بسم الله الرحمن الرحيم

فرم گردآوری اطلاعات پایان نامه ها

کتابخانه مرکزی دانشگاه علامه طباطبایی

عنوان: برسی تاثیر برنامه آموزشی مستقیم (رو در رو) بر سنجش مهارت نگارش در زبان دوم
نویسنده / محقق: هومن بیژنی
مترجم:
استاد راهنما: سرکار خانم دکتر معرفت الدین / جناب آقای دکتر فهیم **استاد مشاور / استاد داور:** جناب آقای دکتر تاج
کتابنامه: دارد **واژه نامه:** ندارد
نوع پایان نامه: بنیادی □ توسعه ای □ کاربردی ■
مقطع تحصیلی: کارشناسی ارشد **سال تحصیلی: ۱۳۸۷- ۸۸**
محل تحصیل: تهران **نام دانشگاه:** علامه طباطبایی **دانشکده:** ادبیات فارسی و زبانهای خارجی
تعداد صفحات: ۱۸۰ **گروه آموزشی:** آموزش زبان انگلیسی
کلید واژه ها به زبان فارسی: ارزشیابی، آموزش درجه بندی کننده، پایایی، پایایی ارزیاب ها، اعتبار
کلید واژه ها به زبان انگلیسی: Assessment, Rater training, Reliability, Inter-rarer reliability, Validity

Allameh Tabataba'i University
Faculty of Persian Literature and Foreign Languages

Evaluating the Effectiveness of a Face-to-Face Training Program on L2 Writing Assessment

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Masters of Arts in Teaching English as a Foreign Language (TEFL)

Thesis Advisor:
Dr. Fahimeh Marefat

Thesis Reader:
Dr. Zia Tajeddin

By:
Houman Bizhani

Tir 1388 / July 2009

We hereby approve that this thesis
By

Houman Bizhani

Entitled "Evaluating the effectiveness of a face-to-face training program on L2 writing assessment" be accepted in partial fulfillment of the requirements for the degree of Masters of Arts in TEFL.

Committee on the Oral Examination:

Advisor: Dr. Fahimeh Marefat

.....

Reader: Dr. Zia Tajeddin

.....

Examiner: Dr. Mansoor Fahim

.....

Head of the Department of English Language and Literature

Dr. Zia Tajeddin

.....

July 2009

Dedicated to:

My dear mother
who encouraged and inspired me with affection

ACKNOWLEDGMENTS

As I look back on the process of planning, running and writing the research for this thesis, I become aware how fortunate I was in having had the opportunity to work with so many intelligent, generous, and helpful people.

I would like to thank, first of all, Dr. Fahimeh Marefat, my advisor, who had been a wonderful supporter of my research and without her sincere help, approval, and influence, this study could have never been penned.

Moreover I would like to express my deepest gratitude to Dr. Zia Tajeddin, my reader, for his generous contribution, assistance, supports, patience, and constructive comments in this study.

Special thanks go to Dr. Sara Cushing Weigle, UCLA faculty member, for her sincere help and for providing me with the literature in this study.

I wish to acknowledge Dr. Ute Knoch, University of Auckland faculty member, whose great help in providing me with the helpful articles in this research encouraged me a lot.

I am also greatly in debt to Dr. Asadi from Tarbiat Moallem University, whose honest and great help in training raters helped me run this research.

Perhaps my greatest debt is to the 12 raters who gave so generously all their time in rating so many compositions, not once but three times.

Most especially I would like to thank my great mother who has been a constant support and believer in me and has shown incredible patience and good cheer during this study.

Abstract

Rater training is an essential component of any reliable writing assessment in the first or second language studies. However, little is known about the processes by which raters learn to apply the appropriate criteria in making judgments about writing samples. This study examined the process through both qualitative and quantitative approaches. Twelve raters, six inexperienced and six experienced, rated 15 essays on a topic before training, 15 additional essays on the same topic immediately after training, and another 15 essays on the same topic after training using a four-part scoring rubric (The IELTS scale) covering Organization, Structure, Vocabulary, and Punctuation.

The major findings, after the ratings were analyzed, were as follows: (1) Inexperienced raters tended to be less consistent than experienced raters before training but not afterwards. (2) Training was effective and moved all the raters towards having more consistency less severity and less bias in rating. (3) Through time, training loses its effectiveness and raters should be retrained in intervals. (4) Despite training, significant differences in severity among raters remained.

The qualitative data analysis along with the help of quantitative analyses showed that inexperienced raters were more willing to accept

authorities' comments in rating and their ratings were more improved than experienced raters because experienced raters are less likely to accept authorities' comments. The results demonstrated that training helps raters understand the crucial criteria to be used in rating essay compositions. On the other hand the results showed that all raters are capable of rating essay compositions regardless of the background or any type of experience in this area; therefore, what makes difference in rating essay compositions is a good training program. Another implication of this study is that decision makers should not charge a huge amount of money in using experienced raters for rating essay compositions, because experience does not increase rating reliability. In fact they should use inexperienced raters but with a good training program in advance.

List of Tables

List of Figures

چکیده

الف: موضوع و طرح مساله (اهمیت موضوع و هدف): با وجود آنکه اهمیت آموزش در ارزیابی مهارت نگارش مورد بحث قرار گرفته اما تحقیقات ناچیزی در رابطه با درک ارزیاب ها نسبت به امر برنامه های آموزشی انجام شده است. همچنین تحقیقات بسیار کمی در رابطه با تاثیر آموزش ارزیابی مهارت نگارش بر ارزیاب های باتجربه و بی تجربه انجام شده است. از طرف دیگر تحقیقاتی که تاکنون انجام شده اند اکثرا بر مبنای ارزیابی تاثیر آموزش در کوتاه مدت بوده اند و بنابراین تحقیق کمی در رابطه با تاثیر بلند مدت آموزش در ارزیابی مهارت نگارش انجام شده است. به همین ترتیب، تحقیقات و بررسی های بسیار کمی در رابطه با میزان تاثیر آموزش در کاهش درجه سخت گیری و یا آسن گیری ارزیاب ها و همچنین میزان سو گیری آنها انجام شده است. این رساله تمامی مسایل بالا را مورد بررسی قرار داده است.

ب: مبانی نظری شامل مرور مختصری از منابع، چهارچوب نظری و پرسشها و فرضیه ها: لیناکر (۱۹۸۹) بر این باور است که گوناگونی ارزیاب ها امری اجتناب ناپذیر و بخشی از پروسه ی ارزیابی است. او اعتقاد دارد که ارزیاب ها را نمی توان با استفاده از آموزش به میزان یکسانی از درجه سخت گیری رساند. بنابر این عملکرد آموزش نباید به گونه ای باشد که هدف از آن رساندن ارزیاب ها به تطابق با یکدیگر (پایایی ارزیاب ها) باشد بلکه باید به گونه ای باشد که هدف از آن رساندن ارزیاب ها به تطابق با خودشان (پایایی ارزیاب) باشد. بر این اساس پرسشهای زیر قابل طرح اند. **۱-** درک ارزیابها از برنامه رو در رو آموزشی در سنجش مهارت نگارش قبل و بعد آموزش چگونه است؟ **۲-** آیا پایایی ارزیابها بلافاصله بعد از برنامه ی رو در رو آموزشی افزایش می یابد؟ **۳-** آیا بعد از برنامه آموزشی ارزیابها سخت گیر تر یا آسان گیر تر می شوند؟ **٤-** آیا کاهشی در میزان سو گیری ارزیابها بعد از برنامه آموزشی وجود دارد؟ **۵-** آیا تفاوتی بین ارزیابهای باتجربه و بی تجربه در بخش مهارت نگارش قبل و بعد از برنامه آموزشی وجود دارد؟ **٦-** آیا پایایی ارزیابها در طولانی مدت پس از اتمام برنامه ی آموزشی همچنان افزایش داشته است؟

پ: روش تحقیق شامل تعریف مفاهیم، روش تحقیق، جامعه مورد تحقیق، نمونه گیری و روشهای نمونه گیری، ابزار اندازه گیری، نحوه اجرای آن، شیوه گرد آوری و تجزیه و تحلیل داده ها: به منظور پاسخ به سوالات تحقیق از یک طرح تجربی نمای کمیتی استفاده شده است. در این راستا ۶۰ زبان آموز، ۱۲ ارزیاب، ۲ استاد دانشگاه شرکت داشته اند. همچنین ۴۵ نمونه انشاء زبان آموزان، ۲ پرسشنامه، مقیاس درجه بندی آیلتس، دستور نامه زبان آموزان و ارزیابها مورد استفاده قرار گرفت. در مرحله اول این پژوهش از زبان آموزان آزمون نگارش به عمل آمد و پس از تایپ ۱۵ عدد از آنها به ارزیابها برای ارزیابی داده شد. در مرحله دوم، اولین پرسشنامه به ارزیابها داده شد و سپس برنامه آموزشی اجرا شد. در مرحله سوم پرسشنامه دوم به ارزیابها داده شد و سپس ۱۵ برگه نگارش دیگر به ارزیابها داده شد تا ارزیابی کنند. در مرحله چهارم که 2 ماه بعد از پایان برنامه آموزشی اجرا شد، ۱۵ برگه ی نگارش دیگر به ارزیابها داده شد تا ارزیابی کنند. در این پژوهش از روشهای آماری سنجش میزان همبستگی، سنجش میزان ضریب پایایی و آزمون تی استفده شد.

ت: یافته های تحقیق: بعد از آموزش تمامی ارزیاب ها تا حد بسیار بالایی به مطابقت با یکدیگر رسیدند. آموزش همچنین درجه سخت گیری و آسان گیری ارزیاب ها و همچنین میزان سو گیری آنها را تا حد بسیار زیادی کاهش داد. همچنین آموزش بر ارزیاب های بی تجربه بسیار موثر تر از ارزیاب های با تجربه بوده و نهایتا اینکه گذر زمان باعث کمرنگ شدن اثر آموزش بر ارزیاب ها می شود.

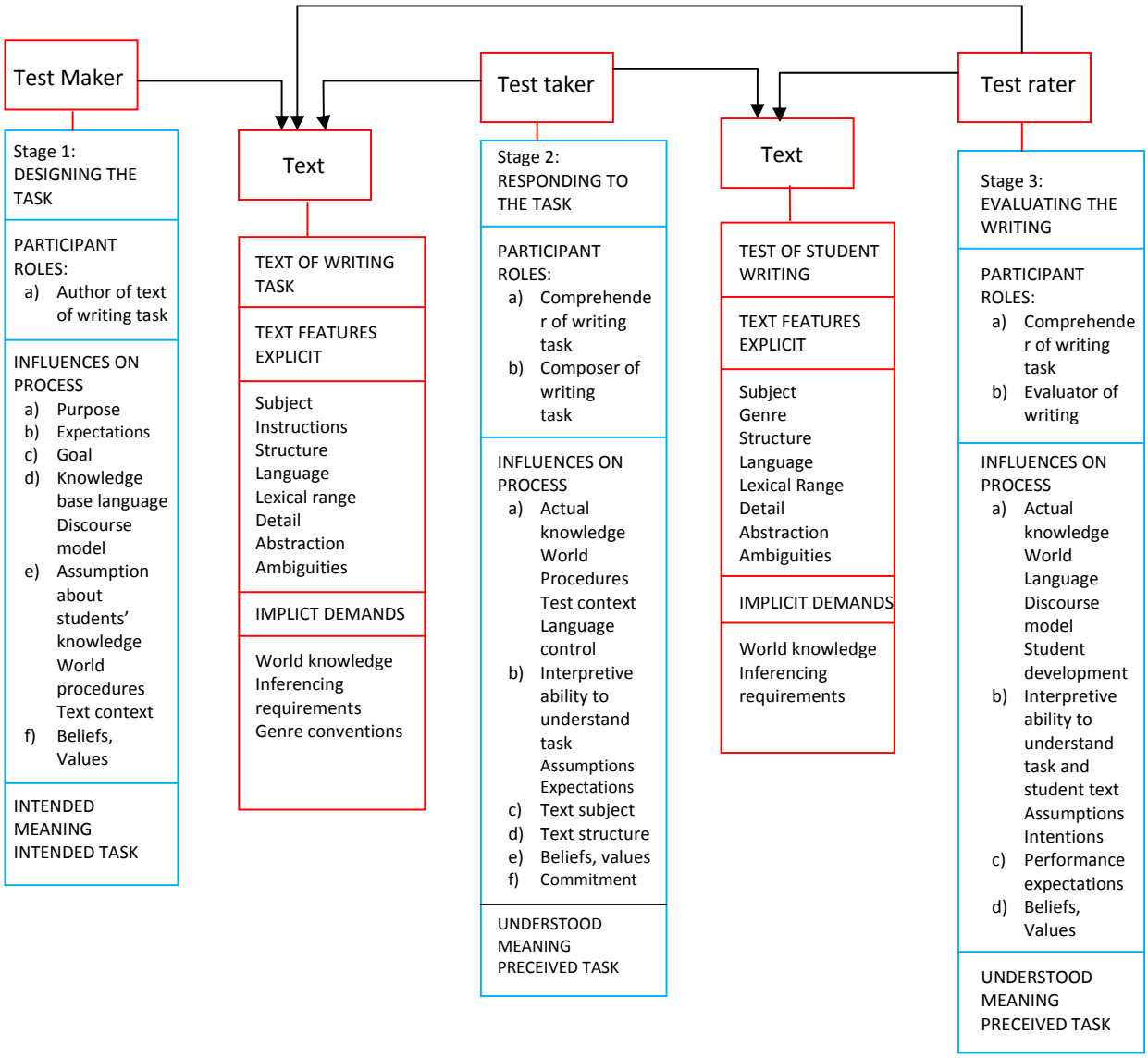
ث: نتیجه گیری و پیشنهادات: این پژوهش نهایتا به افزایش میزان پایایی بین ارزیاب ها و کاهش میزان سخت گیری و یا آسان گیری آنها انجامید. این پژوهش نشان می دهد که استفاده صرف از مقیاس باعث افزایش میزان ضریب پایایی نمی شود بلکه آنچه که میزان پایایی را افزایش می دهد آموزش مناسب است. این پژوهش نشان داد که تجربه و زمینه باعث افزایش میزان ضریب پایایی نمی شود بنابراین سیاستگزاران آموزشی بجای صرف هزینه برای استفاده از ارزیاب های باتجربه باید از ارزیاب های بی تجربه استفاده کنند و در عین حال از از برنامه های آموزشی بهره گیرند. همچنین اینکه آموزش باید به طور تناوبی برای ارزیاب ها تجدید شود. بر این اساس پیش نهاداتی برای تحقیق های بیشتر مطرح می شود که عبارتند از: استفاده از پیش نویس های کلامی بدست آمده از ارزیاب ها برای بررسی ذهنیت آنها در ارزیابی. بررسی تاثیر ویژگی های شخصیتی ارزیاب ها بر ارزیابی ایشان. استفاده از ارزیاب های بومی و مقایسه آنها با ارزیاب های غیر بومی.

صحت اطلاعات مندرج در این فرم بر اساس محتوای پایان نامه و ضوابط مندرج در فرم را گواهی می نمایم.

استاد راهنما:

سمت علمی:

رییس کتابخانه: نام دانشکده:


Test Maker
Stage 1:
DESIGNING THE TASK
PARTICIPANT ROLES:
a) Author of text of writing task
INFLUENCES ON PROCESS
a) Purpose
b) Expectations
c) Goal
d) Knowledge base language
Discourse model
e) Assumption about students' knowledge
World procedures
Text context
f) Beliefs, Values
INTENDED MEANING
INTENDED TASK
Text
TEXT OF WRITING TASK
TEXT FEATURES EXPLICIT
Subject
Instructions
Structure
Language
Lexical range
Detail
Abstraction
Ambiguities
IMPLICT DEMANDS
World knowledge
Inferencing requirements
Genre conventions
Test taker
Stage 2:
RESPONDING TO THE TASK
PARTICIPANT ROLES:
a) Comprehender of writing task
b) Composer of writing task
INFLUENCES ON PROCESS
a) Actual knowledge
World
Procedures
Test context
Language control
b) Interpretive ability to understand task
Assumptions
Expectations
c) Text subject
d) Text structure
e) Beliefs, values
f) Commitment
UNDERSTOOD MEANING
PRECEIVED TASK
Text
TEST OF STUDENT WRITING
TEXT FEATURES EXPLICIT
Subject
Genre
Structure
Language
Lexical Range
Detail
Abstraction
Ambiguities
IMPLICIT DEMANDS
World knowledge
Inferencing requirements
Test rater
Stage 3:
EVALUATING THE WRITING
PARTICIPANT ROLES:
a) Comprehender of writing task
b) Evaluator of writing
INFLUENCES ON PROCESS
a) Actual knowledge
World
Language
Discourse model
Student development
b) Interpretive ability to understand task and student text
Assumptions
Intentions
c) Performance expectations
d) Beliefs, Values
UNDERSTOOD MEANING
PRECEIVED TASK

Chapter One

Introduction

1. Introduction

1.1. Overview of the study

The ability to write effectively is becoming increasingly important in our global community, and instruction in writing is thus assuming an increasing role in both second and foreign language education. As a result, the ability to write in a second language is widely recognized as an important skill for educational, business, and personal reasons (Weigle, 2002). With the significance attached to writing in SL and FL contexts, testing writing is considered an important issue. As the role of writing in second language learning increases, there is an even greater demand for valid and reliable ways to test this ability.

In recent years, essay examinations have become a standard practice in assessing the writing skills of both first and second languages. Because such tests require subjective evaluations of writing quality, a great deal of research emphasis has been placed on achieving an acceptable level of interrater reliability in order to show that essays can be scored as fairly and constantly as possible. However, this emphasis on reliability has been at the expense of decreasing test validity (Charney, 1984); that is, the procedure for achieving higher reliability may not lead to valid judgments of writing quality. One issue which is at the heart of both reliability and validity in essay scoring is that of rater training.

In the past 30 years, holistic writing assessment has become the norm in evaluating the writing skills in both first and second language. In holistic assessment, examinees are asked to write compositions on one or more topics, and then they are scored by raters. Because these tests are scored subjectively, it is essential that raters be carefully trained to conform to some standards (Weigle, 1994b). These standards are given to raters through scoring rubrics that describe the typical characteristics of writing samples at different levels. It is evident that without training, ratings tend to be highly unreliable. A large body of literature, beginning with the work of Diederich, French, and Carlton (1998) and continuing through the present day with the work of Elder,