



Kouros Akef
Karim Hajhashemi

A Model for Developing Rating Scale Descriptors

Assessing EFL Writing Process

 **LAMBERT**
Academic Publishing

Kouros Akef
Karim Hajhashemi

A Model for Developing Rating Scale Descriptors

**Kourosh Akef
Karim Hajhashemi**

**A Model for Developing Rating Scale
Descriptors**

Assessing EFL Writing Process

LAP LAMBERT Academic Publishing

Impressum/Imprint (nur für Deutschland/only for Germany)

Bibliografische Information der Deutschen Nationalbibliothek: Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

Alle in diesem Buch genannten Marken und Produktnamen unterliegen warenzeichen-, marken- oder patentrechtlichem Schutz bzw. sind Warenzeichen oder eingetragene Warenzeichen der jeweiligen Inhaber. Die Wiedergabe von Marken, Produktnamen, Gebrauchsnamen, Handelsnamen, Warenbezeichnungen u.s.w. in diesem Werk berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, dass solche Namen im Sinne der Warenzeichen- und Markenschutzgesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürften.

Coverbild: www.ingimage.com

Verlag: LAP LAMBERT Academic Publishing GmbH & Co. KG
Heinrich-Böcking-Str. 6-8, 66121 Saarbrücken, Deutschland
Telefon +49 681 3720-310, Telefax +49 681 3720-3109
Email: info@lap-publishing.com

Approved by: IAU, Science and Research Campus, 2007

Herstellung in Deutschland (siehe letzte Seite)
ISBN: 978-3-659-16513-9

Imprint (only for USA, GB)

Bibliographic information published by the Deutsche Nationalbibliothek: The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at <http://dnb.d-nb.de>.

Any brand names and product names mentioned in this book are subject to trademark, brand or patent protection and are trademarks or registered trademarks of their respective holders. The use of brand names, product names, common names, trade names, product descriptions etc. even without a particular marking in this works is in no way to be construed to mean that such names may be regarded as unrestricted in respect of trademark and brand protection legislation and could thus be used by anyone.

Cover image: www.ingimage.com

Publisher: LAP LAMBERT Academic Publishing GmbH & Co. KG
Heinrich-Böcking-Str. 6-8, 66121 Saarbrücken, Germany
Phone +49 681 3720-310, Fax +49 681 3720-3109
Email: info@lap-publishing.com

Printed in the U.S.A.

Printed in the U.K. by (see last page)

ISBN: 978-3-659-16513-9

Copyright © 2012 by the author and LAP LAMBERT Academic Publishing GmbH & Co. KG and licensors

All rights reserved. Saarbrücken 2012

This book is dedicated to the Sunshine of My Heart,

Sepideh

And To the Apple of My Eye,

Sepelir

ACKNOWLEDGEMENTS

This book would not be writing without the help, endeavors, and encouragements of many people. I am very much indebt of my best friends and colleagues, Dr. Bagheridoost and Dr. Gal'eh, for their kind co-operations and assistance in the revision of the rating scales, as well as for taking the burden of rating such an endless number of writing samples.

I do not really know how I can express my words of appreciation to my dear friend Mr. Esmkhani who scholarly found the necessary sources, diligently worked on the data, and patiently helped me with the SPSS software. Without his collaborations, the statistical analysis in this study would be a nightmare.

And my heartiest gratitude certainly goes to the light of my life, my beloved wife, without whose wholehearted love, caring, supports, and encouragements nothing of this would be possible.

Also, I owe a big thank you to my little son who always missed playing with a caring daddy, and who always put up with his room full of scattered books and papers.

And the last but not the least, I wish to thank my parents without whose cares, hopes, supports, tears, and prayers, I would not be where I am now today.

Kouros Akef

PREFACE

The role of writing in social interactions is becoming more and more evident in modern communities. The increasing use of personal computers and the rapid growth of the Internet have emphasized the importance of writing skill among the four skills. As a result, the ability to speak and write a second language is becoming widely recognized as an important skill for educational, business, and personal reasons.

The use of performance assessment as a measure of writing language skills in English as a second or foreign language has become a common practice all over the world. However, the process of assessment and rating EFL learners' writing abilities in high schools and, especially, in academic environments is still done based on traditionally accepted principles in educational systems. In these assessments, the examinees generally respond to some assigned topics in order to create a grammatically well-formed written product, but their efforts, and the achievements they make for creating this piece of writing are not taken into account in their final scores. Therefore, the aim of this book is to provide an assessment model which can help teachers to observe and evaluate student writers' performances during the writing process.

Relying mainly on the final written products for scoring writing performance disregards the remarkable progress and achievements the students make as they are passing through each stage of the writing process. It is hoped that applying such a model for writing assessment can provide teaching practitioners with a better diagnostic tool for the evaluation of learners who

are more likely to perform differently in different stages of the writing process.

Kouros Akef
Karim Hajhashemi

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	ii
PREFACE	iii
LIST OF TABLES	viii
LIST OF FIGURES	xii
LIST OF APPENDICES	xiii
1 INTRODUCTION	1
1.1 Background and Purpose	1
1.2 Statement of the Problem	11
1.3 Research Questions	13
1.4 Hypotheses	16
1.5 Definitions of the Key Terms	18
1.6 Limitations and Delimitations of the Study	24
1.7 Significance of the Study	25
2 LITERATURE REVIEW	28
2.1 On the Nature of Writing Ability	28
2.2 Second Language Writing	41
2.3 Approaches to Teaching Second Language Writing	42
2.4 The White and Arndt's Model	46
2.5 Generating	52
2.6 Focusing	54
2.7 Structuring	55
2.8 Drafting	57
2.9 Evaluating	58
2.10 Re-viewing	59
2.11 Seow's Model	62
2.12 Drafting	64
2.13 Responding	65

2.14	Revising	65
2.15	Editing	66
2.16	Evaluating	67
2.17	Post-writing	68
2.18	The role of feedback	69
2.19	Language assessment	73
2.20	Writing as performance assessment	80
2.21	Scoring procedures for writing assessment	85
2.22	Establishing a rating procedure	86
2.23	Raters' problem	87
2.24	Criteria for making judgement	88
2.25	Rating scales	89
2.26	Type of rating scales	92
2.27	Primary trait scoring	92
2.28	Holistic scales	94
2.29	Analytic scales	96
2.30	Designing scoring scales	98
2.31	Writing scale descriptors	101
2.32	Using the scales	106
2.33	Dealing with the problems	109
	2.33.1 Raters' training	109
	2.33.2 Evaluating scoring procedures	110
	2.33.3 Assessing reliability of scores	111
	2.33.4 Assessing validity of scoring procedures	112
	2.33.5 Evaluating the practicality of scoring procedures	113
2.34	Conclusion	113
3	MATERIALS AND METHODS	115
3.1	Phase one: Designing scale descriptors	116

3.2	Subjects	116
3.3	Instrumentation	117
3.4	Essay writing prompts	118
3.5	Suggested rating scales	118
3.6	Procedures	119
3.7	Teaching Procedure	119
3.8	Data collection	122
3.9	Designing rating scales	123
3.10	Rater Training	157
3.11	Phase two: Statistical procedures	158
4	RESULTS AND DISCUSSION	160
4.1	Description of the Data	161
4.2	Factor Analysis	166
4.3	Interpretation	181
4.4	Discriminant analysis	188
4.5	Interpretation	207
4.6	Multiple Regression Analysis	210
4.7	The problem of collinearity	225
4.8	Interpretation	232
4.9	Conclusion and Discussion	234
5	CONCLUSION, IMPLICATIONS, AND SUGGESTIONS	245
5.1	Implications for language teaching	248
5.2	Implications for language testing	250
5.3	Implications for rater training	255
5.4	Suggestions for further research	256
5.5	Final remarks	259
	REFERENCES	260
	APPENDICES	266

LIST OF TABLES

Table	Page
2-1: Peer responding checklist (Seow 2002, p.318)	65
2-2: Old and new paradigms in second language teaching (Richards & Renandya, 2002, p.335)	76
2-3: Rating scale, Occupational English test for health professionals (McNamara, 2000, p. 41)	91
2-4: Rating Scale about knowledge of syntax (Bachman &Palmer 1996, p.214)	102
3-1: A holistic scale of students' levels of performance on each stage of writing process	125
3-2: A general analytic rating scale for brainstorming scripts	129
3-3: An analytic rating scale for brainstorming scripts	131
3-4: A general analytic rating scale for outline scripts	137
3-5: An analytic rating scale for outline scripts	139
3-6: A general analytic rating scale for drafting scripts	145
3-7: An analytic rating scale for drafting scripts	147
3-8: A general analytic rating scale for editing scripts	151
3-9: An analytic rating scale for editing scripts	153
4-1: The designed rating scale and its components	162
4-2: Correlation coefficient between the raters and the inter-rater reliability estimates	163
4-3: The Analysis of Variance for the rating produced by the raters	164
4-4: Intra-rater reliability estimation for Rater 1	164

4-5: Descriptive statistics for the 20 variables defined for the stages of writing	165
4-6: Variables entered into factor analysis study	166
4-7: KMO and Bartlett's test	167
4-8: Communalities of the variables	170
4-9: Eigenvalues and the variance explained	171
4-10: Communalities after Central idea was dropped from the analysis	174
4-11: Unrotated factor matrix ^a	175
4-12: Pattern matrix for oblique rotation ^a	177
4-13: Structure matrix for oblique rotation	179
4-14: Factor Correlation Matrix	180
4-15: The arrangement of variables based on factor loadings	182
4-16: Variables measuring underlying constructs	184
4-17: Pattern matrix for the four-factor model model	185
4-18: The analysis of case processing summary	190
4-19: Normal distribution of the discriminator variables	190
4-20: The group statistics	191
4-21: Tests of Equality of Group Means	192
4-22: Pooled within-groups matrices	194
4-23: Covariance matrices	195
4-24: The result of Box's M test	196
4-25: Eigenvalues for discriminant functions	197

4-26: Wilk's Lambda and chi -square test	199
4-27: Standardized canonical discriminant function coefficients	200
4-28: Structure matrix	201
4-29: Group Centroids	202
4-30: Classification results	204
4-31: Classification function coefficients	207
4-32: Variables entered/removed	211
4-33: Model summary	212
4-34: Analysis of Variance for model fitness ^b	213
4-35: Regression coefficients	213
4-36: Excluded variables	215
4-37: Collinearity Diagnostics	216
4-38: Outliers or the students who did not fit the model	218
4-39: The result of the first round of the analysis	221
4-40: The result of the second round of the analysis	223
4-41: The result of the third round of the analysis	223
4-42: Multicollinearity diagnostics	226
4-43: Correlation coefficient between variables	227
4-44: Collinearity statistics	228
4-45: Principal components	229
4-46: Collinearity statistics for principal components	230
4-47: collinearity diagnostics for the two extracted component	230

4-48: the principal component summary	231
4-49: the regression coefficients for the two principal components	232
4-50: Pattern matrix of the factors ^a	237
4-51: Variables measuring underlying constructs	238
4-52: The final modified form of the brainstorming sub-scale	239
4-53: The final form of outlining sub-scale	240
4-54: The final form of drafting sub-scale	241
4-55: The final form of the editing sub-scale	242
4-56: General writing rating scale	243

LIST OF FIGURES

Figure	Page
1-1: The characteristics of performance-based assessment (adapted from McNamara, 1996, p. 9)	5
1-2: A model of process writing (White & Arndt, 1991, p.4)	8
2-1: A model of process writing (White & Arndt, 1991, p.4)	47
2-2 Crop Spidergram (White & Arndt, 1991, p. 81)	56
2-3: The Writing Process (Seow, 2002 , p.315)	63
3-1: A brainstorming script created by one of the student in generating ideas stage	133
3-2: A spidergram created by one of the students	135
3-3: An outline created by one of the students	140
3-4: Part of a draft created by one of the students	148
3-5: A final draft created by one of the students	154
3-6: Part of a final draft created by one of the students	156
4-1: The scree plot	172
4-2: partial and simple correlations	179
4-3: Factor plot in rotated factor space	180
4-4: components of language competence, Bachman (1990, p. 87)	186

LIST OF APPENDICES

Appendix	Page
A: Writing Prompts	266
B: Writing Instructions	267
C: Teaching Syllabus	268
D: TOEFL Writing Scoring Guide	273

CHAPTER 1

INTRODUCTION

1.1 Background and Purpose

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.

Over the years, different approaches have been adopted for teaching and assessing writing (Raimes, 1991). Traditionally, writing was viewed as transcribed speech. It was often assumed that the acquisition of spoken language was sufficient for, and had to take precedence over the learning of written language. Therefore, teachers mostly avoided introducing writing early in the process of language learning because they believed that the difference between pronunciation and spelling would interfere with the proper learning of speech (Silva & Matsuda, 2002). The primary focus of this approach was on formal accuracy. Teachers were required to employ a controlled program of systematic habit formation in order for the learners to avoid errors. 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 language learners 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), 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 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. This reaction was mostly due to the prescriptivism and linearity inherited in product-oriented approaches.

Dissatisfaction with the product-oriented approaches paved the way for the emergence of process approach to writing. 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 (Silva & Matsuda, 2002).

Unfortunately, the pure form of the process approach has not won widespread acceptance in the academic environment although many instructors have adapted some of its features in their teaching methodology. In academic contexts, the concern in most fields of study is that a learner should be able to perform academic writing tasks which satisfy the academic community, such as essay exams. These have little to do with a process orientation (Weir, 1993).

In other words, an important issue here 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. However, writing assessment procedures in academic contexts have 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; Bachman, 1991; J. D. Brown & Hudson, 1998; Chapell & Brindly, 2002; Genesee & Upshur, 1996; Gipps, 1994).

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, 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) assessment, whose goal is to “gather evidence about how learners are approaching, processing, and completing real-life tasks in a particular domain” (Huerta-Macías, 1995, p. 9).

McNamara (1996) has also emphasized that performance assessment is essentially a methodological issue and has a distinctive feature which is the process of the performance task. This follows that one needs to adopt an appropriate model based on which he can establish his assessment. One model proposed by Kenyon (1992, cited in McNamara, 1996) seems to be appropriate and also applicable for the purpose of this study. It provides a schematic representation of the features of a typical second language performance test in comparison with a traditional language test:

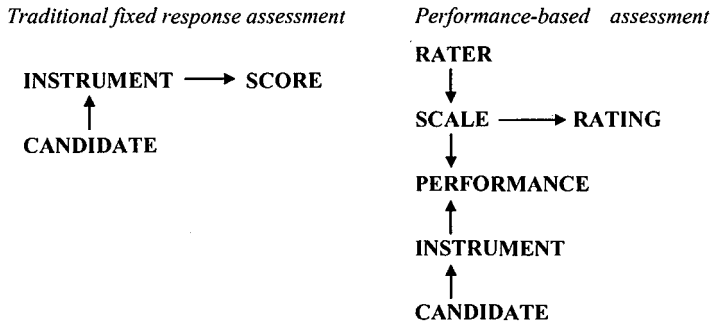


Figure 1-1: The characteristics of performance-based assessment (adapted from McNamara, 1996, p. 9)

As it is clearly shown in this model, in the performance-based assessment the candidate's performance is rated or judged according to a scale.

A true performance-based assessment is distinguished from the traditional measurements in terms of two factors: a performance process of the

examinees which is observed and an agreed judging process (e.g., a rating scale) by which the performance process is judged (McNamara, 1996).

Fitzpatrick and Morrison (1971) offer a general definition of performance tests as “one in which some criterion situation is simulated to a much greater degree than is represented by the usual paper-and-pencil test” (p. 238). Fitzpatrick and Morrison further explain that performance assessments can cover both processes and products, and the term performance assessment is short hand for fuller performance and product evaluation.

The tension between process and product approaches to the teaching of writing, as Nunan (1999) comments, has become a very controversial issue. Product-oriented approaches focus on the examinees’ final product which should be a coherent, error-free text. On the other hand, process approaches focus on the steps involved in drafting and redrafting a piece of work. Proponents of process writing recognize and accept the reality that there will never be a perfect text, but that one can get closer to perfection through producing, reflecting on, discussing, and reworking successive drafts of a text (Nunan, 1999).

However, despite its attractive nature, process writing has been criticized on several grounds. Some scholars, for example, argue that through process writing learners hardly produce factual writing which fosters the development

of critical thinking (Martin, 1985). Others believe that process writing is necessary, but not sufficient, for achieving language proficiency. Learners need structures and models to practice, and they need to learn how to produce an acceptable product on demand (Raimes, 1991). White and Arndt (1991), suggesting a practical pattern for process writing, point out that the process approach aims at helping the learner to develop a set of skills. Therefore, models (products) and procedures (processes) are both needed in writing classrooms.

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 view writing as a complex, cognitive process that requires sustained intellectual effort over a considerable period of time. They suggest that producing a text involves six recursive (nonlinear) procedures (Figure 1.2).

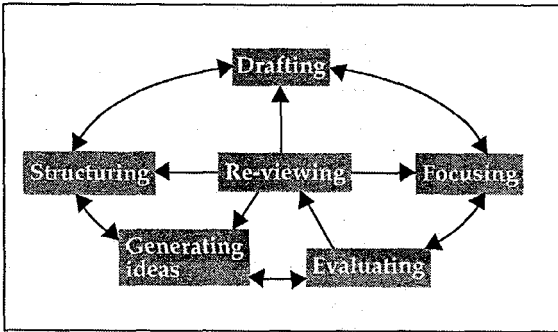


Figure 1-2: A model of process writing (White & Arndt, 1991, p.4)

As it is displayed in Figure 1.2, the six stages of generating ideas, focusing, structuring, drafting, reviewing, and evaluating are all mutually related. That is, each stage can affect its neighboring stages and at the same time receives their effects. For example, in the structuring stage, student writers use the ideas they have generated in the generating ideas stage, and while they are organizing these ideas in the structuring stage, they may generate further new ideas in order to fit into the organization structure they have created. According to White and Arndt (1991), reviewing stage influences all stages and, in turn, it is influenced by the evaluation stage. This means that evaluation and reviewing monitor every stage of the writing process.

Another important element in writing assessment is the rating scale that is used. 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). Of these elements, two parameters are of major concern in scoring: defining the rating scale and ensuring that raters use the scale appropriately and consistently. McNamara (1996) 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.

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 further mention 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 scale definitions, Bachman and Palmer argue that the scale definition should include:

1. the specific features of the language sample to be rated with the scale,
2. the definition of scale level in term of the degree of mastery of these features. (p. 213)

According to Bachman and Palmer, “while inefficiency and unreliability are potential problems [with rating scales], they are by no means insurmountable. With sufficient planning and development, rating scales can be highly reliable and relatively efficient”(p. 219).

1.2 Statement of the Problem

Regarding the above-mentioned issues, this study aims at designing an appropriate model for the assessment of Iranian 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 Iranian 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. In order to carry out this study, the following sub-problems were taken into consideration:

1. To define different stages of the writing process in terms of four operational stages based on the ideas suggested by Raimes (1991), and White and Arndt (1991). These stages include:
 - a) Generating Ideas (*brainstorming*)¹
 - b) Structuring (*outlining*)

¹ Brainstorming is one the techniques which is usually used in the stage of generating ideas. Similarly outlining is also one of the techniques of the structuring stage. For the sake of convenience, and because the researcher’s collaborators were more familiar with the terms brainstorming and outlining, these two terms are used to refer to their corresponding stages of writing throughout this research.

- c) Drafting (*writing and revising*)
 - d) Editing (*reviewing/ evaluating*)
2. To classify the students' scripts into four ability groups or levels such as VERY GOOD, GOOD, FAIR, and POOR based on the quality of performance observed in the student's scripts created in each stage of the writing process.
 3. To determine how many distinguishing features (or independent variables) can be found in the students' scripts that can indicate the level of performance of student writers in each stage of the writing process.
 4. To define scoring descriptors based on the identified features (or variables) for each performance level, as well as to assign scores to each level descriptor.
 5. To use the scoring scales for rating the student writers' scripts created in each stage of the writing process.
 6. To perform statistical analyses to determine the reliability, validity, and the usefulness of the scoring scales.

To sum up, the purpose of this study is to determine whether appropriate scoring scales can be designed for each of the defined stages of the writing process and also to determine to what extent these scoring scales can assess the performances of the Iranian EFL learners in an academic writing task.

1.3 Research Questions

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) observed in each script. Therefore, the following question is posed:

1. Is there any distinguishing feature in the students' scripts at each stage of the writing process?

If the answer to the question were positive, then those features would be used as a basis for the development of a rating scale for that very stage.

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), factor analysis, discriminant analysis, and multiple regression analysis were consulted in order to check the reliability, as well as the validity, of assessments resulted from the application of the rating scales.

An important issue to be dealt with in the quantitative phase of this study was to find out how far the scale descriptors can produce consistent results. Therefore, the next question addresses the inter-rater reliability of the designed rating scales:

2. Is there any statistically significant difference among the ratings made by the three raters for the evaluation of the student writers' scripts?

The inter-rater and the intra-rater reliability estimations and the Analysis of Variance were used to provide answer to this question.

In order 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. For example, in this study, the students' abilities in every stage of the writing process were assessed based on their ratings of such variables as number of ideas, diversity of ideas, relevance, coherence, fluency, grammar and vocabulary. Now, the question is "what underlying constructs do all these variables actually measure?"

3. What underlying constructs are measured by the variables assessed through the application of the rating scale?

The next purpose of this study was to see whether the variables in this study (i.e., the four stages of brainstorming, outlining, drafting, and editing) can be used to efficiently discriminate among the different levels of writing products as assessed by the TOEFL writing rating scale. In other words, the purpose was to determine to what extent the four variables mentioned could discriminate or differentiate among the four levels of TOEFL writing proficiency. Therefore, the question raised here is:

4. Can the student writers' performances be significantly classified in to four levels of performance -- VERY GOOD, GOOD, FAIR, POOR, --

based on the scores they have received in the four stages of brainstorming, outlining, drafting and editing?

The statistical procedure to deal with this question was discriminant analysis.

Yet, another aim of this study in the quantitative phase was to determine to what extent the student writers' scores in every stage of the writing process can predict the student writers' TOEFL based writing scores. In other words, the aim was to find out if the students' scores in each stage of the writing process could predict, as well as explain, the variance in their TOEFL based writing scores. Therefore the relevant question would be:

5. Do the students' scores in each stage of the writing process significantly predict the students' TOEFL based writing scores?

Multiple regression analysis was used to provide answer to the above questions.

1.4 Hypotheses

The first phase of this study is qualitative, hence hypothesis generating. In fact, the first stage of a qualitative study in terms of Selinger and Shohamy (1989) is heuristic, inductive or hypothesis generating; any hypothesis testing occurs at the later stages after the preliminary hypotheses are generated. The

researcher, assisted by two essay-writing instructors, went through half of the data or student writers' scripts and extracted the distinctive features for every stages of the writing process. These identified features form the hypotheses for the first question of this study. Subsequently, the researcher and his three assistants went through the rest of the data to see whether the same features were mined. The degree of match between the features obtained in these two stages show how much the hypotheses hold true.

After the ratings of the student writers' scripts, the issue at stake was to find out if the rating scale can produce consistent results and also if there was any significant difference between the ratings of the raters. Therefore the null hypothesis for the second question of this study is:

H₀₂: There is no statistically significant difference among the ratings made by the three raters for the evaluation of the student writers' scripts.

Exploratory factor analysis is also hypothesis generating in nature. This is because no pre-identified model is imposed on the data. What is done is performing an explanatory factor analysis on the data to see what patterns would emerge. Actually what comes out of such explanatory factor analysis can be used as a possible model to serve as an input to a hypothesis-testing confirmatory factor analysis which requires conducting another line of research on further data.

The hypotheses for other questions will be as follows:

H₀₄: The student writers' performances can not be significantly classified in to four levels of performance -- VERY GOOD, GOOD, FAIR, POOR, -- based on the score they have received at each stages of the writing process.

H₀₅: The students' scores in each stage of the writing process do not significantly predict the students' TOEFL based writing scores.

1.5 Definitions of the Key Terms

The following terms are used in this study. They are defined according to the purpose of this study.

Assessment refers to the act of collecting information and making judgments about language learners' knowledge of a language and their ability to use it (Bachman & Palmer, 1996). *Performance assessment* aims to make inferences more directly from test performance to performance outside the test setting. In performance tests, the assessment is carried out in a context where students are involved in an act of communication (McNamara, 1996, 2000).

Coherence refers to the relationships which link the meanings of sentences in a text. Generally, it refers to the logical and/or the mechanical connection

between the parts of a piece of writing. Coherence is a semantic property of discourses, based on the interpretation of each individual sentence relative to the interpretation of other sentences (van Dijk, 1977).

Cut-off point is the point on a scoring continuum at which a classification decision is made, e.g., whether a subject has passed or failed (McNamara, 2000)

Drafting is one of the stages of the writing process in which student writers start producing stretches of written texts. In this stage, the writers focus on the fluency of writing and are not preoccupied with the grammatical accuracy or the neatness of the draft (Seow, 2002). Here, the writers may run through the cycle of writing-revising-rewriting several times, creating several drafts of their text (White & Arndt, 1991).

Editing is one of the stages of the writing process in which the student writers are engaged in tidying up their texts as they prepare the final draft for evaluation. They edit their own or their peers' works for grammar, spelling, punctuation, diction, sentence structure, and accuracy of supportive material such as quotations, examples, and the like (White & Arndt, 1991).

Feedback is the opinion of instructors about the content and format of student writers' scripts. Student writers use these comments to revise their written materials (Williams, 2003).

Final product refers to the final draft of the writing task.

Generating ideas is one of stages of the writing process in which student writers draw upon their schemata in order to discover a topic, identify the purpose, and discover ideas for a piece of writing. *Brainstorming* is one of the techniques used in generating ideas in which student writers think quickly and without inhibition to produce as many ideas as possible in a given area or on a given topic (White & Arndt, 1991). In this study, the term brainstorming was used for generating ideas. Therefore, for the sake of convenience, brainstorming is used as a general term to refer to the stage of generating ideas throughout this study.

Inter-rater reliability refers to the extent to which pairs of raters agree. It also refers to the degree to which the ratings of one rater are predictable from the ratings of another, based on the scores given by each for a given set of writing performance (McNamara, 2000). In other words, it refers to the tendency of different raters to give the same score to the same script (Weigle, 2002).

Intra-rater reliability refers to the extent to which a single rater is consistent across different scripts of similar quality or the same script on different occasions. It is the tendency of a rater to give the same score to the same script on different occasions (Weigle, 2002).

Moderation refers to the process of reconciling or reducing differences in the judgments and standards used by different raters in the rating procedure; this is usually done at meetings of raters in which writing performances at relevant levels are rated independently and then discussed (McNamara, 2000).

Prompt refers to a writing topic to which the subjects respond in writing.

Raters are those who judge performances on writing, using an agreed-upon rating procedure and criteria in doing so (McNamara, 2000).

Rating procedure is an agreed-upon procedure followed by writing raters in judging the quality of performances (McNamara, 2000).

Rating refers to the assessment of performance, recorded as scores on a rating scale (McNamara, 2000). *Analytic rating* refers to the rating of each aspect of a writing performance separately, for example, grammar, organization,

content, etc. (McNamara, 2000; Weigle, 2002). *Holistic rating* is the rating of a writing performance as a whole (McNamara, 2000; Weigle, 2002).

Rating scale refers to an ordered set of descriptions of typical performances in terms of their quality, used by raters in rating procedures (McNamara, 2000).

Scale descriptors (level descriptors) refer to the statements describing the character of a minimally acceptable performance at a given level (McNamara, 2000).

Script refers to the writing samples created by the student writers (Weigle, 2002).

Single-stage script refers to the script a student produces in one stage (e.g., in generating ideas) after he/she receives complete instructions about performing successfully in that stage of the writing process. A *four-stage script set* refers to the collection of scripts (usually four scripts-- one for every defined stage of the writing process in this study) a student creates as he/she is writing a five-paragraph essay on a single prompt.

Stage refers to each period of time or step in the writing process in which student writers are engaged in certain distinct activities, such as generating ideas, structuring, drafting and editing.

Structuring is one of the stages of the writing process which involves organizational process of grouping ideas together and deciding upon how to sequence them (White & Arndt, 1991). *Outlining* is a technique in the structuring stage in which student writers group their random ideas created in the stage of generating ideas in to an organized and hierarchical list of related items or ideas (Yorkey, 1970). Again for the sake of convenience, the term outlining is used to refer to the structuring stage throughout this study.

Writing fluency refers to a level of proficiency in communication which includes:

- The ability to produce the written language with ease.
- The ability to write with a good but not necessarily perfect command of vocabulary and grammar.
- The ability to communicate ideas effectively.
- The ability to produce continuous pieces of writing without causing comprehension difficulty or the breakdown of communication.

Writing process refers to the recursive, nonlinear approach, comprising of various stages such as generating ideas, structuring, drafting, revising, and

editing, that student writers employ in order to create a piece of writing (White & Arndt, 1991).

1.6 Limitations and Delimitations of the Study

In conducting the present study, the researcher has imposed certain delimitations and faced some limitations. The delimitations included the student writers (subjects) participating in this study, the raters who assisted the researcher in conducting the study, and the time of data collection. The major limitation encountered in this study was the unavailability of the needed computer programs for the data analysis intended for this study.

The subjects (student writers) in this study were EFL students taking essay-writing courses at Islamic Azad Universities, Central Tehran Branch and Karaj Branch.

The groups of participants for this study were selected during three subsequent semesters from September 2004 to February 2006.

The writing tasks consisted of a number of topics for writing five paragraph essays similar to those tasks the students should carry out in the final essay writing examination.

For scoring the student writers' scripts, the raters were selected among the language instructors who were teaching EFL writing courses for at least three years at Islamic Azad Universities, Central Tehran Branch and Karaj Branch.

The major limitation faced in this study was the unavailability of multi-faceted Rasch Computer programs, such as WINSTEPS and FACETS during the time of this study. Originally, the aim of this study was to determine to what extent the designed scale descriptors for each stage could represent the students' levels of writing ability in that very stage. In that case, it would be possible to refine the scale descriptors into the stand-alone criterion statements with known difficulty values for each stage of the writing process. Because of the unavailability of Rasch measurement programs at the time of the study, the researcher decided to use other statistical procedures, such as explanatory factor analysis, discriminant analysis, and multiple regression analysis; in order to determine the reliability and validity of the assessments made through the application of the rating scales in this study.

1.7 Significance of the Study

The role of writing in social interactions is becoming more and more evident in modern communities. The increasing use of personal computers and the rapid growth of the Internet have emphasized the importance of writing skill among the four skills. As a result, the ability to speak and write a second

language is becoming widely recognized as an important skill for educational, business, and personal reasons.

The use of performance assessment as a measure of writing language skills in English as a second or foreign language has become a common practice all over the world. However, the process of assessment and rating EFL learners' writing abilities in high schools and, especially, in academic environments is still done based on traditionally accepted principles in Iranian educational system. Language writing measurements have usually been based on student writers' final products which are mainly scored holistically for accuracy. In these assessments, the examinees generally respond to some assigned topics in order to create a grammatically well-formed written product, but their efforts, and the achievements they make for creating this piece of writing are not taken into account in their final scores.

Therefore, regarding the recent focus of attention to adopting process-oriented approaches to teaching writing skills, it seems necessary to search for an assessment model which can help teachers to observe and evaluate student writers' performances during the writing process. Relying mainly on the final written products for scoring writing performance disregards the remarkable progress and achievements the students make as they are passing through each stage of the writing process. An assessment model which can evaluate the student writer's performance in each stage of the writing process would help

teachers pinpoint where each student's weaknesses or strengths lie. Applying such a model for writing assessment can provide teaching practitioners with a better diagnostic tool for the evaluation of learners who are more likely to perform differently in different stages of the writing process.

In addition, assessment procedures have always influenced the way teachers select and present materials in the classroom. The application of an assessment model which evaluates the student's performance in each stage of the writing process can encourage the use of a process-oriented approach in classrooms. By adopting such a model, the students' final writing scores would be the sum of the scores he could gain at each stage. This can motivate the teacher to gear her teaching approach so that she can better focus on the development of skills the students need in each stage of the writing process.