Wash Back in Language Testing

A Comparison of the Entrance English Major Test of State and Islamic Azad University: A Wash Back Study

ADVISER

Professor Parviz Birjandi

READER

Dr. Seyed Mohammad Alavi

BY

Jahanbakhsh Nikoopour

May 2005

Tehran
Abstract

Public examinations have always been used as instruments and targets of control in the school system (Eckstein and Noah, 1993). Their relationship with the curriculum, with teacher teaching, with student learning, and to individual chances in life is of vital importance in most societies. Consequently, a belief that assessment can leverage educational change has often led to top-down educational reform strategies (Noble and Smith, 1994). There is evidence to suggest that examinations have wash back effect on teaching and learning (Alderson and Wall, 1993). The extensive use of examination scores for various educational and social purposes in societies nowadays has made the effect of wash back a distinct educational phenomenon. This study was based on the basic model of wash back proposed by Hughes (1993). The participants being directly influenced by the Newly Developed English Major Test (NDEMT); that is, high school teachers and those students who attempted the NDEMT were the subjects of his study. Two questionnaires and an observation checklist were used to investigate the wash back effect of the NDEMT on teachers and students. The prime aim of this research put into the following ten hypotheses is to find out whether the NDEMT has wash back on Iranian EFL candidates and high school EFL teachers. (1) The subjects do not agree on the correspondence between the content of high school English courses and that of the NDEMT. (2) The subjects do not agree on the correspondence between the instructional practice of high school English courses and the types of English tests used in the NDEMT. (3) The NDEMT does not have any wash back on the content of the educational program, the evaluation methods of high school English instruction, the test construction and administration procedure, the types of test used in high school classes, the way test results are analyzed, the teachers' performance, the students' attitude toward it, and the teachers' attitude toward it.

The findings of the study showed that all these null hypotheses were rejected. In addition, the results were almost the same in state and Islamic Azad University. Moreover, the positive and negative wash back effects of the NDEMT were put forward. Some strategies were suggested to promote positive wash back effects and some for reducing the harmful ones.
# Table of Contents

**CHAPTER I Introduction** .......................................................... 1  
Background and Purpose ...................................................................... 2  
Statement of the Problem ................................................................. 7  
Definition of Terms ........................................................................... 9  
Wash back ......................................................................................... 9  
Wash back Validity .......................................................................... 10  
Content Validity ............................................................................... 10  
Construct Validity ........................................................................... 12  
Newly Developed English Major Test (NDEMT) .................................. 13  
Research Questions ......................................................................... 14  
PART I: Content and Form of the NDEMT ......................................... 15  
PART II: Wash back effect of the NDEMT ......................................... 15  
Statement of Hypotheses .................................................................... 16  
PART I: Content and Form of the NDEMT ......................................... 17  
PART II: Wash back effect of the NDEMT ......................................... 17  
Significance of Study ......................................................................... 18  

**CHAPTER II Review of Related Literature** ................................. 20  
Testing: Background ......................................................................... 21  
Arguments in Support of Konkoo ....................................................... 21  
Background and Setting .................................................................... 27  
Standardized Tests: Some Findings ................................................... 31  
What is Wash back? .......................................................................... 42  
Wash back Effect Revisited ............................................................... 50  
Wash back and Language Tests ........................................................ 57  
Participants in Wash back Process .................................................... 64  
Wash back Research ......................................................................... 70  
Research Findings ............................................................................ 75  
Test Impact: Background .................................................................. 76  
Language Teaching and Wash back .................................................. 79  
Identification of Wash back ............................................................... 88  
Wash back Processes ...................................................................... 91
CHAPTER III Methodology ........................................ 104
Introduction ................................................................... 104
Subjects .......................................................................... 105
The Demographic Features of Subjects ......................... 106
Students ......................................................................... 106
Teachers .......................................................................... 108
Instrumentation ............................................................. 109
Pilot Study .................................................................... 109
Reliability of Student Questionnaire ......................... 110
Reliability of Teacher Questionnaire ......................... 111
Reliability of Observation Checklist ......................... 112
Construct Validity of the Questionnaire .................... 112
Actual Study ................................................................ 113
Reliability of Student Questionnaire ......................... 114
Reliability of Teacher Questionnaire ......................... 116
Reliability of both Questionnaires (Teacher & Student) 118
Construct Validity of the Instruments ....................... 119
Design .......................................................................... 124
Procedure ...................................................................... 125

CHAPTER IV Data Analyses .......................................... 127
Introduction ................................................................... 128
Data Analyses .............................................................. 130
Chi-square Analysis for Test Content ....................... 166
Chi-square Analysis for Wash back Effect .................. 169
Statistical Analysis of Observation Checklist ............... 173
Demographic Features of Subjects ............................ 174

CHAPTER V Results & Conclusions ............................... 192
Introduction ................................................................... 193
Positive Washback of the NDEMT ............................ 195
Negative Washback Effects of NDEMT ..................... 197
Pedagogical Implications ............................................. 200
Suggestions for Further Research.................................................................206
Bibliography..................................................................................................208

Appendix A

Appendix B

List of Tables

CHAPTER THREE
Table 3-1: frequency tabulations for age (students)
Table 3-2: frequency tabulations for gender (students)
Table 3-3: frequency tabulations for type of education (students)
Table 3-4: frequency tabulations for field of study (students)
Table 3-5: frequency tabulations for interest (students)
Table 3-6: frequency tabulations for language level (students)
Table 3-7: frequency tabulations for the NDEMT
Table 3-8: frequency tabulations for age (teachers)
Table 3-9: frequency tabulations for gender (teachers)
Table 3-10: frequency tabulations for university degree (teachers)
Table 3-11: frequency tabulations for work experience (teachers)
Table 3-12: frequency tabulations for grade of teaching (teachers)
Table 3-13: frequency tabulations for interest (teachers)
Table 3-14: frequency tabulations for job language level (teachers)
Table 3-15: frequency tabulations for mean hours of teaching (teachers)
Table 3-16: Reliability Results of Student Questionnaire (pilot study)
Table 3-17: Reliability Results of Student Questionnaire (pilot study)
Table 3-18: Reliability Results of Teacher Questionnaire (pilot study)
Table 3-19: Reliability Results of Teacher Questionnaire (pilot study)
Table 3-20: Principal Component Factor Analysis (Pilot study)
Table 3-21: Rotated Factor Analysis (Pilot study)
Table 3-22: Reliability Results of Student Questionnaire (actual study)
Table 3-23: Reliability Results of Student Questionnaire (actual study)
Table 3-24: Reliability Results of Student Questionnaire (actual study)
Table 3-25: Reliability Results of Teacher Questionnaire (actual study)
Table 3-26: Reliability Results of Teacher Questionnaire (actual study)
Table 3-27: Reliability Results of Teacher Questionnaire (actual study)
Table 3-28: Reliability Results of Teacher & Student Questionnaire
Table 3-29: Reliability Results of Teacher & Student Questionnaire
Table 3-30: Reliability Results of Teacher & Student Questionnaire
Table 3-31: Total Variance Explained
Table 3-32: Factor Loadings
Table 3-33: Factor Loadings
Table 3-34: Factor Loadings

CHAPTER FOUR

Table 4-1: frequency tabulations for" further study in translation major" (students)
Table 4-2: frequency tabulations for" further study in teaching major" (students)
Table 4-3: frequency tabulations for" further study in literature" (Students)
Table 4-4: frequency tabulations for" higher acceptance probability" (Students)
Table 4-5: frequency tabulations for" evaluating proficiency level" (Students)
Table 4-6: frequency tabulations for" ease in competition" (students)
Table 4-7: frequency tabulations for" better job opportunity"(students)
Table 4-8: Descriptive statistics for student questionnaire
Table 4-9: Descriptive statistics for teacher questionnaire
Table 4-10: Inferential statistics for student questionnaire
Table 4-11: Inferential statistics for teacher questionnaire
Table 4-12: Inferential statistics for teacher and student questionnaires
Table 4-13: Inferential statistics for Azad University students
Table 4-14: Inferential statistics for state University students
Table 4-15: Inferential statistics for state and Azad University students
Table 4-16: Analysis of Variance for student questionnaire
Table 4-17: Repeated Measures Analysis of Variance
Table 4-18: Probabilities for Post Hoc Tests
Table 4-19: Repeated Measures Analysis of Variance
Table 4-20: Probabilities for Post Hoc Tests
Table 4-21: Repeated Measures Analysis of Variance
Table 4-22: Probabilities for Post Hoc Tests
Table 4-23: Repeated Measures Analysis of Variance
Table 4-24: Probabilities for Post Hoc Tests
Table 4-25: Repeated Measures Analysis of Variance
Table 4-26: Probabilities for Post Hoc Tests
Table 4-27: Test of means against reference constant
Table 4-28: Test of means against reference constant
Table 4-29: Test of means against reference constant
Table 4-30: Test of means against reference constant
Table 4-31: Test of means against reference constant
Table 4-32: Friedman ANOVA and Kendall coefficient of concordance
Table 4-33: Friedman ANOVA and Kendall coefficient of concordance
Table 4-34: Friedman ANOVA and Kendall coefficient of concordance
Table 4-35: Descriptive statistics for students' questionnaire
Table 4-36: Descriptive statistics of wash back on language program
Table 4-37: Reliability Results for Observation Checklist
Table 4-38: Frequency Table (age)
Table 4-39: Frequency Table (gender)
Table 4-40: Frequency Table (education)
Table 4-41: Frequency Table (record)
Table 4-42: Frequency Table (teaching)
Table 4-43: Frequency Table (interest)
Table 4-44: Frequency Table (level)
Table 4-45: Frequency Table (average)
Table 4-46: Frequency Table
Table 4-47: Frequency Table
Table 4-48: Frequency Table
Table 4-49: Frequency Table
Table 4-50: Frequency Table
Table 4-51: Frequency Table
Table 4-52: Frequency Table
Table 4-53: Test of means against reference constant
Table 4-54: Test of means against reference constant
Table 4-55: Friedman ANOVA and Kendall coefficient of concordance
Table 4-56: Friedman ANOVA and Kendall coefficient of concordance
Table 4-57: Friedman ANOVA and Kendall coefficient of concordance
Table 4-58: Friedman ANOVA and Kendall coefficient of concordance
Table 4-59: Friedman ANOVA and Kendall coefficient of concordance
Table 4-60: Friedman ANOVA and Kendall coefficient of concordance
Table 4-61: Friedman ANOVA and Kendall coefficient of concordance
Table 4-62: Friedman ANOVA and Kendall coefficient of concordance

List of Figures

CHAPTER THREE
Figure 3-1: Screen plot (pilot study)
Figure 3-2: Screen plot
Figure 3-3: A Basic Model of Wash back (Hughes, 1993)

CHAPTER FOUR
Figure 4-1: Subjects' views on test content facet
Figure 4-2: Subjects' views on test format facet
Figure 4-3: Subjects' views on test content and format facet
Figure 4-4: Subjects' views on wash back-affected facet
Figure 4-5: Subjects' views on test content and wash back
Figure 4-6: Graphic data on test content facet
Figure 4-7: Graphic data on test format facet
Figure 4-8: Graphic data on wash back facet
Figure 4-9: Graphic representation of data on test content and wash Back facet
Figure 4-10: The wash back of NDEMT on the English classes
Figure 4-11: Graphic data for teachers' methodology
Figure 4-12: Graphic data on Content of educational programs
Figure 4-13: Graphic data on students' learning strategy
Figure 4-14: Graphic data on teachers' method of evaluation
Figure 4-15: Graphic data on teachers' attitude
Figure 4-16: Graphic data on students' attitude
Figure 4-17: Graphic data on students' affective domain


of the International language in Educational Conference

1994 (pp.31-34). Hong Kong: University of Hong Kong.

Mahdavi Educational, Cultural and Charity Institute.

Boyle, J. & Falvey, P. (1994). English Language Testing in
Hong

Kong: Chinese University Press.

Bracey, G.W. (1998). Put to the test: An educator’s and
consumer’s

guide to standardized testing. Bloomington: In Phi Delta

premier in assessment literacy. Washington, DC: American
Youth Policy Forum. from http://www.cse. ucla.edu / CRESST/
files /Bracey Rep.pdf

Brown, H.D., (1994) Principles of Language Learning and
Teaching. Prentice Hall Regents, U.S.A.

Performance on Cloze Test Items? Proceedings of Third
Conference on Second Language Research in Japan.


*Language Testing*, 1, 1: 70-81.


The Scores or Ruining the Schools. Portsmouth, Nh: Heinemann.


Centrum voor onder wijsonderzoek van de in primary and secondary education. Amsterdam: Stichting Universiteit van Amsterdam.
چکیده
عنوان: بررسی مقابله ای جامع تأثیر گذاری امتحان زبان ورودی دانش‌گاه‌های دولتی و آزاد اسلامی بر نظام آموزشی زبان در ایران

امتحانات عمومی همواره به عنوان ابزار و اهداف کنترل در نظام آموزشی بوده اند. ارتباط آنها با برنامه درسی، با تربیت معلم، پایداریگری دانش آموز و با فردیت های فردی در زندگی در تمام جوامع اهمیت حیاتی دارد. بنابراین به عنوان عامل تغییر به صورت راهبردهای اصلاحی آموزشی مورد استفاده قرار گرفته است. شواهدی وجود دارد که امتحانات بر روی یادگیری و آموختن تأثیر گذاری دارند. استفاده گسترده ازمونه امتحانات برای اهداف اجتماعی و آموزشی امرره در جامعه این ویژگی تأثیر گذاری آزمون را یک چهاردهم آموزشی منحصر به فرد ساخته است.

این مطالعه بر اساس مدل اصلی پیشنهادی تاثیر گذاری آزمون که توسط هیوز (1993) ارائه شده انجام گرفته است. افراد مورد مطالعه کسانی هستند که مستقیما تحت تأثیر آزمون منحصراً زبان بوده وند دانش آموزانی که در این امتحان شرکت کرده اند. دو بررسی انجام شده و چک چک لیست مشاهده نیز برای بررسی تأثیر گذاری آزمون بر روش امتحان، ونوع آزمون استفاده شده است. هدف اصلی تحقیق در قالب ده فرضیه این بوده که آزمون منحصراً زبان بر روی داوطلبان رشته زبان و دبیران زبان انگلیسی دبیرستان ها تأثیر دارد.

افراد مورد مطالعه در مورد مطالعه امتحانات بین محتواي زبان انگلیسی دبیرستانی و محتوای آزمون منحصراً زبان موافقته ندارند.

1- افراد مورد مطالعه در مورد آموزش دبیرستان و نوع آزمون های استفاده شده در کلاسهای دبیرستان، نحوه تجزیه و تحلیل نتایج، عملکرد معلم، نگرش دانش آموز به آزمون و نگرش معلم به آزمون تأثیر ندارد.
نتایج مطالعه نشان داده است که همه فرضیه‌های صفر رد شده‌اند. همچنین نتایج در دانشگاه دولتی و آزاد اسلامی تقریبا یکسان بوده است. بعلاوه تاثیرات مثبت و منفی آزمون نیز مشخص شد و راهبردهایی برای افزایش اثرات مثبت و کاهش اثرات منفی آن پیشنهاد شده است.