and the issues of rater training.

**Conclusion**

Since the results of language tests should most often be reported in a form of scores, rating scales are inevitable parts of any assessment procedure which deals with the evaluation of students' performance skills in prompt-type tasks such as speaking and writing in which students are required to give extended responses. In these prompt-type tasks, the quality of response is usually judged in terms of levels of ability demonstrated by the students in completing the assigned task, via the use of rating scales defined and developed for the evaluation purpose.

In this regard, Bachman and Palmer (1996) maintain that in developing analytic rating scales there should be the same number of separate rating scales as there are distinct component in the construct definition. McNamara (1996) also notes that the scales which are used in assessing performance tasks, such as writing, represent the theoretical basis upon which the test is founded. Hence, the rating scales

[ Downloaded from ndea10.khu.ac.ir on 2026-11-06 ]

represent test makers' or scale developers' attitudes regarding what skills or abilities are being measured by the test. For this reason, according to McNamara (1996), the development of rating scales and the descriptors for such scales are of critical importance for the validity of the assessment.

Furthermore, Weigle (2002) also states that analytic rating scales can be more reliable than holistic or primary trait scales because more components are included for scoring students scripts. Therefore, the above-mentioned issues suggest that the rating scale designed in this study can provide a more valid and reliable instrument for essay writing instructors.

In addition, the designed rating scale in this study offers some other advantages for assessing students' writing ability. As it was mentioned earlier, the designed rating scale can serve as a practical and functional tool for formative evaluation of students' performances during a writing course. Most writing instructors like to assess the effectiveness of their instruction formally or informally on a continuous basis. In a process-oriented writing classroom, after the teacher has completed the necessary instructions on a given stage of the writing process for example, generating ideas, outlining, drafting or editing, she may like to know how effectively their students have acquired the necessary skills, and whether it is an appropriate time to move on to the next stage of the writing process. In this regard, the present rating scale can make a detail assessment of students' performance in every stage of the writing process.

In performance assessment, the assumption is that more reliability and validity would be obtained if a rating scale is developed that describes the features of the writing text in a valid way and if raters are adequately trained to understand the content of the rating scales (Lumley, 2002). Recent studies (Weigle, 1998; Lumley, 2002) have demonstrated that rater training is successful in making raters more self-consistent. The main effect of rater training is to reduce random errors in raters' judgments. The suggested rating scale can offer a handy instrument for rater training in order to increase self-consistency of the raters, as well as the agreement between them. The scale descriptors which describe the writing components involved in every stage of the writing process can help the raters to develop a common understanding about these elements and, thus, form a more consistent base for their judgments. The suggested rating scale can also provide rich materials for rater trainings. The four rating sub-scales with their descriptors can show to

[ Downloaded from ndea10.khu.ac.ir on 2026-11-06 ]

what extent different raters can place different emphases on different components of the writing ability, and how differently they may interpret these components. The suggested rating scale may help researchers and rater trainers to investigate the rater-item interaction.

The modern way of life and the growth of technology have increased the importance of writing in second and foreign language teaching context. As process oriented approaches and task based syllabi has gained popularity over the language curricula, the need for more sophisticated methods of writing assessment, as well as scoring procedures is felt more than ever. Designing of the present rating scale with its sub-scales has been an attempt for the fulfillment of this very need. Yet, it may not be without shortcomings. It is hoped that its application and further lines research reveal its usefulness for the assessment of students' writing performance.

**Acknowledgements**
The present article is based on a PhD dissertation presented in partial fulfillment of the requirements for a doctoral degree in TESL at Islamic Azad University, Science and Research Campus. This study was made possible by the invaluable instructions and guidelines of Dr. P. Birjadni, and Dr. A. Mirhassani, the readers of the PhD dissertation; as well as Dr. M. Alavi,. The researchers would also like to express their words of appreciation to their colleagues and friends Dr. E. Baggeridoost, and Dr. N. Ghal'eh, the raters, and Mr. Esmkhani, the SPSS consultant.

*Received 5 May, 2009*
*Accepted 23 August, 2009*

**References**

Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.

Bachman, L. F. (1991). What does language testing have to offer? *TESOL Quarterly,* 25 (4), 671-704.

Bachman, L. F. (2000). Modem language testing at the turn of the century: Assuring that what we count counts, *Language Testing,* 17(1), 1-42.

[ Downloaded from ndea10.khu.ac.ir on 2026-11-06 ]

Bachman, L. F. & Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.

Bailey, E. P. & Powell, P. A. (1989). *The Practical writer with readings*. New York: Holt, Rinehart and Winston, Inc.

Brown, J. D. & Hudson, T. (1998). The alternatives in language assessment, *TESOL Quarterly,* 32(4), 653-675.

Chapell, C. A. & Brindly, G. (2002). Assessment. In N. Schmitt (Eds.), *An introduction to applied linguistics* (pp. 267-288). London: Arnold.

Cudeck, R. (2000). Exploratory factor analysis. In H. A. Tinsley & S. D. Brown (Eds.), *Handbook of applied multivariate statistics and modeling* (pp. 265-296). New York: Academy Press.

Genesee, F. & Upshur, J. A. (1996). *Classroom-evaluation in second language education*. Cambridge: Cambridge University Press.

Gipps, C. V. (1994). *Beyond testing: Towards a theory of educational assessment*. London: The Flamer Press.

Huerta-Macias, A. (1995). Alternative assessment: Responses to commonly asked questions, *TESOL Quarterly,* 5, 8-10.

Huerta-Macias, A. (2002). Alternative assessment. In J. C. Richards & W. A. Renandya (Eds.), *Methodology in language teaching: an anthology of current practice*. Cambridge: Cambridge University Press.

Lumley T. (2002). Assessment criteria in a large-scale writing test: what do they really means? *Language Testing*, 19(3), 246-276.

Lynch, B. K. & McNamara, T. F. (1998). Using G-Theory and many facet Rasch measurement in the development of performance assessments of ESL speaking skills of immigrants, *Language Testing,* 15(2), 158-188.

[ Downloaded from ndea10.khu.ac.ir on 2026-11-06 ]

McNamara, T. F. (1996). *Measuring second language performance*. London and New York: Longman.

McNamara, T. F. (2002). *Language Testing*. Oxford: Oxford University Press.

North, B. & Schneider, G. (1998). Scaling descriptors for language proficiency scales, *Language Testing, 15*(2), 217-203.

Nunan, D. (1999). *Second language teaching and learning*. Boston: Heincle & Heincle Publishers.

Raimes, A. (1991). Out of the woods: Emerging traditions in the teaching of writing, *TESOL Quarterly,* 25 (3), 407-30.

Raimes, A. (2002). Ten steps in planning a Writing Course and Training teachers of writing. In J. C. Richards & W. A. Renandya (Eds.), *Methodology in language teaching: an anthology of current practice.* Cambridge: Cambridge University Press.

Seow, A. (2002). The writing process and process writing. In J. C. Richards & W. A. Renandya (Eds.), *Methodology in language teaching: an anthology of current practice.* Cambridge: Cambridge University Press.

Silva, T. & Matsuda, P. K. (2002). Writing. In N. Schmitt (Eds.), *An introduction to applied linguistics* (pp. 251-266). London: Arnold.

TOEFL Writing Scoring Guide (2007). *Educational testing service*. Retrieved from: *http://www.ets.org/portal/site/ets/menuitem.1488512ecfd5b8849a7713bc3921509/?vgnextoid=e7d72d3631df4010VgnVCM10000022f9190RCRD&vgnextchannel=db35197a484f4010VgnVCM10000022f 5190RCRD*

Weigle S. C. (1998). Using FACETS to model rater training effects, *Language Testing,* 15(2), 263-287

Weigle S. C. (2002). *Assessing writing*. Cambridge: Cambridge University Press.

[ Downloaded from ndea10.khu.ac.ir on 2026-11-06 ]

Weir, C. (1993). *Understanding & developing language tests*. Hertfordshire: Prentice Hall.

White, R. & Arndt, V. (1991). *Process writing*. London: Longman.

Williams, J. G. (2003). Providing feedback on ESL students' written assignments, *The Internet TESL journal (Online)* IX (10). Retrieved from: http;//iteslj.org/Techniques/WilliamsFeedback.html

Zamel, V. (1983). The composing process of advanced ESL students: six case studies, *TESOL Quarterly,* 17(2), 165-187.

**Appendix A**

**The suggested rating scales (sub-scales)**

The final revised formed of the rating scales (sub-scales) are presented here:

Rating scales for each stage of writing process

**Brainstorming sub-scale**

| | | | | | |
|---|---|---|---|---|---|
| The extent to which a certain number of ideas is generated. | 0 | 1 | 2 | 3 | 4 |
| The extent to which ideas are developed into concepts. | 0 | 1 | 2 | 3 | 4 |
| The extent to which ideas address different aspects (negative, positive, etc.) of the assigned topic (prompt). | 0 | 1 | 2 | 3 | 4 |
| The extent to which ideas are diverse representing facts, opinions, and examples. | 0 | 1 | 2 | 3 | 4 |

Total score: ☐

[ Downloaded from ndea10.khu.ac.ir on 2026-11-06 ]

**Outlinig sub-scale**

| | | | | | |
|---|---|---|---|---|---|
| The extent to which the outline is detailed, consisting of subordinate and coordinate ideas. | 0 | 1 | 2 | 3 | 4 |
| The extent to which the outline presents relevant and logical arrangements (groupings) of subordinate and coordinate ideas. | 0 | 1 | 2 | 3 | 4 |
| The extent to which the outline has subordinate ideas to support coordinate ideas. | 0 | 1 | 2 | 3 | 4 |
| The extent to which the outline reflects application of ideas in the brainstorming stage. | 0 | 1 | 2 | 3 | 4 |
| The extent to which the outline has been applied to the creation of the drafts in the drafting stage. | 0 | 1 | 2 | 3 | 4 |

Total score: ☐

**Drafting sub-scale**

| | | | | | |
|---|---|---|---|---|---|
| The extent to which the draft is a fluent piece of writing: the extent to which it reflects the ability to produce a smooth flow of ideas; the ability to produce continuous pieces of writing without causing difficulty or break down of communication. | 0 | 1 | 2 | 3 | 4 |
| The extent to which the draft presents a central idea to be communicated to the audience. The extent to which the central idea is stated in the thesis statements and it is supported by the topic sentences of the body paragraphs. | 0 | 1 | 2 | 3 | 4 |
| The extent to which the draft provides relevant information about the central idea of the prompt through supporting sentences. The extent to which the draft sticks to the main idea and no deviation from the main topic of discussion is observed. | 0 | 1 | 2 | 3 | 4 |

[ Downloaded from ndea10.khu.ac.ir on 2026-11-06 ]

| | | | | | |
|---|---|---|---|---|---|
| The extent to which the draft can attract the readers, attentions, present blueprints, has body paragraphs with topic sentences and supporters, and has a conclusion which is linked to the introductory paragraph and the thesis statement. The extent to which the final draft presents an organized and clear progression of ideas appropriately linked. | 0 | 1 | 2 | 3 | 4 |

Total score: ☐

**Editing sub- scale**

| | | | | | |
|---|---|---|---|---|---|
| The extent to which the final draft contains structures with few noticeable grammatical errors or word order problems. | 0 | 1 | 2 | 3 | 4 |
| The extent to which the final draft features a precise and effective use of vocabulary. | 0 | 1 | 2 | 3 | 4 |
| The extent to which the final draft is coherent. The meanings of the sentences are linked logically via the mechanical (cohesive) devices throughout the text. | 0 | 1 | 2 | 3 | 4 |
| The extent to which the final draft presents relevant and adequate information regarding the assigned topic. The extent to which no gaps or redundant information is observed. | 0 | 1 | 2 | 3 | 4 |
| The extent to which the final draft features no inaccuracies in spelling and punctuation. | 0 | 1 | 2 | 3 | 4 |

Total score: ☐

Since based on the statistical analyses performed, editing and drafting were highly loaded on Factor 1, namely Language Knowledge., combining the drafting and editing sub-scales would result in a general rating scale for assessing the students' final products. The next table displays the results of this combination.

[ Downloaded from ndea10.khu.ac.ir on 2026-11-06 ]

**The writing rating scale**

| | | | | | |
|---|---|---|---|---|---|
| **Fluency**: the extent to which the script is a fluent piece of writing; the extent to which it reflects a smooth flow of ideas; the extent to which the script shows the abilities of the writer to produce continuous pieces of writing without causing difficulty or break down of communication. | 0 | 1 | 2 | 3 | 4 |
| **Development:** The extent to which the draft presents a central idea to be communicated to the audience. The extent to which the central idea is logically expanded through the essay. | 0 | 1 | 2 | 3 | 4 |
| **Relevance**: The extent to which the draft provides relevant information about the central idea of the prompt in supporting sentences. The extent to which the draft sticks to the main idea. The extent to which no deviation from the main topic of discussion is observed. | 0 | 1 | 2 | 3 | 4 |
| **Organization**: The extent to which the draft attracts the readers' attentions, presents a short summary of the rest of writing (the blueprints), has body paragraphs with topic sentences and supporters, and has a conclusion which is linked to the introductory paragraph and the thesis statement. The final draft presents an organized, and clear progression of ideas appropriately linked. | 0 | 1 | 2 | 3 | 4 |
| **Grammar**: The extent to which the final draft contains structures with few noticeable grammatical errors or word order problems. | 0 | 1 | 2 | 3 | 4 |
| **Vocabulary**: The extent to which the final draft features the precise and effective use of vocabulary. | 0 | 1 | 2 | 3 | 4 |
| **Coherence**: The extent to which the draft is coherent. The extent to which the meanings of the sentences are linked logically and via the use of the mechanical (cohesive) devices throughout the text. | 0 | 1 | 2 | 3 | 4 |
| **Mechanics**: The extent to which the final draft features no inaccuracies in spelling and punctuation. | 0 | 1 | 2 | 3 | 4 |

Total score: ☐

[ Downloaded from ndea10.khu.ac.ir on 2026-11-06 ]

**Appendix B**
**The writing tasks**
Writing task in this study included a number of writing prompts plus a series of writing instructions which guided the students to perform at each stage of the writing process. One example is provided here.

***1. Writing Prompt***
***Prompt:*** *Rapid growth of global communication has become a threat for the cultural identity of small and weak countries. Do you agree or disagree with the above statement. Use specific reasons and examples to support your opinion.*

***2. Writing Instructions***
Please follow the following stages while writing on the topic. Use a separate answer sheet for each stage. Label each stage on your answer sheet.

**A.** ***BRAINSTORMING:*** *Focus on the topic and write down whatever ideas that come to your mind freely and randomly.*
**B.** ***OUTLINING:*** *Use your ideas in Stage (A) and try to group the relevant ideas together. (1) You may create a spidergram first (optional), (2) write an outline for a five-paragraph essay using the ideas you have generated in Stage (A). You may also generate some new ideas, if necessary, in order to complete your outline.*
**C.** ***DRAFTING:*** *Now, start writing your essay using the outline you have created in Stage (B). Create at least two drafts: one for a one-paragraph essay, and another for a five-paragraph essay.*
**D.** ***EDITING:*** *Edit, revise, and finalize your five-paragraph draft. Apply necessary corrections regarding grammar, vocabulary, unity, coherence, spelling, and punctuation marks. After you have made all the corrections, write the final draft of your essay.*

[ Downloaded from ndea10.khu.ac.ir on 2026-11-06 ]