

The Iranian EFL Journal September 2010 Volume 6 Issue 3

Iranian EFL Journal



**ISSN On-line: 1836-8751
ISSN Print: 1836-8743**

The Iranian EFL Journal

September 2010

Volume 6

Issue 3

Chief Editors

Dr. Paul Robertson

Dr. Rajabali Askarzadeh Torghabeh

Iranian EFL Journal



Publisher

Dr. Paul Robertson

Time Taylor International Ltd.

Senior Associate Editor:

Dr. Rajabali Askarzadeh Torghabeh
Ferdowsi University of Mashhad
Iran

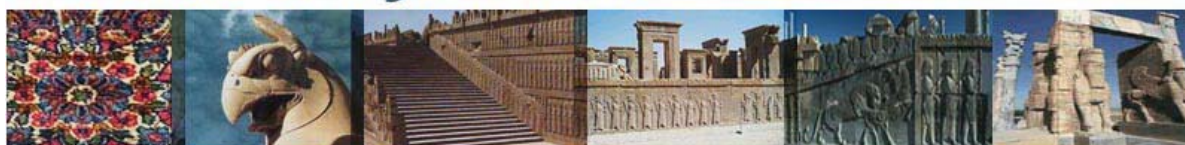
Advisors

Dr. Roger Nunn The Petroleum Institute Abu Dhabi UAE	Dr. John Adamson Shinshu Honan College Japan	Professor Dr. Z.N. Patil Central Institute of English and Foreign Languages Hyderabad, India
	Senior Statesmen	
	Professor Rod Ellis New Zealand University of Auckland New Zealand	

	Associate Editors	
Professor Dr. Dan Douglas Iowa State University USA	Dr. Reza Pishghadam Ferdowsi University of Mashhad Iran	Dr. Behzad Ghonsooly Ferdowsi University of Mashhad Iran
Prof. Dr. Rana Nayar Panjab University India	Dr. Abdolmahdy Riazi Shirza University Iran	Dr. Salmani Nodushan University of Zanjan Iran
Editorial team		
Dr. Pourya Baghahi Azad University of Mashhad Iran	Dr. Zohre Eslami Rasekh Texas A & M University USA	Dr. Azizullah Fatahi Shar-e Kord University Iran
Dr. Christopher Alexander University of Nicosia Cyprus	Dr. Parvaneh Tavakoli London Metropolitan University England	Dr. Seyyed Ayatollah Razmjou Shiraz University Iran
Dr. Shamala Paramasivam University of Putra Malaysia	Dr. Manizheh Yuhannaee University of Isfahan Iran	Dr. Antony Fenton Soka University Japan
Dr. Esmā'eel Abdollahzadeh Iran University of Science and Technology Iran	Dr. Ingrid Mosquera Gende Bettatur University College of Tourism Tarragona Spain	Dr. Abbas Zare'ee Kashan University Iran

The Iranian EFL Journal September 2010 Volume 6 Issue 3

Iranian EFL Journal



The Iranian EFL Journal Press
A Division of Time Taylor Publishers
QC Pavilion
Cebu

<http://www.Iranian-efl-journal.com>
Iranian.efljournal@yahoo.com

This E book is in copyright.

No reproduction may take place without
The express written permission of the
Iranian EFL Journal

No unauthorized copying

All rights reserved. No part of this book may be reproduced, stored
In a retrieval system or transmitted in any form or by any means
Electronic, mechanical, photocopying or otherwise, without the prior
written permission of the Iranian EFL Journal.

Chief Editor: Dr. Paul Robertson
Senior Associate Editor:
Dr. Rajabali Askarzadeh Torghabeh
Ferdowsi University of Mashhad
Iran

ISSN On-line: 1836-8751
ISSN Print: 1836-8743



Table of Contents

1. Foreword: Dr. Paul Robertson and Dr. Rajabali Askarzadeh Torghabeh	6
2. Mansoor Fahim, Parviz Maftoon and Hossein Hashemnezhad	7 - 27
Types and Amounts of Original vs. Modified Descriptions of Scaffolding in EFL Writing Ability	
3- Azar Hoseini Fatemi and Hesam Shahriari	28 - 44
To Teach or Not to Teach: On the Didactic Aspect of Accent Training in the EFL Classroom	
4- Nasser Rashidi and Mansoureh Sajjadi	45 - 69
The Effect of Time Intervals on the Keyword Mnemonic Technique Used in Teaching Vocabulary to Iranian Female EFL Learners	
5- Mohammad Reza Hashemi and Elham Yazdanmehr	70 - 85
A Process-oriented Evaluation of EFL Writing Tasks in FCE/CAE Preparation Courses versus IELTS/TOEFL Courses of Iran	
6-Masoumeh Akhondi, Faramarz Aziz Malayeri and Arshad Abd. Samad	86 - 108
Assessing Speaking: Manipulating the Facet of the Length of Prompt in an Oral Proficiency Interview Setting	
7- Amin Marzban and Mohammad Ali Ayatollahi	109 - 120
Accurate Recall Concerns in Language Classrooms: The Role of Clause Order in Memory Constraints	

The Iranian EFL Journal September 2010 Volume 6 Issue 3
Iranian EFL Journal



Foreword

Welcome to the third edition of the year 2010. The Iranian EFL Journal has had strong growth over the last few years with a monthly readership now exceeding 1,800 readers. For a new journal examining the topic of English second language acquisition from a local perspective, the growth and readership has been pleasing. Statistically, readers are coming from almost 80 countries. In the third issue of volume 6 we present 6 articles for your reading. In the first article, the authors Mansoor Fahim, Parviz Maftoon, and Hossein Hashemnezhad have investigated the effects of scaffolding perspectives on Iranian EFL sophomores' writing, and estimated the kinds and amounts of two scaffolding assessment techniques in Iranian EFL context. In the second article, Azar Hoseini Fatemi and Hesamoddin Shahriari Ahmadi have addressed the pedagogical issues on the integration of accent training into the EFL classroom. In the third article, Nasser Rashidi and Mansoureh Sajjadi have investigated the effect of time intervals on vocabulary retrieval of Iranian female EFL learners who used the keyword method to encode the information entering the long term memory. Mohammad Reza Hashemi and Elham Yazdanmehr have examined EFL writing tasks between two groups of the most popular ESOL exam preparation courses in Iran: IELTS and TOEFL versus FCE and CAE. In the next article, Masoumeh Akhondi, Faramarz Aziz Malayeri, and Arshad Abd. Samad have compared the interviewees' performances in two experimental groups receiving different lengths of the same prompt and the third group who started an impromptu interaction with the interviewers. In the last article, Amin Marzban and Mohammad Ali Ayatollahi have explored whether the nonstandard complex sentences can constrain sentence processing.

We hope you enjoy this edition and look forward to your readership.



Title

Types and Amounts of Original vs. Modified Descriptions of Scaffolding in EFL Writing Ability

Authors

Mansoor Fahim (Ph.D)

Islamic Azad University, Science and Research Branch, Tehran, Iran

Parviz Maftoon (Ph.D)

Islamic Azad University, Science and Research Branch, Tehran, Iran

Hossein Hashemnezhad (Ph.D candidate)

Islamic Azad University, Science and Research Branch, Tehran, Iran

Bio Data

Mansoor Fahim, associate professor of TEFL at Allameh Tabataba'i University, Tehran, Iran from 1981 to 2008. At present, he runs Research methods, Psycholinguistics, Applied Linguistics, Second Language Acquisition, and Seminar classes at M.A. level and First Language Acquisition, Psycholinguistics, and Discourse Analysis courses at Ph.D. level. He has published several articles and books mostly in the field of TEFL and has translated some books into Farsi.

Parviz Maftoon, associate professor of TEFL at Islamic Azad University, Science and Research Branch, Tehran, Iran. He received his Ph.D. degree from New York University in 1978 in Teaching English to Speakers of Other Languages (TESOL). His primary research interests concern EFL writing, second language acquisition, SL/FL language teaching methodology,

and language syllabus design. He has published and edited a number of research articles and books. He is currently on the editorial board of some language journals in Iran.

Hossein Hashemnezhad, Ph.D candidate of TEFL at the Islamic Azad University, Science and Research Branch, Tehran, Iran. He is a lecturer at the Islamic Azad University, Khoy Branch and currently, the Dean of the College of Humanities in the Islamic Azad University, Khoy Branch. His research interests are language teaching as well as assessment. He has published and presented papers in national and international conferences and journals.

Abstract

The purpose of the present study is twofold; first, to investigate the effects of scaffolding perspectives on Iranian EFL sophomores' writing, then to estimate the kinds and amounts of two scaffolding assessment techniques in Iranian EFL context. Regarding the original definition of scaffolding, the researchers used Lidz's twelve component behaviors of adult mediating instruction as typical example of interactions between a tutor and students. Concerning the modified definition of scaffolding, portfolio and conference as two new educational and assessment tools, were taken as the second type of scaffolding in this study. Data collection consisted of a survey adopted from four standardized checklists. Data analysis identified the dominant use of traditional assessment in Iranian EFL context. Further, EFL teachers in general use all number of scaffolding techniques for a limited amount and EFL teachers under study applied all number of scaffolding techniques for a fair amount. Finally implications were drawn for language teachers, policy makers, syllabus designer, and L2 writing instruction.

Key words: Scaffolding, Portfolio, Conference, Traditional assessment, EFL.

Introduction

Scaffolding, as stemming from Socio-Cultural Theory (SCT) and Zone of Proximal Development (ZPD), is the dialogic process by which one speaker assists another in

performing a function that he or she cannot perform alone. *Scaffolding Theory* was first introduced in the late 1950s by J. Bruner a cognitive psychologist. He used the term to describe young children's oral language acquisition. Carter and Nunan (2001, p. 226) define scaffolding as the support given to language learners to enable them to perform tasks and construct communications which are at the time beyond their capability. Van Lier (2004, p. 147) regards scaffolding as assisted performance. He adds that a scaffold on a building permits work to be conducted that would not be possible without the scaffold. But the scaffold is temporary: As soon as it is no longer needed it is dismantled. Scaffolding, as a teaching strategy, originates from Lev Vygotsky's sociocultural and his concept of the zone of proximal development. Vygotsky's sociocultural theory proposes that social interaction plays a fundamental role in the development of cognition. He theorized that learning occurs through participation in social or culturally embedded experiences. In Vygotsky's view, the learner does not learn in isolation. Instead learning is strongly influenced by social interactions, which take place in meaningful contexts. Children's social interaction with more knowledgeable or capable others and their environment significantly impacts their ways of thinking and interpreting situations. A child develops his or her intellect through internalizing concepts based on his or her own interpretation of an activity that occurs in a social setting. The communication helps the child develop inner or egocentric speech.

In its original use (Bruner, 1975, cited in Puntambekar & Hubscher, 2005), scaffolding described interactions between a parent and a child or between a tutor and a student. Puntambekar and Hubscher (2005, p. 1) state that scaffolding is no longer restricted to interactions between individuals-artifacts, resources, and environments themselves are also being used as scaffolds. Scaffolding construct is being applied more broadly, to include the support provided in technology tools, peer interactions, and discussions. With an increase in project-based and design-based environments for teaching science and math in the context of a classroom, the notion is now increasingly being used to describe the prompts and hints provided in tools to support learning. Regarding Bruner's original definition of scaffolding, the interactions between a tutor and a student, the researchers used Lidz's (1991, cited in De Guerrero & Villamil, 2000, p. 53) scale for evaluating scaffolding and measuring mediated instruction. Concerning Puntambekar and Hubscher's (2005) modified definition of scaffolding, *portfolio* and *conference* as two new educational and assessment tools, are regarded as the second type of scaffolding in this research. In order to examine

the role of scaffolding in this study, the researchers look to *portfolio* and *conference* as two scaffolding and assessment tools.

Alternative assessments as scaffolding tools

In recent years, there has been a growing interest in the application of assessment procedure using evaluation without tests which is called alternative assessment, authentic assessment, or informal assessment. Alternative assessment has been described as an alternative to traditional, standardized testing and all of the problems found with such testing. These forms of assessment are more learner-centered in that, in addition to being an assessment tool, they provide learners with a tool to be more involved in their learning, and give them a better sense of control for their own learning.

[They] also...provide teachers with useful information that can form the basis for improving their instructional plans and practices (Richards & Renandya, 2002, p. 335).

According to Birjandi, Bagheridoust, and Mossalanejad (2000), in alternative assessment we are replacing an alternative to conventional ways of monitoring students' language progress and performance. It is an ongoing process involving the student and teacher in making judgments about the students' progress in language using non-conventional strategies. In its simplest form, alternative assessment is a turn to a something new, a search for a wider palette of choices. Brown and Hudson (1998) view procedures like portfolios, conferences, diaries, self-assessments, and peer assessments not as *alternative assessments* but rather as *alternatives in assessment*.

Research in scaffolding and alternative assessment (portfolio-conferencing)

Mirzaee (2008), adopted a Vygotsky-inspired sociocultural approach, sought to explore the effect of the ZPD-sensitive discourse scaffolding on EFL learners' co-construction and microgenetic appropriation of L2 metadiscourse performing collaboration writing tasks. The participants in the study were assigned to four different treatment conditions: (i) teacher-directed formal teaching, (ii) comprehensible-input enhancement, (iii) the non-ZPD (random) L2 discourse scaffolding, and (iv) the ZPD-sensitive L2 discourse scaffolding. The findings indicated a substantial effect for the ZPD participants' microgenetic appropriation and use of English metadiscourse resources. More important, this improvement was found to statistically more significant than the other groups' development.

Nunes' (2004) article, *portfolios in the EFL classroom: disclosing an informed practice*, provides an overview of the work carried out over the period of one year with a group of 10th

grade students in a Portuguese high school. It argues that by using portfolios in EFL classrooms, the teacher cannot only diagnose the learners' skills and competences, but also become aware of their preferences, styles, dispositions, and learning strategies, thus being able to adopt a more learner-centered practice.

In another study Banfi (2003) provides an overview of work conducted with groups of advanced language students to improve their linguistic, academic, and professional skills. By means of a number of connected tasks, existing skills were put into practice and developed, while at the same time new ones were introduced. The gradual development in the course of the three portfolios, together with the regular feedback provided, allowed for a smooth process that was conducive to considerable degrees of improvement in a relatively short space of time.

Firooz Zareh (2006) conducted a research aimed at investigating the relationship between alternative assessment techniques and Iranian students' learning. To achieve this goal, a group of freshmen from Islamic Azad University were selected to participate in the study. Throughout the whole semester, the two techniques of *conferencing* and *self-assessment* were put into practice in the experimental group. The findings of the study assured the inclusion of alternative assessment techniques in assessment and instruction.

Hashemnezhad (2007) conducted a cross-sectional survey study to investigate the types of speaking assessment techniques used by Iranian English teachers at university level. The findings demonstrated two main points. On the one hand techniques which Iranian English teachers used were based basically upon traditional assessment. On the other hand although they applied traditional method of assessment, the ranking of techniques did not follow a rational one, that is, the techniques did not harmonize well with students' language proficiency level.

Purpose of the study

The purpose of the present study is to investigate the kinds and amounts of both original and modified versions of scaffolding on Iranian EFL writing ability. The potential of process-oriented writing assessment is often implied but seldom examined empirically in EFL assessment contexts in terms of multiple types of evidence such as two scaffolding techniques of portfolio and portfolio/conference and their different types of revision and feedback provided to learners. Since the importance of scaffolding in Iranian EFL settings, both original and modified versions, is simply ignored, the researchers introduced Lidz's

(1991, cited in De Guerrero and Villamil, 2000, p. 53) twelve component behaviors of adult mediating instruction as interactions between a tutor and students and two scaffolding techniques of portfolio and conference as typical examples of modified version of scaffolding. To cover the second definition, they have a glance over the dichotomy of traditional versus alternative types of assessment. It is perceived that Iranian EFL university teachers use more traditional types of assessment. This is the reason why the researchers use two alternative techniques of assessment, *portfolio* and *conference*, as two scaffolding tools and try to examine the efficacy of these two scaffolding and rescaffolding instruments on students' achievement. Based upon the afore-mentioned discussions, the following research hypotheses stand out:

- (1) Iranian EFL university teachers generally use more alternative assessment techniques than traditional ones.
- (2) Iranian EFL teachers generally use a few number of scaffolding techniques to a limited extent and specifically used considerable number of scaffolding techniques to a fair extent in this study.
- (3) Iranian EFL teachers use portfolio in their writing classes and the most frequent one is a loose-leaf notebook.
- (4) Iranian EFL teachers use oral conferencing feedback in their writing classes and the most frequent one is conferencing between teacher and individual students.

Method

Participants

Participants to fill out Checklist Number One were 20 M.A university lecturers. The sample group (n=20) was all the on hand EFL university lecturers. Participants to fill out Checklist Number Two, Three, and Four were 80 EFL sophomores at Islamic Azad University.

Instrumentations

Lidz's scale for evaluating scaffolding

Lidz (1991, cited in de Guerrero and Villamil, 2000) developed a scale for measuring mediated instruction by a rather recent application of the concept of scaffolding based on the ZPD concept and Feuerstein's work on dynamic assessment. Lidz argues that there are two types of learning: (i) direct, or learning without receiving assistance from others, and

(ii) mediated, or learning through the interactional processes between the developing human organism and an experienced, intentional partner who mediates the world to the learner by framing, selecting, focusing, and providing feedback. Lidz defines scaffolding as the mediator's adjusting the complexity and maturity of the teaching interaction to facilitate the child's mastery of the task providing support when necessary; and providing encouragement and prompts to the child to move ahead when ready. Lidz's (1991, cited in de Guerrero and Villamil, 2000) introduced twelve component behaviors of adult mediating instruction as:

1. *Intentionality*: Consciously attempting to influence the child's actions. This involves making effort to keep the interaction going, engage the child's attention, inhibit impulsive behavior, and maintain goal orientation.
2. *Meaning*: Promoting understanding by highlighting for the child what is important to notice, marking relevant differences, elaborating detail, and providing related information.
3. *Transcendence*. Helping the child make associations to related past experiences and projects him or her into the future.
4. *Joint regard*: Trying to see the activity through the child's eye; looking at an object that has been brought into focus by the child; using "we" to talk about the experience.
5. *Sharing of experiences*: Telling the child about an experience or thought that the mediator had and of which the child is not aware.
6. *Task regulation*: Manipulating the task to facilitate problem solving; stating a principle of solution or inducing strategic thinking in the child.
7. *Praise/Encouragement*: Communicating to the child, verbally or nonverbally, that he or she has done something good; keeping high the child's self-esteem.
8. *Challenge*. Maintaining the activity within the limits of the child's ZPD. This implies challenging the child to reach beyond his or her current level of functioning, but not so much that the child will feel overwhelmed and get discouraged.
9. *Psychological differentiation*: Keeping in mind that the task is the child's and not the mediator's; that the goal is for the child to have a learning experience, not the adult. Avoiding competitiveness with the child.
10. *Contingent responsiveness*: The ability to read the child's behavior and to respond appropriately. It can be compared to a well-coordinated dance between two partners who are very much in tune to one another.
11. *Affective involvement*: Expressing warmth to the child; giving the child a sense of caring and enjoyment in the task.

12. *Change*. Communicating to the child that he or she has made some change or improved in some way.

Checklists

Checklists were the second type of materials used in the research. There were four different checklists with various functions. For more information of this section see Research Procedure Phase One.

Alternative assessment tools (portfolio and portfolio/conference)

The type of portfolio which was used in the study was *growth portfolio*, which demonstrates an individual's development and growth over time. Development can be focused on academic or thinking skills, content knowledge, self-knowledge, or any area that is important in our setting. The instructors reminded the students of following Song and August's (2002) Portfolio Model. Following this model, students' portfolios included a *cover letter, four revised paragraphs with two drafts for each paragraphs, and a writing exam*.

The type of conferencing was *collaborative counseling* or *conferencing* in which the teacher worked individually with students in developing their papers. It took 3-5 minute conversations or discussions about students' completed work or about work in progress.

Research procedure

The general procedures in this study are divided into two main phases:

Phase 1 Warming Up

Preparation of instruments (checklists)

Initially, the researchers prepared the four checklists from various sources. Checklist Number One (*Kinds of assessment techniques most Iranian EFL teachers' use of traditional and alternative types of assessment*) was researchers made. From the review of literature they had, especially Farhady, Jafarpoor, and Birjandi (1995) and Gennese and Upshur (1996), the researchers developed a writing assessment checklist. Farhady et al., (1995) mentioned sixteen specific techniques within three main models of writing, readiness for writing, beginning writing, and free writing. These techniques are suitable for assessing EFL writing. These are the techniques that EFL teachers traditionally and recently use in their writing classes. Gennese and Upshur (1996), classified evaluation

without tests (alternative assessment) into six techniques of *observation, portfolio, conference, journals, questionnaires, and interviews*. The researchers combined traditional and alternative types of assessment and developed a checklist comprising twenty yes-no questions.

Checklist Number Two (*kinds and amounts of scaffolding techniques Iranian EFL teachers apply*) Was developed by Lidz's (1991, cited in De Guerrero and Villamil, 2000, p. 53) twelve component behaviors of adult mediating instruction. Lidz's scale consists of 12 component behaviors in the adult. The researchers modified this scale and developed a new scale including a checklist with 3-point Lickert scale. This newly developed scale included 12 Yes-No questions to find out if EFL teachers used any of the 12 techniques. Further, the researchers developed a 3-point Lickert scale to know how much scaffolding was used in the classroom.

Checklist Number Three (*kinds and use of portfolio as a scaffolding tool in EFL classes*) was developed by the researchers from the review of literature they had. Genesee and Upshur (1996) state the kinds and importance of portfolio assessment. The researchers prepared a checklist comprising six Yes-No questions. The first question asked whether the teacher used portfolio in the classroom. The next five items covered the types of portfolio used in the classroom.

Checklist Number Four (*Kinds and Use of Conference as a rescaffolding tool in EFL classes*) was developed by the researchers from the review of literature they had. They examined Genesee and Upshur (1996) and Montgomery and Baker, (2007) on the importance and types of conference in the classroom. They prepared a checklist comprising 13 Yes-No questions. The first question asked whether the teacher used conference in the classroom. The next 12 items covered the types of conferences used in the classroom.

Validation of four checklists

Validation of data collection instruments is an extremely important step in every research; however, it is often only briefly reported in research studies. In this study, the researchers estimated both validity and reliability of instruments. Of the five various types of validity, the researchers used *content validity*. To establish the content validity of the instruments, they went through the three stages of *EFL lecturers' opinions, the ideas of English to Persian translators, and piloting with some target population*.

In order to obtain experts' opinions on the relevance of items to the purpose of the checklists, possible wording, and interpretation problems and the instructions, all the

research four checklists were initially given to four EFL experts. To ease the students' understanding, checklists Numbers Two, Three, and Four were translated into students' official language, Persian. To obtain experts' opinions on the relevance of items to the purpose of the checklists, and possible interpretation problems, the four Farsi drafts of checklists were given to two English to Persian translators as well as university teachers.

At the same time, the research four checklists were also piloted with forty members of the target population. After piloting, wording and conceptual problems were discussed and additional ideas were invited. As a result of the content validity check, some changes were implemented, of which the most important ones were eliminating irrelevant items, collapsing related statements, and addressing a number of wording problems.

Regarding the reliability of checklists, the researchers selected *internal consistency reliability*, from the three recommended methods of parallel form, internal consistency, and test-retest for estimating the reliability of checklists. In order to check the internal consistency of the four checklists, the researchers used *Cronbach's alpha* (Cronbach alpha coefficient). The coefficients obtained were as follows:

Checklist Number One, Part One = 0.730 highly accepted

Checklist Number One, Part Two = 0.889 highly accepted

Checklist Number Two = 0.532 moderately accepted

Checklist Number Three = 0.949 highly accepted

Checklist Number Four = 0.902 highly accepted

- *Introducing the textbook, classroom requirements, the concepts of portfolio and conference*

The researchers introduced *College Writing: From Paragraph to Essay* (Zemach & Rumisek, 2003) as writing textbook. They emphasized the importance of process approach in writing in which students' improvements was continuously assessed in every group. Then, they introduced two scaffolding concepts of portfolio and conference as two key elements in process writing. The researchers reminded the students of following Song and August's (2002) portfolio model. Following this model, students' portfolios included a cover letter, four revised paragraphs with two drafts for each paragraph, and a writing exam. The following steps were applied in portfolio and portfolio/conference session.

1. At their orientation, new students in the two classes of portfolio and portfolio/conference were introduced to the concept of the portfolio and

portfolio/conference and were given a copy of the document and enough explanation and discussion.

2. The next session, students attended another meeting where they developed a plan for working on the portfolio and portfolio/conference. The teacher and students discussed the details and inspected portfolio samples. The teacher introduced a "portfolio structure" which was a modified version of portfolio structure prepared by Lynch and Shaw (2005, pp. 295-296). According to this structure, students could browse to gain insights into this portfolio process

Phase 2: Checklist distribution

Distributing checklist Number One

Checklist One (*EFL teachers' use of traditional and alternative assessment*) was prepared and distributed to 20 EFL university teachers *prior* to research in order to find the kinds of assessment techniques EFL teachers generally use (traditional vs. alternative types of assessment).

Distributing checklist Number Two

Checklist Two (kinds and use of scaffolding techniques Iranian EFL teachers apply) was distributed to students both prior to and after the study. This checklist was distributed to 80 EFL students. Students were asked to answer to 12 Yes-No questions in order to ensure whether or not their teachers had applied such techniques during previous writing classes. The purpose of the modified checklist (distributed at the end of the study) was twofold: First, to ask whether their teachers had used any of the mentioned techniques during the term, and second, to what extent had the teachers used these techniques?

Distributing checklists Number Three and Four

After final exam (post test), students were approached and asked if they would be willing to fill out two short checklists (Checklists Number 4 and 5). It was assumed that after final exams would be an ideal time for giving the survey because the students would not know their final grade or worry that rating their teachers would affect their grade in the course. Both checklists were distributed to 80 EFL students. Checklist Number Three (*kinds and use of portfolio as a scaffolding tool in EFL classes*) comprised six Yes-No questions.

The first question asked whether the teacher used portfolio in the classroom. The next five items covered the types of portfolio used in the classroom. Checklist Number Four (kinds

and use of conference as a rescaffolding tool in EFL classes) comprised 13 yes-no questions. The first question asked whether the teacher used conference in the classroom. The next 12 items covered the types of conferences used in the classroom.

Results

As it was previously stated, this study was an attempt to find evidence to accept or reject the four hypotheses presented earlier. This section investigates the hypotheses empirically one by one and reports the findings.

Hypothesis 1: Iranian EFL teachers use' of alternative assessment techniques

Figure 1 below illustrates that the means of traditional and alternative assessment techniques in EFL writing are different. The mean for traditional is 28.0 and for alternative is 7.5. This indicates that in general Iranian EFL teachers use more traditional assessment techniques than alternative ones.

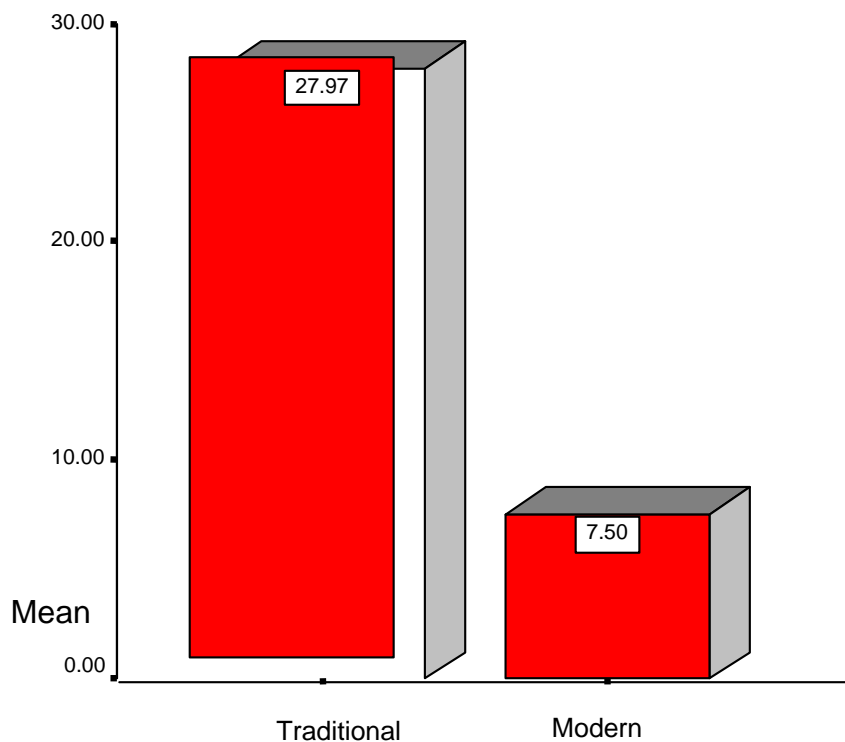


Figure 1 Graphical representation of Iranian EFL teachers' use of traditional vs. alternative techniques in EFL writing

In order to compare the degree of significance between traditional and alternative assessment techniques in writing, a Paired Samples Test was applied to the result. Table 1 shows that there are significant differences between Iranian EFL teachers' use of

traditional vs. alternative assessment techniques in writing (t-observed $t = 10.365$, $p = .000 < .005$).

Table 1 Results of paired samples test for EFL teachers' use of traditional vs. alternative assessment techniques in writing.

	Paired Differences					
	Mean	Std. Deviation	Std. Error Mean	t	df	Sig. (2-tailed)
Pair 1 Traditional- Modern	20.4688	12.49009	1.97486	0.365	39	.000

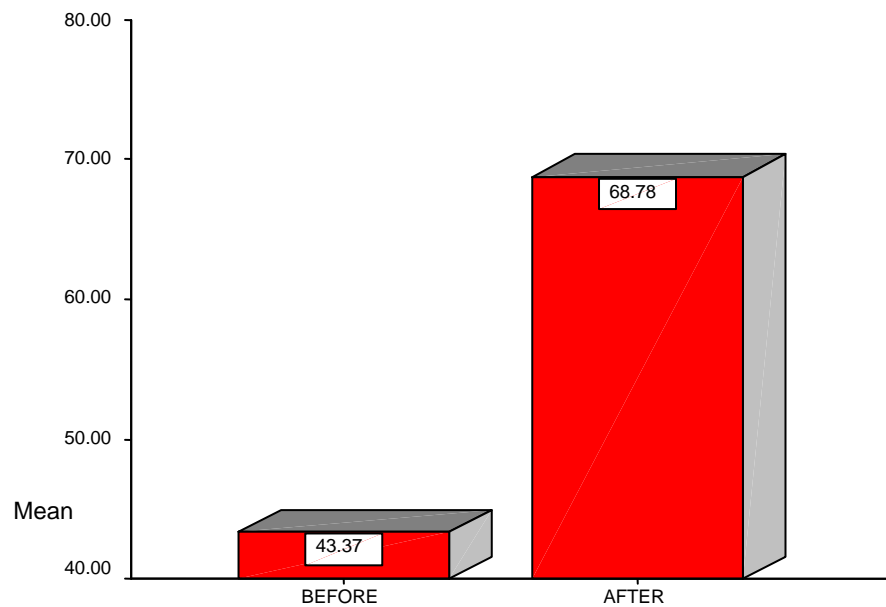


Figure 2 Graphical representation of Iranian EFL teachers' use of scaffolding techniques before and after treatment

Hypothesis 2: *Iranian EFL teachers' use of scaffolding techniques in general and specifically in this study*

The researchers distributed the present checklist both prior to and after the study. Figure 2 below illustrates that the means of scaffolding before and after study is different. The total mean for the use of scaffolding, both before and after the mentioned techniques, are 43.37 and 68.78.

Table 2 below represents the mean for the use of every 12 scaffolding techniques both before and after the present study. It indicates that EFL teachers in general and

specifically in this study used all kinds of scaffolding techniques but the amount of scaffolding was increased considerably during the study. The total mean for 12 scaffolding techniques was increased from 43.3681 (before study) to 68.7847 (after study). The following table shows a significant increase in every 12 scaffolding techniques after the study.

Table 2 Paired sample statistics: The Mean for the use of every 12 scaffolding techniques both before and after the study

				Mean	N	Std. Deviation	td. Error M.
Pair				43.3681		10.51953	1.17612
Total				68.7847		9.30508	1.04034
					80		
1. Intentionalit	Pair	Before	QB1	2.0750	80	.85351	.09543
	1	After	QA1	3.1875		.92905	.10387
2. Meaning	Pair	Before	QB2	2.5375	80	1.44908	.16201
	2	After	QA2	3.3875		.77122	.08622
3. Transcendence	Pair	Before	QB3	1.7875	80	.72380	.08092
	3	After	QA3	3.1250		.91920	.10277
4. Joint Regard	Pair	Before	QB4	2.0125	80	.78746	.08804
	4	After	QA4	2.7500		.94802	.10599
5. Sharing of Experience	Pair	Before	QB5	2.1125	80	.87140	.09743
	5	After	QA5	2.9875		.90699	.10140
6. Task Regulation	Pair	Before	QB6	1.8500	80	.79715	.08912
	6	After	QA6	3.1750		.86822	.09707
7. Praise	Pair	Before	QB7	2.4000	80	.97565	.10908
	7	After	QA7	3.2875		.88866	.09936
8. Challenge	Pair	Before	QB8	2.7375	80	.97752	.10929
	8	After	QA8	3.2000		.84793	.09480
9. Psychological Differentiation	Pair	Before	QB9	1.9250	80	.86822	.09707
	9	After	QA9	2.5125		.95459	.10673
10. Contingent Responsively	Pair	Before	QB10	2.7250	80	1.00599	.11247
	10	After	QA10	2.8625		.96448	.10783
11. Affective Involvement	Pair	Before	QB11	2.7875	80	.95060	.10628
	11	After	QA11	3.3625		.84560	.09454
12. Change	Pair	Before	QB12	2.6625	80	1.04268	.11658
	12	After	QA12	2.9250		1.00347	.11219

In order to show the degree of significance between two groups of study, Paired Sample Statistics was calculated. The result of statistics in Table 3 below reveals that totally there are significant differences between Iranian EFL teachers' use of scaffolding techniques

before and after treatment (t-observed, $t = -18.958$, $p = .000 < .05$), except questions Numbers 10 (Contingent responsively) and 12 (Change).

Table 3 Paired sample tests

		Paired	Differences			
		Mean	Std. Deviation	t	df	g. (2- tailed)
Pair Total	fore-After	-25.4167	11.99133	-18.958	79	.000
Pair 1	QB1 QA1	-1.1125	1.21169	-8.212	79	.000
Pair 2	QB2 QA2	-.8500	1.53524	-4.952	79	.000
Pair 3	QB3 QA3	-1.3375	1.211117	-9.877	79	.000
Pair 4	QB4 QA4	-.7375	1.18795	-5.553	79	.000
Pair 5	QB5 QA5	-.8750	1.22604	-6.383	79	.000
Pair 6	QB6 QA6	-1.3250	1.11122	-10.665	79	.000
Pair 7	QB7 QA7	-.8875	1.26285	-6.286	79	.000
Pair 8	QB8 QA8	-.4625	1.13566	-3.643	79	.000
Pair 9	QB9 QA9	-.5875	1.16591	-4.507	79	.000
Pair 10	QB10 QA10	-.1375	1.29989	-.946	79	.347
Pair 11	QB11 QA11	-.5750	1.27065	-4.047	79	.000
Pair 12	QB12 QA12	-.2625	1.43856	-1.632	79	.107

Note:

Pair 1: Intentionality; Pair 2: Meaning; Pair 3: Transcendence; Pair 4: Joint regard; Pair 5: Sharing of experience; Pair 6: Task regulation; Pair 7: Praise; Pair 8: Challenge; Pair 9: Psychological differentiation; Pair 10: Contingent responsively; Pair 11: Affective involvement; Pair 12: Change

Hypothesis 3: Use of portfolio as a scaffolding tool in EFL writing classes

Figure 3 shows that Iranian EFL teachers used portfolio in this study and the most frequent type is Question Number Three (a loose-leaf notebook).

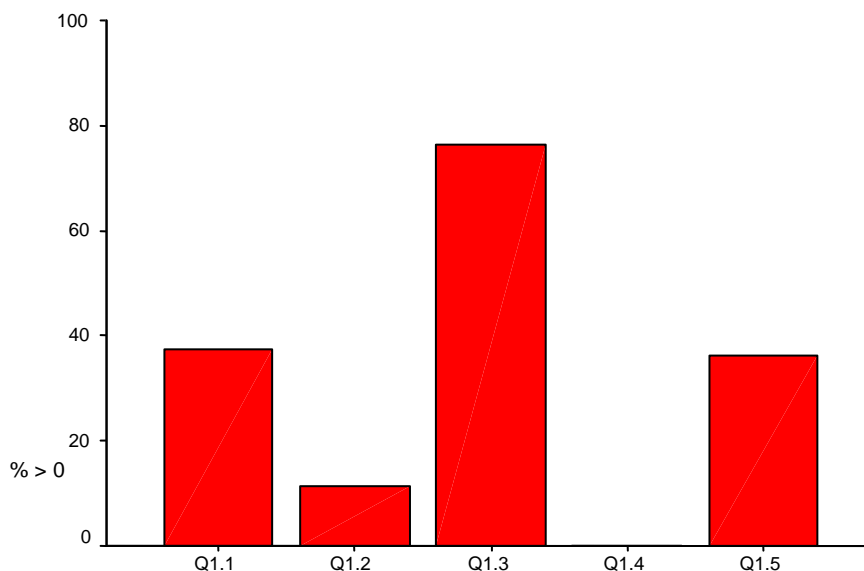


Figure 3 The most frequent types of portfolio in the study

Question Number 1: Traditional writing folder

Question Number 2: Bound notebook with separate sections

Question Number 3: Loose-leaf notebook

Question Number 4: Combination folder and a big brown envelop

Question Number 5: a notebook divided into two sections

In order to find out whether the differences between different types of portfolio are statistically significant or not, a Cochran Test was applied to the result. The result of test shows that there are significant differences among five different kinds of portfolio (Cochran Q is 121.027 and $P = .000 < 0.001$).

Hypothesis 4: Use of conference as a scaffolding tool in EFL writing classes

Figure 4 below shows that Iranian EFL teachers used conference in this study and the most frequent types are Questions Numbers One, Eight, and Five.

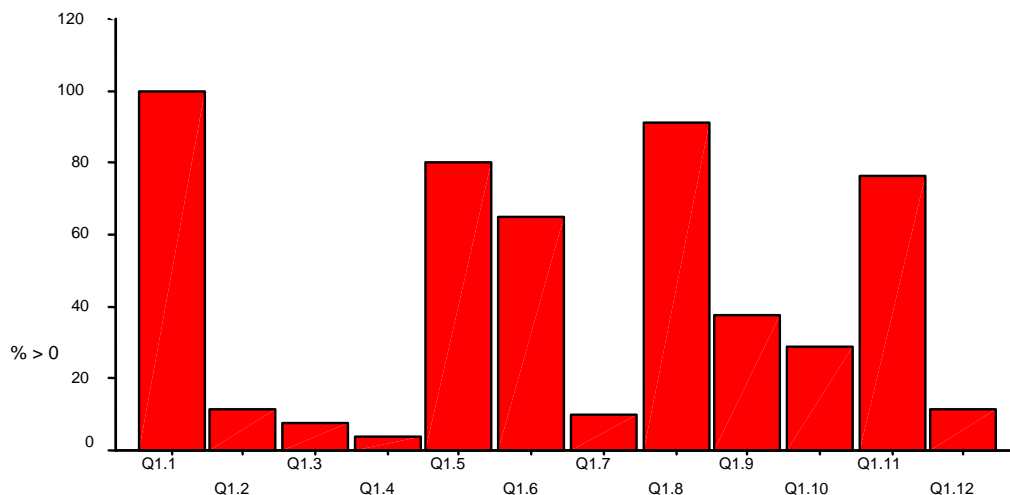


Figure 4 The most frequent types of conference in the study

Question Number 1: Conversation between teacher and individual students

Question Number 8: Teacher uses conferencing for grammar as one of the students' local issue

Question Number 5: conference is conversation about work in progress

In order to find out whether the differences between different types of conferences are statistically significant or not, a Cochran Test was applied to the result. The result of Cochran test indicates that there are not significant differences among twelve different kinds of conferences (Cochran Q is 459.391 and $P = .000 < 0.001$).

Discussion

As mentioned earlier, there were four hypotheses in this study. Of the four hypotheses formulated at the beginning of this study and investigated empirically, hypothesis number one is rejected and hypotheses numbers two, three, and four, are confirmed.

The first hypothesis relates to checklist Number One. As mentioned earlier, the checklist had two major parts. The first part included sixteen techniques which EFL teachers generally and commonly use in writing classes. The second part included six techniques which evaluated alternative type of writing. The analysis of the first part revealed that Iranian EFL writing teachers generally use traditional types of assessment more than alternative ones.

The second hypothesis covered checklist Numbers Two, the kinds and use of scaffolding techniques Iranian EFL teachers apply. This checklist consists of 12 component behaviors in the adult with 3-point Lickert scale. The researchers distributed

the checklist both prior to and after the study. Generally speaking, Iranian EFL writing teachers generally use scaffolding techniques for a limited amount but during the experiments EFL writing teachers used all types of scaffolding techniques for a fair extent.

Regarding the use of portfolio in this study, EFL writing teachers used portfolio as a scaffolding tool and the most frequent type of portfolio was “loose-leaf portfolio”. Concerning the use of conference, EFL teachers used conference in this study and the most frequent type is “conversation between teacher and individual students”.

Conclusion

The teaching and assessing of writing has thus become a specialized component of English language teaching, one that has come to occupy a prominent place in research and pedagogy due in part to the ever-growing number of students enrolled in higher-education institutions in English speaking countries and also to the recognition of changes in global realities. In fact, the field of second language writing is an area affecting the lives of hundreds of thousands at institutions around the world where they must submit high quality written work in a language they did not learn as native speakers. Equally important, the ability to speak and write a second language is becoming widely recognized as an essential tool or an important skill for people of all works of life in today’s global community for various reasons (Weigle, 2002). The findings of the current study may have important implications, both theoretical and pedagogical, for L2 writing instruction.

Theoretical implications

The findings of the study provided support for EFL learners’ writing ability as a typical example of productive ability. The teaching and assessing of writing has thus become a specialized component of English language teaching, one that has come to occupy a prominent place in research and pedagogy due in part to the ever-growing number of students enrolled in higher-education institutions in English speaking countries and also to the recognition of changes in global realities.

Pedagogical implications

Reform of education is not simply reform of school system but reform of the behavior and thinking of the wider social teaching-learning process that guides moral-political

ideas and behavior. Far-reaching curriculum innovation involves fundamental shifts in the values and beliefs of the individuals concerned (Burns, 1996). Decisions related to EFL writing will affect teachers, policy makers and syllabus designers and finally EFL writing instruction.

Implications for teachers

Teachers should introduce newly introduced educational tools, such as portfolio and conferencing to students. These tools may act as scaffolding and motivate the learners to study and learn well. Self-evaluation can be encouraged in student portfolios, self-review checklists, and teacher and peer responses. In addition, verbalizing the writing process step-by-step can be effective, as it affords both students and teachers the opportunity to consider writing dialogically.

Following Bruner's (1975) original definitions of scaffolding and Puntambekar and Hubscher's (2005) modified definition of scaffolding, the researchers used the two types of scaffolding as: Lidz's (1991) scale (twelve component behaviors of adult mediating instruction), and alternative assessment (portfolio and portfolio/conference).

Regarding the use of alternative assessment, especially portfolio, this scaffolding assessment tool gives students a sense of greater control over their writing outcomes, and appears to contribute to increase their final grades. On the other hand, one of the most positive aspects of the portfolio-based method of evaluation is its movement away from the teacher-as-examiner/grade giver, and its emphasis on the teacher-as-coach/facilitator/mentor.

Implications for policy makers and syllabus designers

Viewing writing as a process-oriented activity helps the syllabus designers and material developers write books which encourage students to engage in brainstorming activities, outlining, drafting, rewriting, and editing. This is congruence with what Vygotsky (1978) argues that people learn by doing. Thus it is important to give learners many opportunities to do and learn from their writing.

Implications for L2 writing instruction

EFL students are often anxious about writing and need to be encouraged to see it as a means of learning, rather than demonstrating learning. Instead of considering writing as a goal of language instruction, it would be better to focus on it as a means of

developing language competence in such a way that the emphasis shifts from learning to write and moves in the direction of writing to learn. Teacher-students interactions through feedback, scaffolding, and conferencing are a kind of collaborative group work that may lead to greater opportunities for students to negotiate meaning as they work with peers in improving a written text. In pedagogical practice, viewing writing as a process-oriented activity encourages students to engage in multiple drafting and consider writing as occurring in stages that may differ to some extent among different writers.

References

- Banfi, C. S. (2003). Portfolios: Integrating advanced language, academic, and professional skills. *ELT Journal*, 57(1), 34-42.
- Birjandi, P., Bagheridoust, E., & Mossalanejad, P. (2000). *Language testing: A concise collection for graduate applicants*. Tehran: Mahdavi Publication.
- Brown, J. D., & Hudson, T. (1998). The alternatives in language assessment. *TESOL Quarterly*, 32(4), 653-675.
- Burns, A. (1996). Collaborative research and curriculum change in the Australian migrant English program. *TESOL Quarterly*, 30 (3), 591-598.
- Carter, R., & Nunan, D. (Eds.) (2001). *Teaching English to speakers of other languages*. Cambridge: Cambridge University Press.
- De Guerrero, M. C. M., & Villamil, O. S. (2000). Activating the ZPD: Mutual scaffolding in L2 peer revision. *The Modern Language Journal*, 84(1), 51-68.
- Farhady, H., Jafarpoor, A., & Birjandi, P. (1995). *Testing language skills from theory to practice*. Tehran: SAMT Publications.
- Firooz Zareh, A. (2006). *The effectiveness of alternative assessment and traditional methods of testing on Iranian EFL adult learners' reading proficiency*. Unpublished master's thesis. Allameh Tabataba'i university, Tehran, Iran.
- Genesee, F., & Upshur, J. A. (1996). *Classroom-based evaluation in second language education*. Cambridge: Cambridge University Press.
- Hashemnezhad, H. (2007). *Speaking assessment techniques used by Iranian EFL teachers at university level*. Paper presented at The Fourth Conference on issues in English Language Teaching (IELTI 4), Tehran, Iran.
- Lynch, B., & Shaw, P. (2005). Portfolios, power, and ethics. *TESOL Quarterly*, 39 (2), 263-295.

- Mirzaee A. (2008). *On the effect of the ZPD-sensitive dialogic discourse scaffolding on the microgenetic appropriation of metadiscourse in task-based EFL writing*. Unpublished doctoral dissertation. Allameh Tabataba'i university, Tehran, Iran.
- Montgomery, J. L., & Baker, W. (2007). Teacher-written feedback: Student perceptions, teacher self-assessment, and actual teacher performance. *Journal of Second Language Writing, 16* (2), 82-99.
- Nunes, A. (2004). Portfolios in the ESL classroom: Disclosing an informed practice. *ELT Journal, 58*(4), 327-335.
- Puntambekar, S., & Hubscher, R. (2005). Tools for scaffolding students in a complex learning environment: What have we gained and what have we missed? *Educational Psychologist, 40*(1), 1-12.
- Richards, J. C., & Renandya, W. A. (Eds.) (2002). *Methodology in language teaching: An Anthology of current practice*. Cambridge: Cambridge University Press.
- Song, B., & August, B. (2002). Using portfolio to assess the writing of ESL students: A powerful alternative? *Journal of Second Language Writing, 11*(1), 49-72.
- Van Lier, L. (2004). *The ecology and semiotics of language learning: A sociocultural perspective*. Massachusetts: KAP.
- Vygotsky, L. S. (1978). *Mind in Society: The Development of higher psychological processes*. Cambridge: Harvard University Press.
- Weigle, S. C. (2002). *Assessing writing*. Cambridge: Cambridge University Press.
- Zemach, D. E., & Rumisek, L. A. (2003). *College writing: From paragraph to essay*. Oxford: Macmillan Publications.

Title

To Teach or Not to Teach: On the Didactic Aspect of Accent Training in the EFL Classroom

Authors

Azar Hoseini Fatemi (Ph.D)

Ferdowsi University of Mashhad, Iran

Hesamoddin Shahriari Ahmadi (Ph.D. Candidate)

Ferdowsi University of Mashhad, Iran

Bio Data

Azar Hoseini Fatemi, assistant professor, Department of English, Ferdowsi University of Mashhad, Iran. She is currently teaching EFL courses at Ferdowsi University of Mashhad. Her areas of interest include EFL/ESL courses.

Hesamoddin Shahriari Ahmadi Ph.D candidate in TEFL at Ferdowsi University of Mashhad, Iran. He is currently teaching English courses at Ferdowsi University of Mashhad. His areas of interest include EFL/ESL teaching.

Abstract

This study addresses the pedagogical issues on the integration of accent training into the EFL classroom. Investigations into the area of teaching pronunciation have downplayed the role of accent training in favor of functional intelligibility. Attempts to teach native-like accents to EFL learners have been abandoned mainly due to the perceived impossibility of such feats and also the occasionally-reported advantage of local-accent comprehensibility (Wilcox, 1978; Ekong, 1982). This shift in focus has also

been fueled by the emergence of 'English as an International Language'. The present study aims at discovering the Iranian EFL learners' attitudes towards various English accents. To this end, 112 participants from three proficiency levels participated in this study. An attitudinal survey using bipolar adjectives was used for eliciting the data. The test of Analysis of Variance (ANOVA) was used to analyze the data. By showing that learners' beliefs and attitudes may run contrary to those of policy makers, the present study argues that in deciding on what accent to teach in the classroom, apart from learners' needs, we should also take their wants into account. This implies that instead of imposing our views upon our learners, maybe we ought to rethink our approach towards accent training in the English classroom, and opt for a more liberal stance.

Key words: Accent training, Needs analysis, English as an international language, learner attitude.

Introduction

Owing to the fact that accent is one of the first noticeable features of oral communication, for the foreign language teacher, deciding on what accent to teach in the classroom requires very careful consideration. Teaching a native-like accent, on the one hand, can benefit the learners by equipping them with the covert prestige varieties of the English language. On the other hand, it can deprive them from the greater intelligibility local accents have been said to offer (Wilcox, 1978; Ekong, 1982). The methods have also been known to provide different advice on what accent to teach in the classroom. At the end of the day, EFL instructors either choose to teach English with a native-like accent (i.e. American, British) or decide to train their students to use the language with a local accent. Prior to attempting to look into the pros and cons of each approach, we will begin our discussion by trying to understand the nature of accent.

What is accent?

People generally have a 'common sense' view of accent. Before considering a technical description of accent from the point of view of phoneticians, sociolinguists and psycholinguists, a general description of the term is attempted. The Merriam-

Webster online dictionary describes accent as a “way of speaking typical of a particular group of people and especially of the natives and residents of a region”. From this definition, it can be understood that an accent is a feature or manifestation of speech, which differs depending on the community or region to which the speaker belongs. This feature can, therefore, be regarded as a factor in determining the similarities and differences in speech between speakers, depending on whether they share the same language. For a more detailed definition of accent, it is useful to look at definitions by specialists of various fields, who have also defined accent in terms of their own area of specialty. Phoneticians define accent as a specific pronunciation, determined by the phonetic habits of the speakers’ native language transferred to his or her use of another language (O’Grady et al, 2005). The phonetic view of accent chiefly focuses on the ability and/or inability to produce certain sounds due to their absence or presence in the native language inventory of a speaker.

Sociolinguists adopt a more embracing approach towards the study of accent. Becker (1995), for example, defines accent as part of an individual’s language which serves to specify the speaker’s region of origin or national/ethnic identity regardless of the language being spoken. Contrary to the realm of phonetics and phonology which is quite narrow in its scope of examining accent, sociolinguistics investigates accent in terms of its phonetic, lexical and grammatical variations in diverse social contexts. As clear from the above definition, accent, from the sociolinguistic perspective, serves as a means for identifying people’s attachment to a speech community.

In the context of sociolinguistics, it is of utmost importance to specifically distinguish between accent and dialect. The former describes where a speaker is from regionally or socially (Yule, 1985). While, the latter implies systematic differences in the way language is spoken by different groups.

Accent has also been of great interest to psycholinguists. This importance stems from the perceptions which originate from different accents, both within the speaker and the hearer. That is, people speaking with a particular accent, known to be prestigious, will be recognized as being more attractive both by themselves and by those around them. These stereotypes shaping our attitudes towards accents and their speakers can be the outcome of experience, rumors, books, and most importantly, the media (Alford and Strother, 1990). According to Dixon and Mahoney (2004), speakers with standard accents are regarded to be more intelligent, proficient and fluent than those speaking with a nonstandard accent.

Accent and second language acquisition

Psycholinguists and phonologists who specialize in accent generally hold that the difficulty of learning to pronounce a foreign language is cognitive rather than physical, and that it deals with the manner in which raw sounds are categorized or conceptualized in speech (Fraser, 2000). Accent is also often used to refer to the speech of a person who speaks a language non-natively (Fromkin, Rodman and Hyams, 2003). This means that both native and non-native speakers of a language could be said to have accents. The field of second language acquisition concerns itself with this sense of the term. Southwood and Fledge (1999) define a foreign accent as: “Non-pathological speech produced by second language learners that differs in partially systematic ways from the speech characteristics of native speakers of a given dialect” (p. 335).

As in the case of native accents, foreign accents also result in different attitudes on the part of the speaker and hearer. Most of the studies carried out on perceptions towards accented speech by foreign language speakers have included native speaker subjects and were performed in ESL settings. One of the earliest studies in this area was by Brennan and Brennan (1981) who investigated the attitudes of Anglo-American and Mexican-American speakers towards accented speech. The findings of this study revealed that speakers with lower degrees of accented speech were believed to be of higher social status than those with stronger foreign accents. Other studies, such as that carried out by Ryan and Carranza (1977), showed that there is a high correlation between respondents’ ratings of accentedness and attribution of status and even possible occupation.

In another study on hearer attitudes towards non-native speech, Johnson and Frederick (1994) demonstrated that pronunciation errors resulted in more negative judgments in comparison with grammatical errors, which are more critical to understanding any given speech sample. Stressing the significance of accent, Cargile and Giles (1997) found that pronunciation was a more determining factor in shaping attitudes towards speech than its content. That is, a foreign accent was shown to generate negative attitudes regardless of the tone of the message. Another group of studies have concerned themselves with non-native speakers’ attitudes towards accented English. Manzano (1997) examined the listening comprehension ability and attitudes of Puerto Rican university students with regards to lectures delivered by university professors with varying degrees of English-accented speech. The findings point out that all accents were regarded to be equally

comprehensible. However, the Standard American accent was shown to have the highest ratings with respect to attractiveness, dynamism and superiority. Other studies of similar nature (Chiba, Matsuura and Yamamoto, 1995; Forde, 1995) have also revealed that advanced EFL learners display negative attitudes towards their own regional accents of English. EFL learners also exhibited a marked preference for accents they were more familiar with. The Standard American English was chosen more often due to its cultural influence (music, television, film, etc.).

Considering all that has been said, one is enticed to conclude that language instructors should try to teach English with a native accent (e.g., Standard American or Standard British accent), due to the greater levels of prestige and attractiveness attributed to such accents both by native and non-native speakers of English. Nevertheless, the emergence of English as an international language has made many specialists and language teachers think twice.

Teaching English as an international language

The fast-growing and inevitable globalization of our world has its own consequences which are reflected upon our daily lives. The field of ELT is also significantly influenced by this revolution, and it is important for teachers of language to be aware of these changes and innovations. English as an international language has downplayed the importance of speaking with a native-like accent in the classroom in a number of ways.

First of all, in our modern world, with the increase in the scientific exchange of information, the rapid rise in the rate of global tourism and the perpetual growth in international trade and commerce, English language is no longer used for communicating with native speakers (Warschauer, 2000). The English language is now commonly believed to be shared with a mixed group of non-native speakers. As a result, a variety of possible interactions have been made possible using the English language, not all of which include a native speaker as an interlocutor. Hence, it would seem illogical to have learners take up American or British English when they may only want to communicate with non-native speakers and not necessarily native ones.

Globalization has also given rise to awareness towards linguistic imperialism. This has resulted in an uprising against what is believed to be linguistic hegemony. English as an international language is seen as preserving local identities and values. This movement has had a tremendous influence on taking away all the value assigned to native accents

and sharing it with local varieties. It has also rectified the classical division of English speakers into native-, second language-, and foreign language speakers (Jenkins, 2005).

The importance of needs analysis

Since Munby (1978) first introduced needs analysis to language teaching, much attention has been directed towards learners' needs. The communicative approach advocated the development of a syllabus based on learner needs and the analysis of the target situation. Needs analysis later went on to become one of the vital and most determining parts of English for specific purposes (Dudley-Evans and St. John, 1998).

Needs have been defined and classified into different categories. In other words, contrary to the time when it was first introduced, needs are no longer viewed as a unitary term. A multitude of classifications have been proposed. Brindley (1989) have divided needs into a subjective and objective category. Berwick (1989) drew a distinction between 'perceived' and 'felt' needs. Hutchinson and Waters (1987) also categorized needs as 'necessities', 'lacks' and 'wants'. Therefore, it would no longer suffice for language planners and syllabus designers to regard viewing the target situation as a form of comprehensive needs analysis. This statement holds true for the decision of which accent to aim for in the classroom.

In Hutchinson and Water's (1987) classification of needs, 'wants' refer to subjective needs, as perceived by the learners themselves. It is of great importance for language teachers to not prioritize and impose their own understanding of needs upon their learners and the course. Although it is possible for learners' wants to be in conflict with the views of other interested parties. Nevertheless, due to the close relationship between wants and the learners' level of motivation, it is of utmost importance to include this crucial aspect into our process of needs analysis.

As can be seen, when deciding which accent to teach in class, the teacher is faced with a dilemma. On the one hand, native accents seem to be more prestigious and attractive and consequently, they can possibly result in better job opportunities, improved social life and even higher academic achievement. On the other hand, local accents preserve the speakers' national and individual identity and could also be seen as effort which could be spent on other achievements such as the improvement of comprehensibility and communicative competence. One possible solution to this problem could be to find out about the learners' own attitudes and beliefs towards various accents. That is, we

should take the students' 'wants' into account. The present study aims to find out about Iranian EFL learners' attitudes towards various accents of the English language.

Method

Participants

The participants of this study were 112 (67 F, 45M) learners of English as a foreign language in the city of Mashhad, Iran. The students were studying at private language institutes in this city. Age was controlled by including only those participants who were between 18 to 30 years of age. It is worth mentioning that the first language of all learners was Farsi. All participants were initially asked to take part in a specimen of the paper-based version of the Test of English as a Foreign Language (TOEFL). Based on the results of this test, participants were divided into three proficiency groups: Beginners, Intermediate and Advanced students (B, I, A). Participants with a score of over 1 standard deviation above the mean were considered to be advanced. Those with a score of between -1 and +1 standard deviations from the mean were classified as intermediate, and those with a score of below -1 standard deviation were regarded as beginners. The beginner, intermediate and advanced groups consisted of 25, 68 and 19 participants, respectively. The descriptive statistics for the scores obtained on the pretest can be seen in figure 1.

Table 1. Descriptive values for the TOEFL test

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
TOEFL	112	79.00	135.00	106.22	15.72
Valid N (listwise)	112				

Instruments

The stimulus for this study was provided by three different male speakers. The three speakers each had a different accent: The standard North American, standard British, and the local Farsi accent of English. These accents were selected on the basis of their prevalence in language institutes and availability of teachers and textbooks in Iran. All

speakers read out the same neutral text on the topic of glaciers. The passage was selected on the basis of neutrality with the intention of preventing learners' responses from being influenced by the topic, and, hence, being biased. The text was 52 words long, and each reading took approximately one minute to complete. This was believed to be a suitable length, since a shorter passage would not have provided sufficient stimulus for the participants' judgments, and a longer one would have jeopardized the involvement of the listener. None of the readings included hesitation or rephrasing. The speech samples were randomly organized and copied into an audio disc.

The attitudinal survey used in this study consisted of two parts. The first part used a Likert Scale with 5 bipolar adjectives to discover the learners' attitudes towards the stimulus providers and their accents. These adjectives were: unpleasant/pleasant, unfriendly/friendly, not prestigious/ prestigious, uneducated/educated and comprehensible/incomprehensible. The respondents were asked to choose a point on the scale between 1 to 5 for each of the bipolar adjectives. For instance, on the pleasant vs. unpleasant item, a participant who had chosen 4 considered the speaker to be more pleasant than that the one which had chosen 1. The highest possible score for each speaker on this part was 25 and the minimum was 5. In the second section of the survey, participants were provided with three statements about their perceptions regarding pronunciation and accent in foreign language learning. In this part, they were required to express their level of agreement with what was stated on a Likert Scale, with 1 expressing their disagreement, and 5 articulating their agreement with the statement. The maximum achievable score on this part was 15 and the minimum was 3. The statements in the second section of the survey were as follows:

- 1. I believe that it is important for me to learn to speak English with a native accent*
- 2. I believe that achieving a native accent in English is achievable for me.*
- 3. I believe my teacher should spend more time teaching me how to speak with a native accent rather than focusing on grammar and vocabulary.*

Procedure

The participants of this study were carefully briefed by the researchers on what was expected from them. The survey was administered on multiple occasions and in separate venues. However, great care was taken to ensure equal conditions on each administration. The completion of the survey lasted approximately 15 minutes. On

every administration, following each recording, the participants were given time to complete the five attitudinal items for that speaker. Having completed this procedure for all four recordings, the participants were asked to state their level of agreement/disagreement with each of the three statements in the second part.

Results

The test of analysis of variance (ANOVA) was used to see if there were significant differences between the three ability groups (i.e., B, I, A) with regards to their attitudes. The findings of the first section of the survey have been summarized in Tables 2 and 3 below.

Table 2. ANOVA: Attitude by level of proficiency

		Sum of Squares	df	Mean Square	F	Sig.
AMERICAN	Between Groups	769.75	2	384.87	14.46	.00
	Within Groups	2901.23	109	26.61		
	Total	3670.99	111			
BRITISH	Between Groups	280.80	2	140.40	3.77	.02
	Within Groups	4056.25	109	37.21		
	Total	4337.06	111			
IRANIAN	Between Groups	816.24	2	408.12	21.07	.00
	Within Groups	2110.86	109	19.36		
	Total	2927.10	111			

The results of the comparison of means show that there is a significant difference between the three ability groups since the observed value of $F=14.46$ is greater than the critical value. In order to determine where the exact differences lie, a post-hoc analysis was carried out.

Table 3. Tukey HSD: Post-hoc comparison

Dependent Variable	(I) GROUP	(J) GROUP	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
AMERICAN	1.00	2.00	-.15	1.20	.99	-3.02	2.71
		3.00	6.86(*)	1.57	.00	3.13	10.60
	2.00	1.00	.15	1.20	.99	-2.71	3.02
		3.00	7.02(*)	1.33	.00	3.84	10.20
	3.00	1.00	-6.86(*)	1.57	.00	-10.60	-3.13
		2.00	-7.02(*)	1.33	.00	-10.20	-3.84
BRITISH	1.00	2.00	-1.49	1.42	.55	-4.88	1.90
		3.00	2.81	1.85	.28	-1.59	7.22
	2.00	1.00	1.49	1.42	.55	-1.90	4.88
		3.00	4.30(*)	1.58	.02	.54	8.06
	3.00	1.00	-2.81	1.85	.28	-7.22	1.59
		2.00	-4.30(*)	1.58	.02	-8.06	-.54
FARSI	1.00	2.00	.16	1.02	.98	-2.28	2.60
		3.00	-7.07(*)	1.33	.00	-10.25	-3.88
	2.00	1.00	-.16	1.02	.98	-2.60	2.28
		3.00	-7.23(*)	1.14	.00	-9.94	-4.52
	3.00	1.00	7.07(*)	1.33	.00	3.88	10.25
		2.00	7.23(*)	1.14	.00	4.52	9.94

* The mean difference is significant at the .05 level.

The results of the analysis reveal that beginners and intermediate students favored the Standard American accent significantly more than the Standard British and Farsi accents of English. There was no meaningful difference in attitude towards the three accents among intermediate students. Finally, advanced learners showed significantly more pleasant attitudes towards the Farsi accent of English in comparison with the Standard British and Standard American accents. From another perspective, Standard American was found to be the most favored accent followed by the Standard British and finally, the local Farsi accent of English.

In the second part of the survey, each of the three statements was rated by participants in the three groups. The test of Analysis of Variance (ANOVA) was once again used to determine whether the three groups differed in their beliefs towards the statements. The results of the analysis for the first item can be found in Figures 4 and 5.

Table 4. ANOVA: Proficiency level by level of agreement; Item 1

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	46.63	2	23.31	18.76	.00
Within Groups	135.43	109	1.24		
Total	182.06	111			

The results of the comparison of means reveal that based on the critical value, the three groups differed in their level of agreement with this item.

Table 5. Tukey HSD: proficiency by Item 1

(I) GROUP	(J) GROUP	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1.00	2.00	.62(*)	.26	.04	.00	1.24
	3.00	2.04(*)	.33	.00	1.23	2.84
2.00	1.00	-.62(*)	.26	.04	-1.24	-.00
	3.00	1.41(*)	.28	.00	.72	2.09
3.00	1.00	-2.04(*)	.33	.00	-2.84	-1.23
	2.00	-1.41(*)	.28	.00	-2.09	-.72

* The mean difference is significant at the .05 level

Presenting the multiple comparison of means for the first statement in the second section of the survey, the results indicate the points of difference. The findings show that with regards to the first item, beginners agreed significantly more with this statement than intermediate and advanced learners of English. Advanced learners disagreed with this statement significantly more than any other group. Figures 6 and 7 below describe the results of the same analysis for the second statement.

Table 6. ANOVA: proficiency level by level of agreement, Item 2

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	20.90	2	10.45	12.88	.00
Within Groups	88.37	109	.81		
Total	109.27	111			

The results of the comparison of means for the second statement in the second section of the survey show that once again the three proficiency groups differed in their level of agreement on this item.

Table 7. Tukey HSD: proficiency by Item 2

(I) GROUP	(J) GROUP	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1.00	2.00	.3800	.21061	.17	-.12	.88
	3.00	1.3537(*)	.27405	.00	.70	2.00
2.00	1.00	-.3800	.21061	.17	-.88	.12
	3.00	.9737(*)	.23366	.00	.41	1.52
3.00	1.00	-1.3537(*)	.27405	.00	-2.00	-.70
	2.00	-.9737(*)	.23366	.00	-1.52	-.41

* The mean difference is significant at the .05 level.

The results of the comparison of means for the second statement in the second section of the survey indicate that beginner and intermediate learners did not differ significantly in their beliefs towards the second statement. However, advanced students disagreed significantly more with this statement than both intermediate and elementary learners.

Figures 8 and 9 show the results of the ANOVA for the third statement.

Table 8. ANOVA: proficiency level by level of agreement, Item 3

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	11.08	2	5.54	7.58	.00
Within Groups	79.62	109	.73		
Total	90.71	111			

The comparison of means for the third statement in the second section of the survey points out that the three proficiency groups also differed in their level of agreement on this item.

Table 9. Tukey HSD: proficiency by Item 3

(I) GROUP	(J) GROUP	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1.00	2.00	.38	.19	.13	-.08	.86
	3.00	1.01(*)	.26	.00	.39	1.62
2.00	1.00	-.38	.19	.13	-.86	.08
	3.00	.62(*)	.22	.01	.09	1.14
3.00	1.00	-1.01(*)	.26	.00	-1.62	-.39
	2.00	-.62(*)	.22	.01	-1.14	-.09

* The mean difference is significant at the .05 level.

According to the multiple comparison of means for the third statement in the second section of the questionnaire, the results for the second and third item are very similar. Once again, there was no significant difference between the beliefs of elementary and intermediate students. Nevertheless, respondents in both groups agreed with the third statement significantly more than their advanced counterparts.

The findings of the study indicate that most Iranian EFL learners exhibit more positive attitudes towards the Standard American accent of English. This tendency was seen among elementary and intermediate learners. Advanced learners, however, had significantly lower attitudes towards the Standard American accent. Beginners and intermediate learners found the Standard American accent more pleasant, friendly and prestigious. They also rated the speaker of Standard American English to be the most educated and comprehensible. The marked positive attitude towards this accent could be attributed to the widespread use of textbooks drawing heavily on American English. Another possible reason for this trend could be the effect of mass media, such as television, movies and radio programs on learner attitudes.

Following the Standard American accent, Standard British was found to draw the most positive attitudes towards itself. Nevertheless, it was the only accent for which there was not significant difference among elementary, intermediate and advanced learners. It was regarded by most advanced learners to be the most prestigious and its speaker was rated as being more educated than that of any other accent.

The participants of this study had the least positive attitude towards the local Farsi accent of English. However, it is noteworthy that despite its overall lack of popularity, it was found to be more intelligible than the Standard British accent. Perhaps the most

interesting finding of this study was the marked preference among advanced learners towards this accent of English. Although further studies would have to be conducted in order to investigate the reason for this outcome, one could speculate that during their years of studying English, advanced learners become more aware of the utilitarian purpose of learning a foreign language, and, hence, gain more positive attitudes towards their local accent. This preference could also be seen as the result of advanced learners' commonly failed attempts to learn to speak English with a Standard American or Standard British accent.

As the participants in this study became more proficient, they place significantly less importance on learning English with a native accent. Once again, it could be speculated that with experience in learning a foreign language comes the awareness that there are more important aspects to attend to than accent. It may be for this very reason that advanced learners attached the least importance to speaking with a native accent than either elementary or intermediate learners.

Elementary and intermediate learners of English in the present study were not significantly different in their beliefs on the possibility of acquiring a native accent. Advanced learners, on the other hand, believed that it was very improbable for them to speak English with a native accent. This may also have been the result of their experience. As opposed to elementary and intermediate learners who were optimistic towards learning to speak with a native accent, advanced learners believed that achieving a native accent in the English language was significantly less achievable for them.

Advanced learners who participated in this study were of the opinion that less time should be spent on pronunciation practice and accent training in the classroom as opposed to learning grammar, vocabulary and other language skills and components. Elementary and intermediate learners held the belief that more time should be spent on accent training. The importance placed on accent training in the classroom by elementary and intermediate learners may be due to their views on the possibility of achieving a native English accent.

Conclusion

In a time when accent training seems to be effort put to waste, it appears that learners, particularly those in the elementary and intermediate levels, still look forward to

speaking English with a native accent. This subjective need on the part of learners runs contrary to the current significance attached to functional intelligibility and local accent varieties. The question which arises is whether these needs are worthy of our attention or whether they should be neglected for the sake of more compelling needs which learners themselves are not always aware of.

How much value should language planners and instructors set on what students themselves believe to be important for their learning? Dudley-Evans and St. John (1998) introduce a number of sources for gathering information for the purpose of needs analysis. These sources include people studying and working in the field, employers, colleagues, ex-students, documents relevant to the field and the learners, clients and the learners. According to this list, learners are only one of the many sources for collecting information while performing needs analysis. For this reason, it could be argued that the students' perceptions of their needs should not outweigh the other sources which are also crucial to the performing of needs analysis.

Conversely, it could be argued that learners are at the very core of the learning process. That is, unless their needs are addressed, their motivation, which is central to their successful learning, will be drastically reduced. What is more, acknowledging learner needs is a vital component of the liberal education movement. But it could also be claimed that learners are not always aware of their real needs, similar to a child who will not take prescribed medicine due to being unaware of its benefits.

Policies in which learners' perceived needs are rarely taken into account dominate the language education system in Iran. Language schools often impose the variety of English which they deem fit upon their learners. Decisions regarding which variety to teach chiefly depend on the availability of textbooks, materials and the overall preference of teachers. It is common for learners favoring a particular variety to abandon their preference to the domination of one particular type among teachers and policy-makers.

One solution to this problem is to increase and foster the rate of interaction and negotiation between stakeholders. In other words, instead of dictating learners' needs without allowing them to air their views or relying entirely on what the learners perceive to be beneficial for them, we should encourage dialogue among learners, teachers, language planners and all other parties involved in the language learning process.

References

- Alford, R. L., & Strother, J. B. (1990). Attitudes of native and nonnative speakers towards selected regional accents of U.S. English. *TESOL Quarterly*, 24(3), 479-495.
- Becker, P. (1995). *The etiology of foreign accent: Towards a phonological component of identity*. Unpublished master's thesis, Southern Illinois University: Carbondale.
- Berwick, R. (1989). Needs assessment in language programming: from theory to practice. In R. K. Johnson (Ed). *The Second Language Curriculum*. Cambridge: Cambridge University Press.
- Brennan, E. M., & Brennan, J. S. (1981). Accent scaling and language attitudes: reactions to Mexican American English speech. *Language and Speech*, 24(3), 207-221.
- Brindley, G. (1989). *Assessing achievement in the learner-centered curriculum*. Sydney: National Center for English Teaching.
- Cargile, A. C. & Giles, H. (1998). Language attitudes toward varieties of English: An American-Japanese context. *Journal of Applied Communications Research*, 26, 338-356.
- Cargile, A. C. (2000). Evaluations of employment suitability: Does accent always matter? *Journal of Employment/Counseling*, 37, 165-177.
- Chiba, R., Matsuura, H., & Yamamoto, A. (1995). Japanese attitudes towards English accents. *World Englishes*, 14, 77-86.
- Dixon, J.A. & Mahoney, B. (2004). The effect of accent evaluation and evidence on a suspects' perceived guilt and criminality. *Journal of Social Psychology*, 144(1), 63-73.
- Dudley-Evans, T. & St John, M. J. (1998). *Developments in English for specific purposes: A multi-disciplinary approach*. Cambridge: Cambridge University Press.
- Ekong, P. (1982). On the use of an indigenous model for teaching English in Nigeria. *World Language English* 1(2), 87-92.
- Forde, K. (1995). A study of learner attitudes towards accents of English. *Hong-Kong Polytechnic University Working Papers in ELT & Applied Linguistic*, 1, 59-76.
- Fraser, H. (2000). *Coordinating improvements in pronunciation teaching for adult learners of English as a second language*. Canberra: DETYA.

- Fromkin, V. Rodman, R. and Hyam, N. (2003). *An introduction to language*. Boston, MA: Heinle.
- Jenkins, J. (2005). *World Englishes: A resource book for students*. London: Rutledge.
- Johnson, R., & Frederick, L. (1994). *Native speakers' perceptions of nonnative speakers: Related to phonetic errors and spoken grammatical errors*. Paper presented at the annual meeting of the teachers of English to speakers of other languages, Baltimore, MD.
- Manzano, I. T. (1997). *The effect of accent on listening comprehension and attitudes of ESL students*. Unpublished master's thesis, University of Puerto Rico, Mayaguez Campus.
- Munby, J. (1978). *Communicative syllabus design*. Cambridge: Cambridge University Press.
- O'Grady, W., Archibald, J., Arnoff, M., & Rees-Miller, J. (2005). *Contemporary linguistics: An introduction*. (5th edition). Bedford: St. Martin's.
- Ryan, E. B., & Carranza, M. A. (1977). Ingroup and outgroup reactions to Mexican American language varieties. In H. Giles (ed.), *Language, Ethnicity and Intergroup relations*, (pp. 59-82). London: Edward Arnold.
- Southwood, M. H., & Fledge, J. E. (1999). Scaling foreign accent: Direct magnitude estimation versus interval scaling. *Clinical Linguistics and Phonetics*, 13(5), 335-349.
- Warschauer, M. (2000). The changing global economy and the future of English teaching. *TESOL Quarterly*, 34(34), 511-535.
- Wilcox, G. K. (1978). The effect of accent on listening comprehension: A singapore study. *English Language Teaching Journal* 32, 118-127.
- Yule, G. (1985). *The study of language: An Introduction*. Cambridge: Cambridge University Press.

Title

The Effect of Time Intervals on the Keyword Mnemonic Technique Used in Teaching Vocabulary to Iranian Female EFL Learners

Authors

Nasser Rashidi (Ph.D)

Shiraz University, Shiraz, Iran

Mansoureh Sajjadi (M.A in TEFL)

Islamic Azad University, Abadeh Branch, Iran

Bio Data

Nasser Rashidi, an academic member of the Department of Foreign Languages and Linguistics of Shiraz University, Shiraz, Iran. His area of research is TEFL including language teaching and testing, discourse analysis and sociolinguistics. His recent publications include *Practical Contrastive Analysis of English and Persian with Special Emphasis on Grammar* with L. Yarmohammadi (2009) and *Practical Contrastive Analysis of English and Persian with Special Emphasis on Vocabulary* with L. Yarmohammadi (2010).

Mansoureh Sajjadi holds an MA in TEFEL from the Department of Foreign Languages of Azad University, Shiraz branch, Iran. She teaches English courses and IELTS at different institutes in Shiraz. She is a faculty member of the Department of English, Azad University, Abadeh Branch.

Abstract

This research investigates the effect of time intervals on vocabulary retrieval of Iranian female EFL learners who used the keyword method to encode the information entering the long term memory. The design of this study required two groups of learners, so sixty students (all females) served as

subjects. They had enrolled in conversation classes level 10 at Payam Language Institute in Shiraz., Iran. The students were randomly assigned to control and experimental groups. The treatment in this study was based on Atkinson's human information processing system (1975) and took advantage of using keyword method during encoding information in transferring information from short term memory to make the retrieval easier. The study however, compared learners' immediate recall with their delayed recall and involved two conditions either specific to control group and experimental group. The experimental group enjoyed some special Flashcards that involved pictures and keywords on them. Keywords selected were all root-related to the meanings of the words that students were assigned to memorize. Participants in the experimental group were asked to learn their words using keywords provided and make relationship among the meanings of words, keywords and pictures on Flashcards as it had been used in the classroom by the teacher while teaching. They had 15 minutes to process 10 words and then there were two tests to take. One production test and one comprehension test. Two days later a surprise recall test was given. Tests sheets were exactly the same as before. To examine the effects of the treatment a t-test was run. Results of data analysis showed that mnemonic keyword instruction resulted in higher levels of recall and comprehension. The findings can be used in both materials development and teaching vocabulary in reading comprehension classes.

Key Words: keyword method, encoding, retrieval, mnemonic device, recall, comprehension.

Introduction

There are many things we can do to make the learning process more memorable for our learners. Using pictures, interesting contexts and stories can help memory and giving the students the opportunity to practice the new vocabulary in personalized and meaningful tasks are also essential tools. The idea is that if the students are asked to analyze and react personally to new information, it will help them process the language

more deeply, facilitating their ability to retain it in their long term memory.

Theoretical Framework

What is your mobile phone number?

As the question is put to us we suddenly realize that we can't remember our numbers? We repeated them to ourselves over and over a few days before, it is there somewhere in our minds but we just can't get the information. Much to our embarrassment, we have to resort to finding it in our diaries.

Everyone one who has tried to force himself to learn information will know the problem. Despite repeating it again and again and being able to hold it in our short term memory, keeping information in our long term memory is much more difficult. Despite the fact that there seems to be an infinite amount of information that can be stored there, getting things into our long term memory and back again can be a difficult process.

Cue-dependent forgetting is an idea based on experiments where subjects found they forgot lists of learnt words but could remember them if they were given appropriate information to help them recall with a cue, (it's a fruit when they were trying to remember the word pear) students performance was greatly enhanced.

As teachers we need to take this into account and find ways of helping students to combat the effects of memory decay and give them the tools to improve their retrieval ability. We also need to think of ways of making the experience of learning vocabulary more memorable and of recycling the information that we teach. There are many things we can do to make the learning process more memorable for our learners, using pictures, interesting contexts and stories can help memory and give students the opportunity to practice the new vocabulary, so personalized and meaningful tasks are also essential tools. The idea is that if students are asked to analyze and react personally to new information it will help them process the language more deeply, and facilitate their ability to retain it in their long- term memory. This research study tries to introduce another way we can use to make the learning process more memorable for our learners.

Based on human information processing system, Atkinson's (1975) keyword method is used as a mnemonic technique in encoding the information entering the long term memory to make the retrieval easier. The keyword method is one type of linguistic mnemonics.

Means of communication have grown increasingly more efficient in recent years, exposing numerous weaknesses in human's ability to express ideas verbally. One limiting factor of communication is the lack of an adequate vocabulary acquisition and it has received a great deal of attention, including the keyword method. Keyword method acts as a mnemonic technique often referred to as "memory trick." Mnemonics work by developing a retrieval plan during encoding so that a word can be recalled through verbal and visual clues; mnemonics help learners because they aid the integration of new materials into existing cognitive structures and because they provide retrieval clues.

The integration of new materials into existing cognitive structures is called meaningful learning. As Ausubel (1968) put forth, the meaningful learning is a process of relating new material to relevant established entities in cognitive structure. As new material enters the cognitive field, it interacts with, and is appropriately subsumed under, a more inclusive conceptual system. The meaningful learning theory suggested by Ausubel is the theory behind this research study.

Learning and memory are very closely related and the terms used to describe one process are often used to describe the other. The term learning is often used to refer to the processes involved in the initial acquisition and encoding of information. However, the distinction is not hard and fast. After all, information is learned only when it can be recalled later, and retrieval cannot occur unless information was learned (Ausubel 1968).

The process of retrieval is considered in the human information processing system. (Atkinson and Shiffrin 1968, 1971). The human information processing system is a simple and plausible view of the human memory system. The system's basic structural features are three memory stores: sensory store, short term store, and long-term store. Information enters through sensory store. It is lost almost immediately unless it is rehearsed, then it maybe sent to short-term memory. If it is appropriately encoded, the information will be stored in long term memory. Material retrieved from long term memory is processed back through short term memory. As stated previously mnemonics work by developing a retrieval plan during encoding. One way of encoding information is the keyword method of Atkinson (1975). Research has consistently indicated that the use of mnemonic devices substantially enhances higher levels of retention in immediate recall of second language words in comparison with other

learning strategies. (Atkinson, 1975; Wang and Thomas, 1992; Van Hall and Candia Mahn, 1997).

This study therefore tries to answer the following questions:

1. Does the keyword method affect the relationship between Iranian female EFL learners' vocabulary memorization and vocabulary retrieval?
2. If the answer to the first question is yes, what is the effect of time intervals on the results?

Review of Literature

In the mid-seventies, Atkinson and Raugh (1975) found remarkable results in using the keyword method to teach Russian vocabulary to college students. While later studies have not tended to find such dramatic results, nevertheless a large number of studies have demonstrated the advantage of the keyword mnemonic to learn vocabulary. Some researchers have become interested in the strategy. Others have suggested a number of limitations.

The keyword method is undeniably an effective way for accelerating the EFL learners speed of learning of suitable material, nor is there any doubt that it improves immediately recall (McDaniel, Pressley, and Dunay, 1987) . However, what people want is long-term recall, and it is there that the advantages of the keyword method are most contentious. while many studies have found good remembering a week or two after learning (Hall, Wilson, and Patterson, 1981) using the keyword mnemonic, others have found that remembering is no better one or two weeks later whether people have used the keyword mnemonic or another strategy. Some have found it worse (Gruneberg, 1998). It has been suggested that, although keyword may be a good retrieval cue initially, over time earlier associations may remain their strength and make it harder for EFL learners to retrieve the keyword image. Note that it is not the keyword itself that fails to be remembered. It is the image. The weakness then, is in the link between keyword and image. This is the link that you must strengthen to use the mnemonic method.

For practical purposes, it is merely sufficient to remember that, for long-term learning, you must strengthen this link between keyword and image (or sentence)

through repeated retrieval. A number of factors may affect the strength of a keyword mnemonic. One that is often suggested is whether or not the mnemonic is supplied for the students, or thought up by them (Old, n.d.). Intuitively, we feel that a mnemonic you've thought up yourself will be stronger than one that is given to you. One study (Wang and Thomas, 1995) that compared the effectiveness of keywords provided versus keywords that are self generated, found that participants who were required to make up their own keywords performed much worse than those who were given keywords. This doesn't answer the question of the relative durability, but it does point to how much more difficult the task of generating keywords is. This has been confirmed in other studies. As Wang et al. (1992) say the quality of the keyword mnemonic may affect its durability. Mnemonics that emphasize distinctiveness, that increase the vividness and concreteness of the words to be learned, are remembered less well overtime than mnemonics that emphasize relational and semantic information. This is why the emphasis in recent time is on making interactive images or sentences, in which the keyword and definition interact in some way. Having bizarre images seem to help remembering immediately after learning (when there is a mix of bizarre and less unusual images), but doesn't seem to help particularly over the long term (Wang and Thomas, 1995).

Hall et al. (1981) suggest that keywords that are semantically as well as acoustically related to the word to be learned might prove more durable. However, the evidence does not explain why the immediate benefits of mnemonic devices fail to extend to long-term retention. In addition, research on mnemonic has drawn mostly on the assessment of phonetic related languages such as English as opposed to image-related languages such as Chinese (Wang and Liu, 2003). Brown and Perry (1991) identified three main types of mnemonic strategies: (1) keyword, (2) semantic, and (3) keyword semantic. According to Ellis (1997), the core of the keyword method is to create a form and semantic connection between the target word and its translation. A semantic strategy attempts to evoke an association between the target word and the learners existing semantic system. A keyword semantic strategy involves a combination of the aforementioned strategies. Research has shown that keyword mnemonic is an effective device in accelerating learning speed and in boosting immediate recall of second language vocabulary.

However, other research (e.g., Van Hall, and Candia Mahn, 1997) indicated the keyword mnemonic strategy in relation to a non-mnemonic strategy (e.g., rote

learning). Thus a discussion of the feasibility of keyword mnemonics in learning foreign language vocabulary must consider the various factors affecting the efficacy of the keyword mnemonic strategy in relation to a non-mnemonic learning strategy. Mastropieri and Scruggs (2001) consider the effect of the keyword method on teaching abstract vocabulary. Of course several recent investigations have evaluated the effectiveness of the keyword method in facilitating the vocabulary learning of students. Taylor (1981) provided some initial evidence that the keyword method was highly effective in teaching vocabulary information to a sample of boys, but a methodological flaw inhibited unequivocal conclusion from this study (Scruggs, Mastropieri, and Levin, 1987). Van Hall and Candia Mahn's (1997) discussion of native Dutch speakers learning Spanish sheds light on how efficacy of the keyword method interact with various variables, such as concreteness of the foreign word, the semantic qualities of the keywords and language learning experience. Van Hall and Candia Mahn constructed a factorial design in order to examine the efficacy of the keyword method in learning Spanish in terms of learning condition (mnemonic VS. rote learning). The three factors were language learning experience (multi-lingual VS. monolingual), concreteness of the target word (concrete VS. abstract), and semantic relatedness of the keyword to the target word (semantically-related VS. semantically irrelevant).

In their view, assessment of all of these variables is needed to gauge the success of keyword mnemonic strategies that will magnify immediate recall for all learners. Van Hall and Candia Mahn (1997) contend that the keyword method is so effective for either experienced or inexperienced learners. Although psychological theories provided partial explanation of how memories interact with the properties of words, the mystery is still left unresolved (Wang and Thomas, 1992). Despite the documented success of the keyword mnemonic strategy in the short term retention there remains a practical concern about the efficacy of using keywords for long-term retention.

Teaching Abstract Vocabulary with the Keyword Method: Effects on Recall and Comprehension

Mastropieri and Scruggs's (2001) recent investigations consider the effect of the keyword method on teaching abstract vocabulary. Of course, several recent investigations have evaluated the effectiveness of the keyword method in facilitating

the vocabulary learning of students. Taylor (1981) provided some initial evidence that the keyword method was highly effective in teaching vocabulary information to a sample of boys, but a methodological flaw inhibited unequivocal conclusions from this study (Scruggs, Mastropieri, and Levin, 1987). In two experiments, Scruggs, Mastropieri, Levin, McLoone, Gaffney, and Prater (1985) taught middle school students a list of 14 English words, using either a keyword condition or an experimenter-led drill-and-practice direct instruction condition. In the first experiment, the keyword condition students were provided with keywords and interactive pictures, and in the second experiment, the keyword condition students were asked to generate their own interactive images. In both experiments, keyword condition students greatly outperformed direct instruction condition students; although in the second experiment the performance differential was somewhat lower.

McLoone, Scruggs, Mastropieri, and Zucker (1986) again compared keyword instruction and direct instruction rehearsal condition with adolescents who were learning both native English and Italian words. In the first task, instruction was experimenter directed for both conditions. In the second task, students in both conditions were given instruction in strategy transfer and asked to apply their respective strategies to a new list of words. The keyword condition students again outperformed the direct instruction rehearsal students on both tasks. In a fourth investigation, Veit, Scruggs, and Mastropieri (1986) taught, among other things, Greek root words relevant to the study of prehistoric reptiles e.g., ptero, meaning "winged"; Saur, meaning "Lizard" to middle school students using the keyword method or experimenter-led drill and practice.

The keyword condition students not only learned more meanings but were also more able to apply the information when entire prehistoric reptile names were provided e.g., pterosaur, meaning "winged lizard". Finally, Condos, Marshall, and Miller (1986) evaluated the efficacy of the keyword method when applied to existing curriculum materials and taught by teachers over time in their own classrooms. It was found that students taught via the keyword method scored consistently higher than students of control group on a series of recall tests administered over a 5-week instructional period, and they maintained this advantage over a 10-week delayed recall period.

Effects of Language Learning Experience

Contrary to the common belief that mnemonic strategies will magnify immediate recall for all learners, Van Hall and Candia Mahn (1997) contend that the keyword method is so effective for either experienced or inexperienced learners. In their study of the performance variation exhibited in learners received keyword mnemonics and rote rehearsal, inexperienced learners received either of the learning conditions recalled comparable numbers of learned items. For experienced foreign language learners, rote learners' performance surpassed that of keyword learners. Thus, the keyword method does not seem to magnify as experienced learners retention level. The fact that the retention rate of experienced learners in the vocal rehearsal group surpassed that of the mnemonic group suggests that experienced learners seem to be more sensitive to vocal rehearsal in learning foreign language words. Presumably, their refined phonological knowledge enables them to be more receptive to the vocal information and thereby more able to access the words stored in their memory. The involvement of phonology or vocal rehearsal seemed to play a more significant role in retaining novel foreign words.

Effects of Vocal Rehearsal

A strong link between vocal stimuli and memory has been documented in the psychological and psycholinguistic studies (e.g., Conrad, 1964; Salame and Baddeley, 1982). In asking subjects to perform a read-aloud task, Conrad (1964) noticed that people tend to confuse words with similar pronunciation (acoustic errors) but not words with similar orthography (visual errors). Furthermore, very often people have the experience of reading a passage and hearing the words echoing in their heads. In fact, this kind of automatic phonological coding, triggered by visual input, does not require actual articulatory movement or verbal rehearsal. The above behavioral evidence shows that people read visually presented input and encode it acoustically. In addition, the fact that people can very easily come up with rhymes for words shows that storage of words, whether originally presented visually or acoustically, seems to be organized phonologically in our working memory. Hence, phonological stimuli, as can be obtained in vocal rehearsal, may be made to play a central role in enhancing the level of retention in a phonetic-related language, where the high spelling-sound correspondence might be the keyword recalling. The link between the keyword and novel foreign word is one of the foci in Van Hall and Candia Mahn's (1997) research discussing the

feasibility of mnemonic devices.

Learner-Generated Versus Teacher-Supplied Mnemonics

Psychological theories do provide partial explanation of how memory interacts with the properties of words. Nonetheless, the mystery is still left unresolved. Wang and Thomas's (1992) research study about using mnemonics on long-term retention offers a possible explanation of the encoding and recall variability- the way the keyword is provided: teacher-supplied or subject-generated. Considering that the role of the keyword is to help engage learners to associate new materials with existing knowledge in a meaningful way, Wang and Thomas contend that spontaneous subject-generated encoding is more likely to be retained after an extended interval. Teacher-supplied mnemonics may function effectively in the initial phase because the episodic cues that are incorporated in the keyword image are still available in the working memory during the immediate test of recall. However, the evanescent nature of the teacher-supplied keyword images means the learner will be prone to forgetting when recall is tested after an extended delay. Additionally, learners from different cultural backgrounds might arrive at different interpretations of a given foreign keyword and thus produce different mental images for the foreign words.

Relative Effectiveness of Mnemonic Strategies versus Rote learning

Despite the documented success of the keyword mnemonic strategy in the short term retention, there remains a practical concern about the efficacy of using keywords for long-term retention. Nonetheless, proponents of the keyword method regard rote learning simply as a meaningless repetition of information and criticize the basis of rote-rehearsal and Hebb's (1949) consolidation theory for treating learning as a passive process. According to this theory, short-term memory can be transferred to long-term memory and cause solid structural changes in the brain if the material is rehearsed long enough. More specifically, despite the common belief that short-term memory and long-term memory bear physiological differences, proponents of the theory believe that the transfer of information from short-term memory to long-term memory is simply a matter of time and rehearsal.

Craik and Larkhart (1972) proposed rehearsal theory to modify the insufficiencies of consolidation theory. They contend that rote learning does not necessarily have to be meaningless repetition. According to rehearsal theory, sensory information is processed

at two hierarchically arranged levels of perceptual analysis, from superficial to complex, namely, maintenance rehearsal and elaborative rehearsal. Maintenance rehearsal is the verbal/ visual material being processed in a repetitive manner e.g., the way you try to remember a 7-digit telephone number. In contrast, elaborative rehearsal involves conscious or unconscious organization of input information based on prior experiences (Wang and Liu, 2003).

Method

Participants

This study was conducted with 60 EFI learners studying at Payam language Institute in Shiraz at level 10 that was based on the standards of the institute. Participants were between 15 to 18 years old. At the time of the experiment the subjects, all new comers to level 10, had been exposed to at least 360 hours of instruction in L₂. The reason for selecting students of level 10 is that they are required to know many words to pass the final exam and sit in the next level. The sample of 60 was selected out of 100 students studying at level 10. They were asked to participate in two types of vocabulary test constructed based on vocabularies they had studied at previous levels. They were asked to answer questions of a literal recall test and questions of a comprehension test provided in a matching format. The sample was selected on the basis of the scores on the test. Sixty students with higher scores were selected. They were divided into two groups, control and experimental groups. The researchers paid attention to their scores to make sure that the two groups are completely homogenous. However, due to the difficulties of selecting male subjects including lack of availability the sample consists of only female EFI learners.

Materials

Lists of ten words and definitions were presented to each participant, along with keywords and sentences relating each definition to its keyword. Words used in this investigation were taken from a somewhat larger list provided by Old (n.d.) and included both abstract and concrete words. The list consists of 4 abstract and 6 concrete words. Two additional words one concrete and one abstract were used for practice examples. Keywords selected were all root-related to the meanings of the vocabularies. **(See Appendix A)** This study involves two conditions either specific to control group

and experimental group. Keyword condition is used to name the experimental group and rehearsal condition is used to name the control group. Materials specific to each condition are as follows.

Keyword Condition

Materials for the keyword condition consisted of twelve index cards. On the top of each card were printed the word, keyword in parentheses, and definition. In the middle of each card, the keyword was pictured interacting with its definition in a line drawing. For example, for *oxalis*, meaning clover-like plant, an ox (keyword for *oxalis*) was pictured eating clover-like plants. For abstract words, the keyword was pictured interacting with an instance of the definition. For example, for *Octroi*, meaning a tax paid on goods upon entering a town, an octopus (keyword for *octro*) was pictured collecting taxes from persons entering a town. For *vituperation*, meaning abusive speech, a viper (keyword for *vituperation*) was pictured speaking abusively to someone.

Rehearsal Condition (Direct Instruction Condition)

Materials for this condition were the same as those for the keyword condition, with the exception that any keyword or pictured references to keywords were omitted. For example, the *oxalis* picture omitted the ox, and the *octroi* picture presented a person, rather than an octopus, collecting taxes from people entering a town. With this exception, the drawings and lettering were identical to the materials used in keyword condition.

Tests

Two tests were constructed for this investigation. One was a literal recall test, where students were asked to produce the definition of each word, for example, "what does *vituperation* mean?" (See **Appendix B**) The second test was a comprehension test, in which students were asked, in a matching format, to provide the appropriate word, given a novel instance of the word. (See **Appendix C**) In order to minimize the effect of student's stress and lack of concentration on their scores in the literal recall test, there was no time limit and the participants were asked to answer questions in a silent classroom. To refresh student's memory they were asked one by one to recite the definition of an optional word from the list. Then other words were written by them. In order to minimize student's test-wiseness and to minimize the probability of selecting

alternatives by chance, students were asked to answer questions of the comprehension test in a matching format. In both tests the criteria that constitutes the adequate performance was 10. In rating the literal recall test the rater focused primarily on the content of the definitions written by students. There was no attention paid to grammar, cohesion and organization of answers.

Procedure

Following a brief (approximately 5 minutes) explanation of the keyword method using examples, participants were asked to learn their lists of words using the keywords provided. They were given 15 minutes of study time. The students were told that they were going to learn some new words, and that they should try their best because they were going to be given a test at the end of the lesson. Procedures unique to each condition are described below.

Keyword condition

Students were first taught keywords for the two practice examples. They were then shown the mnemonic pictures for each new word and asked to look at them while the experimenter said, for example, "The keyword for chiton is kite chiton means 'loose garment'. Remember this picture of people making kites out of their loose garments." The students were then asked the definition of the word and asked to describe the picture. The students were shown each of the two practice pictures for 30 seconds. After the practice examples had been shown, the students were given a practice production test and comprehension test on the items. Students were given feedback on their answers. Following the practice test, students were taught the keywords and shown the 10 target words, presented in a randomized order. Students were shown each picture for 30 seconds while the experimenter described the keyword and the pictorial strategy as described above. Again, students were asked to provide the definition of each new word once and to describe the interactive picture once. At the end of the training picture, students were given the production recall test followed by the comprehension test. No time limits were placed on the test administration. In order to examine the effect of the keyword method on delayed recall, two days later, a "surprise" recall test was given. Test sheets were exactly the same as before. Again, no time limits were placed on the test administrations.

Rehearsal Condition

This condition was selected as a best available alternative treatment condition (Borkowski and Buchel, 1983), and was implemented as employed in previous vocabulary training studies (Scruggs, Mastropieri, Levin, and Gaffney, 1985). This type of vocabulary teaching strategy was described by Carnine and Silbert (1979) as a commonsensical approach”: They were first told the definition and then were asked to repeat it. In this condition, students were individually taught the two practice examples using the labeled pictures and procedures described above. For example, “chino means loose-fitting clothing; what does chiton mean? Drill and practice were provided on these two words, followed by a practice recall and comprehension test, as in the keyword condition.

Following the practice test, students received a brief preview of the target words, during which they practiced pronouncing the words. This activity replaced the keyword learning activity in the keyword condition. Students were then given instruction in the twelve words, using experimenter-led drill and practice, rapid- paced questioning, and corrective feedback. Students received, as in the keyword condition, 30 seconds on each of the twelve words. In this condition, however, an additional minute was spent on a review of all words. As this review period replaced the time spent on initial learning of keywords in the keyword condition, the rehearsal condition students actually received more time for learning target information. Like the keyword condition two days later, a "surprise" recall test was given. Test sheets were exactly the same as before.

Scoring

Responses on the recall test were awarded 1 point for a completely accurate answer, half a point for an incomplete answer that nonetheless contained some critical aspects of the target response, for example, "speech" for "buncombe", rather than "boring or empty speech". The comprehension test employed identification format in which answers were either correct or incorrect; therefore, no partial credit was given.

Results and Data Analysis

Means, Standard deviation and other statistics for comprehension test and production test used in keyword method and rehearsal method are as follows:

Table 1 Statistics for Keyword Immediate Recall Comprehension Test Scores

Mean	7/1667
Std. Error of Mean	/32899
Median	7
Mode	6/00a
Std. deviation	1/80198
Variance	3/24713
Skewness	-.267
Std. Error of Skewness	0/427
Kurtosis	-.395
Std. Error of Kurtosis	/833
Range	7/00
Minimum	3/00
Maximum	10/00
Sum	215/00

Table 2 Statistics for Keyword Immediate Recall Production Test Scores

Mean	6/3167
Std. Error of Mean	0/41139
Median	6/25
Mode	4/00
Std. Deviation	2/25329
Variance	5/07730
Skewness	-.052
Std. Error of Skewness	0/427
Kurtosis	-1/439
Std. Error of Kurtosis	0/833
Range	7/00
Minimum	2/50
Maximum	9/50
Sum	189/50

Table 3 Statistics for Rehearsal Immediate Recall Comprehension Test Scores

Mean	5/9
Std. Error of Mean	0/34358
Median	6/0000
Mode	6/00
Std. Deviation	1/88186
Variance	3/54138
Skewness	0/587
Std. Error of Skewness	0/427
Kurtosis	-/147
Std. Error of Kurtosis	0/833
Range	7/00
Minimum	3/00
Maximum	10/00
Sum	177/00

Table 4 Statistics for Rehearsal Immediate Recall Production Test Scores

Mean	5/0333
Std. Error of Mean	0/34900
Median	407500
Mode	5/50
Std. Deviation	1/91155
Variance	3/65402
Skewness	1/016
Std. Error of Skewness	0/427
Kurtosis	0/757
Std. Error of Kurtosis	0/833
Range	7/5
Minimum	2/5
Maximum	10/00
Sum	151/00

Data Analysis

To compare the results of keyword method immediate recall comprehension test and rehearsal method immediate recall comprehension test a paired- samples t-test was used. And the results are as follows:

Table 5 Descriptive Statistics and t-test for the Two Sets of Scores

	Mean	N	Std. Deviation	Std. Error Mean
Pair 1 TEST1	7.1667	30	1.80198	.32899
TEST3	5.9000	30	1.88186	.34358

Paired Samples Correlations

	N	Correlation	Sig.
Pair 1 TEST1 & TEST3	30	.941	.001

Paired Samples Test

.001

	Paired Differences					t
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference		
Pair 1 TEST1 - TEST3	1.2667	.63968	.11679	1.0278	1.5055	10.846

Paired Samples Statistics

	d.f	Sig.(2-tailed)
Pair 1 TEST1 - TEST3	29	.001

As shown in the above table the difference between the scores on the two tests was statistically significant.

Keyword Method and Delayed Recall

To examine the effect of the keyword method on delayed recall, two days after the treatment a surprise recall test was given. Test sheets were exactly the same as before. Again, no time limits were placed on the test administrations.

Keyword Method Delayed Recall Comprehension Test

Raw scores are out of ten. Table 5 shows the raw scores.

Table 6 Keyword Method Delayed Recall Comprehension Test Scores

Scores	Frequency	Relative frequency	Percentage	F	Percentile
0	2	%	6%	2	6/6%
1	3	0/1	10%	5	16/6 %
2	10	0/3	30%	15	50%
3	10	0/3	30%	25	83%
4	3	0/1	10%	28	93/3%
5	2	6	9%	30	100%

Rehearsal Method Delayed Recall Comprehension Test

Raw scores are out of 10. Table 6 shows the raw score.

Table 7 Rehearsal Method Delayed Recall Comprehension Test Scores

Scores	Frequency	Relative frequency	Percentage	F	Percentile
0	3	%1	1%	3	10%
1	11	0/36	36%	14	46/6%
2	8	0/26	26%	22	73/3%
3	6	0/2	20%	28	93/3%
4	1	%3	3%	29	96/6%
5	1	%3	3%	30	100%

Delayed recall prediction and hypothesis testing

To compare the results of keyword method delayed recall comprehension test and rehearsal method delayed recall comprehension test a paired- samples t-test was used and the results are as follows:

Descriptive Statistics and t-test for the Two Sets of Scores

Std. Deviation	Mean		
Pair	Test 5	2.5	1/22
1	Test 6	0/9	1/25

sig. (2 ailed)	t	d.f		
Pair	Test 5	29	0/001	5/16
1	Test 6			

Here again as shown in the above table the difference between the scores on the two tests was statistically significant.

Discussion

This study examined the effects of the keyword method over time intervals. Results indicated that immediately following the instruction of words, students in treatment conditions were able to recall a relatively high percentage of word meanings. It means that only the keyword students, however, were able to learn a large number of words subsequent to instruction and maintain their vocabulary to some extent after two days. Of course, students were not very successful in delayed recall but even in this situation those who received the keyword method outperformed the students of control group. The findings of this investigation are in agreement with other research indicating that the keyword mnemonic technique is useful for learning and recalling information (Pressley and Levin, 1978)

Results of this investigation also indicated that keyword mnemonic instruction resulted in higher levels of recall and comprehension than a rehearsal condition. Moreover, this study in part replicated previous findings that students would learn substantially more vocabulary when the teacher uses the mnemonic keyword method in

teaching vocabulary than when more traditional drill- and practice method is used by the teachers.

Mnemonic keyword instruction also resulted in higher levels of comprehension, as measured by an application task in which students were asked to provide appropriate words in novel (to the learner) instances of the words. This finding is of importance, for it may have been argued that the keyword method facilitates literal recall of new vocabulary definitions at the expense of the ability to use the vocabulary in any different context. It does not appear that there is anything in the keyword method that facilitates greater initial comprehension than alternative instructional procedures. The results of this investigation seem to suggest that students taught vocabulary via the keyword method are better able to use new words because they have initially learned more words than could be of use in such an application task. Taken with the results of the Veit et al. (1986) investigations, this investigation seems to indicate that there is an observed comprehension resulting from mnemonic instruction. This finding is further validated by the results of an investigation involving science facts rather than vocabulary learning (Scruggs, Mastropieri, McLoone, Levin, and Morrison, 1987). The fact that mnemonically instructed students have been shown to comprehend and learn information demonstrated that mnemonic instruction has great promise as a viable instructional alternative.

The success of the mnemonic keyword method may be due, in part, to an interaction of the properties of mnemonic instruction with student's imagination. The learning process is something concrete for students and they try to find something in vocabularies to use as a code. They can touch vocabularies even if they are abstract words. And they do not bother themselves memorizing long lists of vocabularies and forget it as soon as they put the list away.

Pedagogical Implications

Teachers too often find themselves in an awkward situation while teaching vocabularies. Sometimes their attempts fail to make a particular vocabulary understood. Teachers especially in language institutes make students check the meanings of vocabulary in advance to prevent misunderstanding and when the instruction is finished via any method they ask students to memorize vocabularies. To make sure that the students indeed do this, the teachers guarantee their words by establishing a date for a quiz.

The most common way of teaching vocabularies in language institutes is to teach vocabularies in specific contexts like the reading part of the units and encourage students to guess the meaning of a new word by first looking at all the other words surrounding it and then considering the general meaning of the phrase or sentence in which it is located. And where necessary, the meanings of words through definitions, mime, synonyms, antonyms, examples are provided and when all attempts fail, translation is used. And finally students get the concept if the teacher has the chance of an evil. But this is learning after all. What does guarantee students' vocabulary retrieval? Keyword method may help teachers in these contexts.

It is sincerely hoped that the findings of the present study offer enough empirical evidence for educators to write vocabulary books using keyword and pictures or to write teacher training books introducing teaching vocabularies via keyword method.

References

- Atkinson, R. C. (1975). Mnemotechnics in second-language learning. *American Psychologist*, 30, 821-828.
- Atkinson, R.C., and Raugh, M.R. (1975). An application of the mnemonic keyword method to the acquisition of a Russian vocabulary. *Journal of Experimental Psychology: Human Learning and Memory*, 104, 126-133.
- Atkinson, R.C., & Shiffrin, R.M. (1971). The relationship between recall, working memory and learning. Retrieved September 8, 2009 from <http://www.articlesbase.com/childhood-education-articles/the-relationship-between-recall-working-memory-and-learning-1122473.html#ixzz0yd2oXcR>
- Atkinson, R.C., & Shiffrin, R.M. (1968). Human memory: A proposed system and its control processes. In K.W. Spence & J.T. Spence (Eds.), *The Psychology of Learning and Motivation: Advances in Research and Theory* (Vol. 2). New York: Academic Press.
- Ausubel, David P. (1968). *Educational psychology: A cognitive view*. New York: Holt, Rinehart and Winston, Inc.
- Borkowski, J.G. & Buchel, F. (1983). Learning and memory strategies in the mentally retarded. In M. Pressley & J.R. Levin (Eds.), *Cognitive strategy research: Psychological foundations*, (pp. 101-128). New York: Springer.
- Brown, T. & Perry, F. (1991). A comparison of three learning strategies for ESL vocabulary acquisition. *TESOL Quarterly*, 25, 655-670.

- Carnine, D., & Silbert, J. (1979). *Direct instruction reading*. Columbus, OH: Merrill
- Condos, M.M., Marshall, K. J., & Miller, S.R. (1986). Effects of the keyword mnemonic strategy on vocabulary acquisition and maintenance by learning disabled children *Journal of Learning Disabilities, 19*(10), 609–13.
- Conrad, R. (1964). Acoustic confusions in immediate memory. *British Journal of Psychology, 55*, 75-83.
- Craik, F. I. M. & Larkhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal behavior, 11*, 671-68.
- Ellis, R. (1997). Explaining individual differences. *The Study of Second Language Acquisition*. Hong Kong: Cambridge University Press.
- Gruneberg, M.M. (1998). A commentary on criticism of the keyword method of learning foreign language. *Applied Cognitive Psychology, 12*, 529-532.
- Hall, J.W., Wilson, K.P. & Patterson, R.J. (1981). Mnemonics: Some limitations of the mnemonic keyword method for the study of foreign language vocabulary. *Journal of Educational Psychology, 73*, 345-57.
- Hebb, D. O. (1949). *The organization of behavior*. New York: Wiley-Interscience.
- Mastropieri, M. A., & Scruggs, T. E. (2001). Using mnemonic instruction to facilitate access to the general education curriculum. Retrieved August 8, 2009, from http://www.k8accesscenter.org/training_resources/Mnemonics.asp.
- McDaniel, M.A., Pressley, M. & Dunay, P.K. (1987). Long-term retention of vocabulary after keyword and context learning. *Journal of Educational Psychology, 79*, 87-9.
- McLoone, B. B., Scruggs, T. E., Mastropieri, M. A., & Zucker, S. F. (1986). Memory strategy instruction and training with learning disabled adolescents. *Learning Disabilities Research, 2*, 45–53.
- Old, S. R. (n.d.). The keyword method applied to root-learning of vocabulary <http://clearinghouse.missouriwestern.edu/manuscripts/282.php&webclearinghouse.net>
- Pressley, M., & Levin, J. R. (1978). Development constraints associated with children's use of the keyword method of foreign language vocabulary learning. *Journal of Experimental Child Psychology, 26*, 359-372.
- Salame, P., & Baddeley, A. D. (1982). Disruption of short-term memory by unattended speech: Implications for the structure of working memory. *Journal of Verbal Learning and Verbal Behavior, 21*, 150-164.

- Scruggs, I.E., Mastropieri, M. A., Levin, J.R., McLoone, B., Gaffney, J.S., & Prater, M. (1985). Increasing content area learning: A comparison of mnemonic and visual-spatial direct instruction. *Learning Disabilities Research, 1*, 18-31.
- Scruggs, I.E., Mastropieri, M. A., Levin, J.R., & Gaffney, J.S. (1985). Facilitating the acquisition of science facts in learning disabled students. *American Educational Research Journal, 22*, 575-586.
- Scruggs, I.E., Mastropieri, M. A., & Levin, J.R. (1987). Implications of mnemonic strategy research for theories of learning disabilities. In H.L., Swanson (Ed.), *Memory and learning disabilities: Advances in learning and behavior disabilities* (pp.225-244). Greenwich, CT: JAI Press.
- Scruggs, I.E., Mastropieri, M. A., McLoone, B.B., Levin, J.R., & Morrison, C.R. (1987). Mnemonic facilitation of learning disabled students' memory for expository prose. *Journal of Educational Psychology, 79*(1), 27-34.
- Van Hall, J. G., & Candia Mahn, A. (1997). Keyword mnemonics versus rote rehearsal: Learning concrete and abstract foreign words by experienced and inexperienced learners. *Language Learning, 47*, 507-546.
- Veit, D.T., Scruggs, I.E., & Mastropieri, M. A. (1986). Extended mnemonic instruction with learning disabled students. *Journal of Educational Psychology, 78*, 300-308.
- Wang, A.Y., & Liu, D. (2003). On information and knowledge representation in the rain. In *Proceedings of the second IEEE international conference on cognitive informatics (ICCI'03)*.
- Wang, A.Y., & Thomas, M. H. (1995). Effect of keywords on long-long retention: Help or hindrance? *Journal of Educational Psychology, 87*, 468-475.
- Wang, A.Y. & Thomas, M.H. (1992). *The effect of imagery-based mnemonics on the long-term retention of Chinese characters. Journal of Educational Psychology,*
- Wang, A.Y. et al. (1992). *Keyword mnemonic and retention of second-language vocabulary words. Journal of Educational Psychology*

Appendix A:

Root- Related Keyword List

A list of vocabulary words follow. Each is accompanied by a more familiar word and a sentence to help you remember the new word. Please try to use the familiar word, or "key word," in order to remember the meaning of the vocabulary words.

1- bellicose- warlike

Keyword: rebellion

A rebellion is often much like a war.

2- anachronism- something out of its time keyword. Chronological

Chronological order is in order of time; this is the opposite of anachronism.

3- oration – a formal speech

Keyword: oral

A Formal speech is given orally.

4- incarnate – in human form

Key word: carnivore

A carnivore eats flesh; a human is made of flesh.

5- celerity – speed

Keyword: accelerate

To accelerate is to gain speed.

6- sororal- sisterly

Keyword: sorority

A sorority is a group of women who act as sisters.

7- laudable – praiseworthy

Keyword: applaud

If something is praiseworthy, people often applaud for it.

8- penumbra – partial shade

Keyword: umbrella

An umbrella produces shade.

9- cataclysm – a devastating flood

Keyword: catastrophes

A devastating flood is definitely a catastrophe.

10- pedagogue – a teacher

Keyword: pedal

A teacher is cleaning the pedals of his bicycle.

Appendix B

Production Test

Please fill in as many of the definitions and keyword as you can remember

1- penumbra

2- anachronism

3- oration

4- incarnate

5- celerity

6- sororal

7- laudable

8- bellicose

9- cataclysm

10- pedagogue

Appendix C

Comprehension Test

Please match each letter with the specific number.

- | | |
|---------------------------------|----------------|
| a. warlike | 1- penumbra |
| b. something
out of his time | 2- anachronism |
| c. a formal speech | 3- oration |
| d. in human form | 4- incarnate |
| e. a teacher | 5- celerity |
| f. a devastating shade | 6- sororal |
| g. partial shade | 7- laudable |
| h. praise worthy | 8- bellicose |
| I. sisterly | 9- cataclysm |
| j. speed | 10- pedagogue |

Title

A Process-oriented Evaluation of EFL Writing Tasks in FCE/CAE Preparation Courses versus IELTS/TOEFL Courses of Iran

Authors

Mohammad Reza Hashemi (Ph.D)

Ferdowsi University of Mashhad, Iran

Elham Yazdanmehr (M.A.)

Ferdowsi University of Mashhad, Iran.

Bio Data

Mohammad Reza Hashemi, assistant professor in Applied Linguistics, Department of English, Ferdowsi University of Mashhad, Iran. He is currently teaching translation studies. His areas of interest include translation studies, discourse analysis and CDA.

Elham Yazdanmehr, M. A. in TEFL from Ferdowsi University of Mashhad, Iran. She is currently teaching general English courses at Ferdowsi University of Mashhad. She is also affiliated with Iran Language Institute and International Communication Center. Her areas of interest include EFL/ESL syllabus-design and material-development.

Abstract

This research mainly aimed to analyze EFL writing tasks between two groups of the most popular ESOL exam preparation courses in Iran: IELTS and TOEFL versus FCE and CAE. According to the criteria of writing task appropriateness based on the process-oriented approach to writing instruction, 114 learner participants were asked to rate EFL writing tasks based on a checklist previously developed and validated. An observation process was conducted of the actual task performance in class to help explain

the significant results. According to learners' goals, needs and motivation in Iran, these courses were initially divided in two and through the statistical procedures, the mean scores of learners' ratings for all items of the checklist were compared between these groups. Significant differences were obtained in terms of several features especially related to writing procedures. The results are followed by a number of suggestions with the aim of improving the quality of writing instruction and task management for teachers, syllabus designers or material developers of these exam preparation courses.

Key terms: EFL writing tasks, task analysis, ESOL exam preparation courses, process-oriented approach to writing instruction.

Introduction

There have been numerous approaches to the teaching of writing in the history of language teaching. These have evolved with the development of different approaches to teaching in general, which have in turn contributed to the changing role and status of writing within English language syllabuses and the English as a Foreign Language classroom. In spite of other general methodological changes, however, writing continues to be one of the most difficult areas for the teacher and learner of English to tackle.

Traditionally, writing was viewed mainly as a tool for the practice and reinforcement of certain grammatical or lexical patterns, a rather one-dimensional activity in which accuracy was all important and content and self-expression were trivial. However, with an increase in attention to students' practical needs born out of functional-notional approaches, the significance of writing certain text types as a skill was highlighted (Holmes, 2006). Among various approaches existing in the realm of writing instruction and learning, one of the sharpest contrasts belonged to the product-based versus the process-oriented approach, which forms the basis of this study.

Writing Process Approach and EFL Writing Tasks

Process approaches to writing are contrasted with product and genre approaches, with models and language-based curricula, and controlled, rhetorical and English for Academic Purpose (EAP) approaches. Process writing mainly criticizes the pre-process sentence-level focus and the other major criticism is the product approach characterized

by single-draft think→plan→write linear procedures, with once-off correction, and the use of target product models of writing (Bruton, 2005).

According to Trupe (2001), instructors who incorporate such attention to process in performing writing tasks have the opportunity to intervene in the students' writing process at any stage they are involved in. Students who are asked to spend more time on a writing assignment will think more about their topic, retain more information, and develop more powerful insights.

Various headings have been given to the different stages in the writing process, possibly the most exhaustive being White and Arndt's 'generating ideas, focusing, structuring, drafting, evaluating and re-viewing'. These stages generally involve different forms of brainstorming, selecting and ordering ideas, planning, drafting, redrafting and revising and editing. What follows is a presentation and introduction of the procedures involved which are further complemented by the participants' roles and also the setting where the writing task performance takes place.

Pre-writing

This stage includes anything the writer does before writing a draft of his document, such as thinking, taking notes, talking to others, brainstorming, outlining and gathering information (MIT center, 1999). When students spend time thinking about the writing process, they will be able to plan their strategies more effectively (Purude University Writing Lab, 2007). Sasaki (2000) conducted a research investigating the writing processes adopted by less-skilled and more skilled EFL learners. The results revealed that in pre-writing stage, the expert writers spent a longer time planning a detailed overall organization unlike the novices who spent a shorter time making a less global plan. Furthermore, studies such as the one carried out by Ojima (2006) attested to the fact that concept mapping as a form of pre-task planning was associated positively with the overall quality of the writing product during in-class compositions.

Draft-writing

This stage also called *drafting*, writing the *rough draft* or *first draft* comes when learners get their ideas on paper by organizing them in sentences and paragraphs. Walsh (2004) calls draft a quick write-out where the writers do not worry about the form or mechanics. As described in MIT center (1999), the draft tends to be writer-centered; it is you telling yourself what you know and think about the topic. In case the

writer has had sufficient pre-planning and organization, the drafting stage can be both a gratifying and an efficient experience.

Writers should not feel forced to write chronologically. Sometimes the conclusion can be an easier place to begin with than the thesis statement. With each writing assignment, students will be able to find a personal system that works best for them (Purude University Writing Lab, 2007).

Revising

This is the process of reviewing the paper on the ideal level. This process may involve changes such as the clarification of the thesis, the reorganization of paragraphs, and the omission of the extra information (Purude University Writing Lab, 2007). Much of the recent research into the process writing is monopolized by a focus on revision, whether individual or peer. Elbow (1998), cautions us against the counter-productive effect of premature revising. Frankenberg-Garcia (1999) stands in favor of providing student writers with pre-text feedback, i.e. before the draft is completed. In terms of the positive impact of feedback, Lee & Schallert (2008) argue that establishing a trusting relationship between teacher and students may be fundamental to the effective use of feedback in revision.

Besides the type and amount of feedback that teachers provide in the revising stage, peer feedback can as well be investigated. Peer response/review has been found to help both college and secondary students to obtain deeper insight into their writing and revision processes, develop a sense of ownership of the text, generate more positive attitudes toward writing, enhance audience awareness, and facilitate their second language acquisition and oral fluency development (Min, 2006). Combined with sufficient teacher and peer feedback, the revision process can have great impact on the improvement of student writing.

The results of a study conducted by Paulus (1999) revealed that while the majority of revisions that students made were surface-level revisions, the changes they made as a result of peer and teacher feedback were more often meaning-level changes than those they made on their own. Another study carried out in Chinese context by Miao. et al. (2006) compared teacher and peer feedback in writing revision. Their results showed that more teacher feedback is incorporated and leads to greater improvement, but peer feedback appears to bring about a higher percentage of meaning-change revision.

Editing

After improving the quality of content in the revising stage, writers need to take care of mechanics including corrections of spelling, grammar, punctuation, etc. This is the last step before publishing the final product. It is called *proofreading* as well which deals with “how you write” (MIT center, 1999). Three kinds of feedback can be given to the students in this stage: teacher, peer and self editing feedback. According to Stanley (2003), good writers must learn how to evaluate their own language – to improve through checking their own text, looking for errors, structure (self-editing). This way, students will become better writers.

However, for a beginner student who starts writing essays, self editing is most probably difficult. Rather, the teacher can provide more guidance during editing and/or proofreading with the student to set an example (Ozagac, 2004). If these types of editing are accompanied by peer type, the greatest impact will be produced. With the help of modern technology, for instance, we could think of e-feedback on the oral form-focused revision by friends and peers. The results of such a study conducted by Tuzi (2004) in an academic writing course, suggested that e-feedback had a greater impact than the oral one on revision and it helped L2 writers focus on larger writing blocks.

Publishing

After checking for surface level mistakes besides the content and general organization, it is time to publish the written products. As suggested by Gardner & Johnson (1997), learners could do this by reading out their written pieces loud for the whole class or for their peers in groups or pairs (cited in NCREL, 2004). Part of the advantages is that they can receive feedback on their completed works immediately after producing them. Moreover, the hearers, actually including the peer learners, besides the teacher could point out issues which might be a common source of problems for other learner writers as well. Therefore, it can act to the benefit of not only the writer but also the whole class.

Besides the writing procedures which have just been mentioned, there exist a number of more general and basic components of tasks. According to Nunan (2004) these are divided into goals, input and then the procedures which are supported by roles and settings. To start with, we need to regard the goal and rationale of the task. As suggested by Nunan (2004), goals may relate to a range of general outcomes (communicative, affective or cognitive) or may directly describe teacher’s or learner’s behavior. Among the required qualities of goals, he underlines their clarity to the

teacher and learner, task appropriateness to the proficiency level of learners and the extent to which the task encourages learners to apply classroom learning to the real world. As Jones & Shaw (2003) also pinpoint, writing tasks need to give all learners opportunity to perform to their utmost abilities.

There is also a need to consider the characteristics of task input. According to Nunan (2004), input refers to “the spoken, written and visual data that learners work with in the course of task completion. It can be provided by teacher, textbook or some other source” (p.47). Nunan is in favor of employing combination of authentic material and specially written input. Given the richness and variety of these resources, teachers are enabled to apply authentic written texts that are appropriate to the needs, interests and proficiency level of their students. Whoever provides the input, at any rate, should bear in mind that providing learners with a sample or samples of target language use before starting the task, as Muller (2006) suggests, enables learners to use it as a scaffold to which they can then add their own ideas.

In an attempt to engage learners’ interest, as favored by advocates of process writing approach, the teacher can provide stimulating topics and deploy activities which help the students to express and develop their ideas and to develop tasks where they have a more genuine purpose to write and a stronger sense of the audience for whom they are writing (Holmes, 2006). As suggested by Massi (2001), through making conditions more authentic than the ones in traditional classroom tasks, an awareness of audience, purpose and intentionality will be reinforced. As recommended in the Annenberg Media (2007), in the selection of topics, attempt should be made that they interest learners of their age, sex, educational level, field of study and cultural background. Furthermore, the topic needs to be something about which students have some sort of knowledge. In writing tasks this can be done by choosing tasks that allow learners to capitalize on their prior experience. Teachers can devise class activities that develop and expand students’ schemata (Ferris & Hedgcock, 2005).

The criteria of EFL writing task appropriateness have been collected from both the general components of writing tasks and also writing procedures derived from the process-oriented approach to writing instruction. These criteria were employed to analyze the tasks performed in a number of ESOL exam preparation courses in Iran which are introduced as the following.

The world's leading range of exams for learners and teachers of English are offered by Cambridge ESOL which works with thousands of stakeholders in developing, administering, making and validating many different types of examinations within a consistent but evolving frame of reference (Milanovic, 2009). These exams are taken by over 3 million people in 130 countries. The four most popular of these exams around the world and in Iran are TOEFL (Test Of English as a Foreign Language), IELTS (International Language Testing System), FCE (First Certificate of English) and CAE (Certificate of Advanced English). Significance of tasks and task analysis is even more important for these candidates since the tasks involved in these courses, as suggested by Oxford (2006), are high-stake tasks producing high levels of anxiety on the part of learners'. Therefore, any attempts to analyze and enhance the quality of such instructional courses especially the writing skill is likely to be rewarding.

Conducting an interview with a number of experienced instructors of these courses in Iran illuminated a conspicuous difference between learners' motivation in these four courses. Although these courses are known and recognized worldwide as *exam preparation* courses, in Iran it appears that learners' goals and motivations differ considerably in these courses. Such difference divides the four courses in two major groups: IELTS and TOEFL at one end, where learners' motivation is more extrinsic, and FCE and CAE courses at the other end where students' goal is to improve their English proficiency and their motivation is, therefore, characterized as more intrinsically-oriented. The EFL writing tasks presented and performed in sample classes of these two major groups were analyzed in this study in search for the answer to the following question:

RQ: Is there a difference between the participants' ratings of EFL writing tasks in TOEFL/IELTS classes on one hand and FCE/CAE on the other?

HO: There is no significant difference between the participants' evaluation of EFL writing tasks in IELTS/TOEFL and FCE/CAE classes.

Methodology

Participants

The sample classes of the ESOL exam preparation courses which were attended in this study consisted of: 3 IELTS, 2 TOEFL, 3 FCE and 3 CAE ones. The participants were 114 learners consisting of 30 IELTS, 20 TOEFL, 37 FCE and 27 CAE adult learners of

intermediate and upper-intermediate levels. They included 51 boys and 63 girls whose first language was Persian and were all above 18 years of age. All the sample classes were attended and observed during the summer of 2009 and in five state and private language institutes of Mashhad in Iran.

Instrumentation

The criterion employed in analyzing writing tasks in the observed classes was a checklist of 20 items which were divided into two major sections, *task prompt* and *task procedures*, as can be seen in the Appendix. The first section (containing the first 8 items) dealt with the key general features contributing to the appropriateness of writing tasks. The second section including the remaining 12 items focused on the processes involved in the writing task performance and also addressed the learners' and teacher's roles. All the items were to be rated by choosing between four options: 0, 2, 4 and 6. The participants were not only supposed to do their ratings by selecting among the options, but they were also asked to provide explanatory notes whenever they felt necessary.

Procedures

What we did initially was to collect the criteria of evaluating writing tasks both in terms of more basic and general qualities and from a process-oriented perspective. In the light of the aforementioned review of related literature, the appropriateness features of the target tasks were derived and categorized into two major parts. In order to adhere to the principles of checklist development, the instructions provided by Bichelmeyer (2003), Stufflebeam (2000) and Scriven (2000) were followed. In order to validate the checklist, we followed the steps suggested by Dr. D. L. Stufflebeam at the Evaluation Centre of Western Michigan University (Personal correspondence).

Data Collection

During the whole data gathering process the researcher was present in the target classes. Each session was observed from the beginning to the end. Among the types of observation stated by Denzin & Lincoln (2005), the one adopted in this study was an *unobtrusive* one where the subjects are not aware that they are being studied. Besides, it was of a *descriptive* type where attempt is made to note down all the details by the observer without preconceptions or taking any points for granted.

At the end of the class time, when the task performance was over, the checklists were distributed among participants and a brief instruction was provided on the purpose of the analysis and how they were expected to do the rating. They were also asked to include any further comments wherever they felt it was needed on the related items. Moreover, learners were ensured that their identity would be kept unknown especially to their teacher. Almost all of the ratings were done between 10 to 15 minutes.

After the class time and in some cases before the class started, the instructors were interviewed briefly to pose their impression about the nature of these exam preparation courses, learners' motivation and also distinctive features of the observed classes (if any). This information provided further acquaintance with the participants' motivation in these preparatory courses which could help us in the grouping we made to the primary four groups of exam preparation courses.

Data Analysis

Two-sample t-test was employed in this research to compare the differences in mean scores of ratings. Statistical Package for Social Sciences (SPSS) was utilized for this aim, setting the alpha level at 0.05. The results are reported in the following section.

Results and Discussion

The question investigated in this research addressed EFL writing tasks in the two groups of ESOL exam preparation courses: Group 1 (IELTS/TOEFL) and Group 2 (FCE/CAE). It aimed to see if there existed any significant differences between ratings of writing tasks in these two groups of classes. Table 4-1 demonstrates the statistically significant results obtained from this comparison:

Table 4-1 Students' ratings of EFL writing tasks in Groups 1 and 2

Item	Group 1 and 2 (students)	N.	Mean	SD	t-value	df	Sig.	p
3	G1 (IELTS/TOEFL)	50	1.59	.911	-3.107	112	0	s.
	G2 (FCE/CAE)	64	4.70	1.352				
5	G1 (IELTS/TOEFL)	50	1.76	1.051	-0.816	112	0	s.

	G2 (FCE/CAE)	64	2.57	1.266				
9	G1 (IELTS/TOEFL)	50	4.39	.975	2.610	112	0	s.
	G2 (FCE/CAE)	64	1.78	1.142				
10	G1 (IELTS/TOEFL)	50	4.82	1.149	3.356	112	0	s.
	G2 (FCE/CAE)	64	1.46	1.305				
13	G1 (IELTS/TOEFL)	50	3.80	1.020	2.590	112	0	s.
	G2 (FCE/CAE)	64	1.21	1.166				
14	G1 (IELTS/TOEFL)	50	4.98	1.090	0.535	112	0.012	s.
	G2 (FCE/CAE)	64	4.44	1.104				
15	G1 (IELTS/TOEFL)	50	3.76	.879	2.136	112	0	s.
	G2 (FCE/CAE)	64	1.62	1.128				
16	G1 (IELTS/TOEFL)	50	1.80	.841	0.971	112	0	s.
	G2 (FCE/CAE)	64	.83	.993				
18	G1 (IELTS/TOEFL)	50	5.27	.974	4.567	112	0	s.
	G2 (FCE/CAE)	64	.70	.961				

According to table 4-1, in cases of both *task authenticity* investigated through item 3 (t-value= -3.107, $p \leq .05$) and *topic familiarity* investigated in item 5 (t-value= -0.816, $p \leq .05$) Group 2 gained significantly higher mean scores than Group 1. Item 3 assessed the extent to which the task encouraged learners to apply classroom learning to the outside world. This was closely related to the writing types and topics covered in these

groups. In four of the sample classes of Group 2 (FCE/CAE), formal, informal and application letters were worked on, and in the other class, descriptive writing was the type.

It appears that these have been much more practical than essay writing, graph writing and process writing which were the dominant writing types in the observed IELTS and TOEFL classes. The authenticity of these tasks did also affect the ratings of item 5 which evaluated the extent to which topics were familiar to learners' previous knowledge or experience. The findings again show higher rating for FCE and CAE classes, which implies that learners were more familiar with the topics covered in these courses.

With respect to task procedures, seven items made the major difference and rejected the null-hypothesis. Two of them assessed the quality of the *pre-writing* stage: item 9 (t-value= 2.610, $p < .05$) and 10 (t-value= 3.356, $p < .05$). Participants in Group 1 rated the existence of this stage considerably higher than Group 2. Learners' evaluation of teacher's feedback in this stage is also much lower in Group 2. According to the observations, this stage was entirely present in all the observed classes of the first group, especially TOEFL courses where it was explicitly assigned by the teachers. In FCE and CAE classes, however, this stage was conspicuously absent.

Although some brainstorming was provided by the textbook, the students were neither required by their teachers nor guided how to do them in pairs or groups. This is clearly reflected in the ratings of the 10th item which assessed *teacher's feedback* in prewriting (t-value= 3.356, $p < .05$). Differences are extended to the *revising* stage too, where the students were asked to assess their teacher's feedback on content in item 13 (t-value= 2.590, $p < .05$). Again it turns out that teacher's feedback in this stage is better evaluated in Group 1 than its counterpart. The observation process indicated that in IELTS classes, teachers directly drew students' attention to the register, style and expectations of the target reader before and while they were engaged in writing. In TOEFL classes, teachers used the board as well to guide the students further. Such an effective role was just observed in one class of Group 2, in which the teacher's overall feedback in all stages was vast and effective.

The items addressing *editing* types were also rated significantly different. The three related items, 14 (t-value= 0.535, $p < .05$), 15 (t-value= 2.136, $p < .05$) and 16 (t-value= 0.971, $p < .05$) obtained higher mean scores in Group 1 in comparison to Group 2, implying that students were more engaged in editing for grammar, spelling and

punctuation in IELTS and TOEFL classes than in FCE and CAE, as far as their own ratings show. The main reason could be the students' goal of taking part in these courses. In Group 1, the students had the real exam ahead and were keenly aware of how a poorly-edited writing would affect their overall band score.

The students participating in these courses also rated the *publishing* stage (item 18) much higher than Group 2 (t-value= 4.567, $p \leq .05$). It indicates that in IELTS or TOEFL classes more time is spent on the learners' reading out their texts than it is usually done in FCE and CAE courses. What we observed in these classes (IELTS and TOEFL) was that this stage was present in all of them. In two of these classes, all the students were asked to read out their texts and in the others, two-third of the students read their works out to the class. In FCE and CAE classes, however, this stage was remarkably missing and that makes the greatest difference observed in the writing procedures between these two groups of exam preparation courses.

Conclusion & Implications

The primary conclusion made out of this research was that EFL writing tasks could be evaluated from two aspects: a) the basic general features of appropriateness and b) the quality of writing process and participants' roles in task performance. EFL writing tasks in ESOL exam preparation courses were then compared according to both general features of appropriateness and those related to the writing procedures.

The researcher's personal experience previously as a learner and currently as an instructor of these exam preparation courses, along with the interviews with other experienced teachers, helped to divide the four most popular of these courses into two groups. This grouping was based on the goals, needs and motivation type of the learners in Iran, and resulted in the formation of Group 1 (IELTS/TOEFL) and Group 2 (FCE/CAE). In the first group, learners' goal was to get prepared for the actual exam which was a couple of months ahead. In the second group, however, learners' motivation was far more intrinsic and their primary goal was to improve their general language proficiency. It was intended to see how these two groups of learners evaluated the quality of writing tasks differently in these two types of ESOL exam preparation classes.

The findings of this research revealed that the students' assessment of writing tasks were different in a number of cases. The sharpest contrast, for instance, was found to be

in the *publishing* stage of task fulfillment which was explained according to the observations made of what actually occurred during the task performance. What follows now is a number of suggestions for instructors and syllabus designers who are working in the Iranian context of ESOL exam preparation courses. They are made according to the findings of this research:

Implications for IELTS and TOEFL Courses

- 1- Teachers need to have a closer monitoring on the students engaged in the *editing* stage of writing procedures, for grammar, spelling and punctuation.
- 2- Teachers should take care of the *authenticity* of writing tasks which are performed in class. Although they are bound to work on a limited pre-established topics which will appear in the real exam, they can try to explain and justify why it is needed for the learners to learn how to write this specific type of writing and how it will help them in their prospective life in L2 society.
- 3- Teachers can enhance *topic familiarity* to the learners by eliciting their familiar experiences in their first language to take advantage of the students' background knowledge and experience.

Implications for FCE and CAE Courses

- 1- Teachers should make sure that the task goal is appropriate to the *proficiency level* of the students. They can provide them with useful vocabulary or draw their attention to the grammatical structures that they need to use in the specific type of writing they are expected to produce during that session.
- 2- Teachers need to have a closer monitoring on students engaged in the *revising* stage especially by providing them with feedback on content and also *editing* stage of writing, especially editing for grammar and spelling.
- 3- Teachers should set for a *pre-writing* stage in class. They have to engage the students sufficiently and encourage them to work in pairs or groups to have some sort of brainstorming. They should also teach learners how to organize their ideas and pre-plan their writings.
- 4- Teachers are strongly recommended to ask at least some of the students to read out their written products in order to engage all the class in commenting and draw the students' attention to the possible common mistakes.

What we actually obtained through the conduction of this research can be of great value to whoever engaged in ESOL exam preparation especially as teachers in Iran. Moreover, the checklist can be employed by any EFL writing instructor to evaluate the tasks she or he is assigning to the students in class (not necessarily ESOL preparatory courses) or even in designing new tasks which could be better fitted with the students' needs, proficiency level, available time and other relevant factors.

References

- Annenberg Media, (2007). Teaching foreign language writing. Available Online: http://www.learner.org/workshops/tfl/resources/s3_teachingwriting2.
- Bichelmeyer, B. A. (2003). Checklist for formatting checklists. Available Online: <http://www.wmich.edu/evalctr>.
- Bruton, A. (2005). Process writing and communicative task based instruction: Many common features but more common limitations? *TESL-EJ*, 9(3), 1-31.
- Denzin, N. K., & Lincoln, Y. S. (2005). *The sage handbook of qualitative research*. Sage Publications Inc.
- Elbow, P. (1998). *Writing with power*. US: Oxford University Press.
- Ferris, D., & Hedgcock, J. (2005). *Teaching ESL composition: Purpose, process and practice*. UK: Routledge.
- Frankenberg-Garcia, A. (1999). Providing student writers with pre-text feedback. *ELT Journal*, 53, 100-106.
- Holmes, N. (2006). The use of a process-oriented approach to facilitate the planning and production stages of writing for adult students of English as a foreign or second language. Available Online: <http://www.developingteachers.com>.
- Jones, N., & Shaw, S. D. (2003). Task difficulty in the assessment of writing: Comparing performance across three levels of CELS. *Research Notes*, 11, 11-15.
- Lee, G., & Schallert, D. L. (2008). Meeting in the margins: Effects of the teacher-student relationship on revision processes of EFL college students taking a composition course. *Journal of Second Language Writing*, 17, 165-182.
- Massi, M. P. (2001). Interactive writing in the EFL class: A repertoire of tasks. *The Internet TESL Journal*, 7(6).

- Miao, Y., Badger, R., & Zhen, Y. (2006). A comparative study of peer and teacher feedback in a Chinese EFL writing class. *Journal of Second Language Writing, 15*, 179-200.
- Milanovic, M. (2009). Cambridge ESOL and the CEFR. *Research Notes, 37*, 2-5.
- Min, H. (2006). The effects of trained peer review on EFL students' revision types and writing quality. *Journal of Second Language Writing, 15*, 118-141.
- MIT Online Writing and Communication Centre (1999). The writing process. Available Online: http://web.mit.edu/writing/Writing_Process.
- Muller, T. (2006). Researching the influence of target language on learner task performance. *The Asian EFL Journal Quarterly, 8*(3), Art-8.
- North Central Regional Educational Laboratory (2004). Stages of the writing process. Available Online: <http://www.ncrel.org>.
- Nunan, D. (2004). *Task-based language teaching*. Cambridge: Cambridge University Press.
- Ojima, M. (2006). Concept mapping as pre-task planning: A case study of three Japanese ESL writers. *System, 34*, 566-585.
- Oxford, R. (2006). Task-based language teaching and learning: An overview. *The Asian EFL Journal Quarterly, 8*(3), Art-5.
- Ozagac, O. (2004). Process writing. Available Online: <http://www.buowl.boun.edu.tr>.
- Paulus, T. M. (1999). The effect of peer and teacher feedback on student writing. *Journal of Second Language Writing, 8*(3), 265-289.
- Purude University Writing Lab (2007). The writing process [Microsoft PowerPoint presentation]. Available Online: <http://owl.english.purude.edu>.
- Sasaki, M. (2000). Toward an empirical model of EFL writing processes: An exploratory study. *Journal of Second Language Writing, 9*(3), 256-291.
- Scriven, M. (2000). The logic and methodology of checklists. Available Online: <http://www.wmich.edu/avalctr/checklists>.
- Stanley, G. (2003). Approaches to process writing. Available Online: <http://www.teachingenglish.org.uk>.
- Stufflebeam, D. L. (2000). Guidelines for developing evaluation checklists: the checklist development checklist (CDC). Available Online: <http://www.wmich.edu/avalctr/checklists>.
- Trupe, A. L. (2001). A process approach to writing. Available Online: <http://www.bridgewater.edu/WritingCenter/Resources/Process.htm>.

Tuzi, F. (2004). The impact of e-feedback on the revisions of L2 writers in an academic writing course. *Computers and Compositions*, 21, 217-235.

Walsh, M. (2004). Process writing in high school EFL: What, how and why. Available Online: http://www.geocities.jp/walsh_sensei.

Appendix

Evaluation Checklist of EFL/ESL writing tasks (Students' version)		Ratings			
The course: IELTS <input type="radio"/> TOEFL <input type="radio"/> FCE <input type="radio"/> CAE <input type="radio"/> Age: Sex: MALE <input type="radio"/> FEMALE <input type="radio"/>		Not at all (0)	A Little (2)	Adequately (4)	To a Great Extent (6)
Task Prompt					
1	A. Task goal - Was the overall goal of the task clear and void of ambiguity to you as a learner?	0	2	4	6
2	- Was the task appropriate to your current English proficiency level?	0	2	4	6
3	B. Task authenticity - To what extent did the task help you to apply classroom learning to the real world?	0	2	4	6
4	C. Task topic - Was the topic of the task stimulating and appropriate to your age and educational level?	0	2	4	6
5	- To what extent was the topic familiar to you and related to your background knowledge?	0	2	4	6
6	D. Task instructions - To what extent were the instructions clear and concise?	0	2	4	6
7	- Were the target reader and the features of the expected response (e.g. word/time limits, register) clarified in the instructions?	0	2	4	6
8	- Were any sample texts provided for you either by the teacher or the textbook?	0	2	4	6
Task Procedures					
9	A. Pre-writing - Did you spend time on brainstorming, gathering information or outlining before starting to write?	0	2	4	6
10	- Did the teacher familiarize you with techniques such as listing or clustering the ideas, or ask you to share your ideas in groups?	0	2	4	6
11	B. Draft-writing - Did you go through the second stage of putting ideas into sentences or paragraphs without concern for mechanics such as spelling or punctuation?	0	2	4	6
12	C. Revising - Did you revise your jotted down ideas to make sure of their sensibility and accurateness to the reader?	0	2	4	6
13	- Did you receive feedback on <u>content</u> from the teacher or perhaps a peer in this stage?	0	2	4	6
14	D. Editing - To what extent did you edit your writing for grammar and structure?	0	2	4	6
15	- To what extent did you edit your writing for word spelling?	0	2	4	6
16	- To what extent did you edit your writing for punctuation, before submitting it?	0	2	4	6
17	- Did you receive feedback on <u>form</u> from your teacher in this stage?	0	2	4	6
18	E. Publishing - Did you read out your texts finally to the class or your peers?	0	2	4	6
19	- To what extent was the teacher's feedback on the completed piece of writing motivating?	0	2	4	6
20	* To what extent did the task performance occur outside classroom environment (e.g. in a library or language lab)?	0	2	4	6

Title

Assessing Speaking: Manipulating the facet of the length of prompt in an Oral Proficiency Interview setting

Authors

Masoumeh Akhondi (Ph.D.)

Putra University, Malaysia

Faramarz Aziz Malayeri (Ph.D.)

Putra University, Malaysia

Arshad Abd. Samad (Ph.D.)

Putra University, Malaysia

Bio Data

Masoumeh Akhodi, holds a Ph.D. in TESL from Putra University, Malaysia. Her main research interests are assessing reading and speaking skills and strategies.

Faramarz Aziz Malayeri, holds a Ph.D. in TESL. He has completed a Post Doctoral Research tenure in assessing writing in ESL context. He is particularly interested in quantitative methods, measurement theory and assessment of language skills and test taking strategies.

Arshad Abd. Samad, is an Associate Professor of applied linguistics at the Faculty of Educational Studies, Putra University, Malaysia. He is interested in language testing, theories of grammar instruction, SLA and CALL. Arshad has published books, book chapters, and journal articles in the areas of his interest.

sAbstract

The present research compares the interviewees' performances in two experimental groups receiving different lengths of the same prompt and the third group who started an impromptu interaction with the interviewers. The subjects for the study consisted of 45 Malaysian undergraduate TESL students randomly assigned to three groups. The content of the prompts were adopted from texts on the real world events to contribute to the authenticity of the test task. The prompts were of two different length; 300 and 600 words for the short and the long prompts, respectively. An immediate, retrospective interview was conducted after the test takers read the written stimulus. Quantitative data analyses, applying one-way ANOVA, demonstrated no statistically significant difference between the two groups of prompt-receivers, although the mean of the target vocabulary used from the prompt was higher in the short prompt group as compared to the mean for the long prompt group. Qualitative analyses revealed that giving interviewees the prompts with higher target vocabulary load (long prompt) has noticeably affected the test takers' performances in a way that the prompts, especially the long prompt, provided the fundamental floor for the respondents to ground their speech on and appear as effective communicators. Yet, the means scored by the two experimental groups were significantly different from the mean achieved by the impromptu group.

Keywords: Oral Proficiency Interview, assessing oral proficiency, prompt, target vocabulary, communicative effectiveness, interactional authenticity, situational authenticity, high-stakes test, low-stakes test, field-specific test.

Introduction

The testing of speaking has a long history (Splosky, 2001). Language test developers have attempted assessing speaking skills and strategies in order to present a comprehensive account of language proficiency (Plough et al. 2010; Huff et al. 2008; Iwashita et al. 2008; Brown et al. 2005; Upshur and Turner, 1995, 1999; Chalhoub-Deville, 1995, Weir, 1990). Oral interview was duly introduced by Foreign Service Institute (FSI), associated with the US Government, later known as Oral Proficiency

Interview (OPI). The implemented procedure has been revised frequently since World War II until the critical introduction of the provisional ACTFL Guidelines (American Council on Teaching Foreign Languages) in 1982 (Chalhoub-Deville and Fulcher, 2003). Since then, the interview ratings were based on the ACTFL Proficiency Guidelines (Omaggio, 1986, Swender, 1999). This was soon followed by the introduction of OPI interviews which advocated conducting an interview based on certain theoretical stands and recommended procedure. However, some alternative techniques such as Guided Role-play, Paired Group Activity, Multiple Examiner, and Group-oral Discussion were introduced to compensate for criticisms on psychometric standards of OPI. The other branch of research in OPI has attempted variation of media to conduct the interview, such as delivery over video, telephone or computer, which is aptly called “Simulated OPI or SOPI” (Stansfield and Kenyon, 1992, Shohamy, 1994). The necessity to present a verbal stimulant across these variations of interview is maintained. This verbal stimulant is called “the prompt”.

The presentation of prompt in an interactive discourse is theoretically intended to provide the interlocutor with a stimulant to trigger the intended speech. Butler et al. (2000) emphasize that, “... speech cannot exist in a vacuum; the examinee must receive input in some form, either spoken or written, in order to know what task he/she is being required to perform.” The necessity of presenting prompt is emphasized across the past research in assessing speaking and managing oral proficiency interviews. However, few attempts, if any, have been taken on researching alternatives to short question/statement formats which are currently used as interview prompts (Brown, 2006, Personal communication). The current interviewers have been researched widely in a variety of behavioral aspects.

These behavioral aspects are believed to be threatening the validity of interview among which the ways the interviewers construct the prompt or pose their questions have been recognized validity threatening (Lazaraton, 1996; Ross, 1996; Brown and Lumely, 1997; Brown, 2003). The nature of prompt could be considered as one of the most important facets in any variation of oral proficiency interview. Moreover, any attempt to confine the current language proficiency interviews under a promising theoretical framework needs to account for the role of prompt. However, the available models clearly lack it. The most famous language proficiency interviews, particularly the IELTS interview, are still emphasizing on presenting short prompts composed of questions or statements to stimulate the intended speech.

Statement of the problem

The available literature on assessing the speaking skill reveals that there is a growing interest in researching how speaking might best be assessed. A number of studies have focused on investigating the validity of scores and rating scales (Meredith, 1990; Halleck, 1992; Reed, 1992), the rater's behavior and their performance (Barnwell, 1989; Thompson, 1995), the interviewer's variation (Brown, 2003), and the interviewee's individual differences, e.g. proficiency level (Young, 1995), the ability to self-adjust (Ross, 1992; Cafarella, 1994) and differences in linguistic backgrounds (Young and Halleck, 1998). While there is now a substantial body of research into the discourse produced within oral language interviews, the majority of analyses are concerned with characterizing interviewer and candidate discourse in general (Brown, 2003).

None of the previous studies are concerned specifically with the length of input and its impact on the candidate's use of target lexicon and syntax provided in the input (Brown, 2006, personal communication). In the present study, attempts are made to discover the role of the length of input in interviewee's speaking performance. In particular, we have examined how the length of the input affects the interviewee's use of the presented target vocabulary and the grammatical structures (prompt contents) in his/her speech. We also have investigated the degree to which the length of input contributes to the interviewee's overall speaking performance. This is assumed that if it could be shown that the length of inputs make any contribution to the interviewees' overall performance, this might have important implications for high-stakes decision making interviews as well as low-stakes proficiency evaluations.

Review of Literature

The oral interviews, in which examiner and candidate take part in an unscripted discussion of general topics, have long been a popular method for assessing second language oral proficiency (Brown, 2003). Ross (1996) believes that they incorporate features of non-test or conversational interaction. Further, they are claimed to have allowed second language learners demonstrate their capacity to interact in an authentic communicative event utilizing different components of communicative competence extemporaneously. However, the two most famous oral proficiency interviews, i.e., the IELTS interview and the TSE (test of speaking proficiency), are still using the

prompted interviews as their mainstream interview procedure. The IELTS speaking interview module consists of 5 phases, of which the first and the last –the opening and the closing– are very short. Other than the middle phase, which consists of a role-play, the speaking module is basically a conversational interview in which candidates are invited to talk on a range of topics covering a variety of functional and discursial skills (such as description, argumentation, narration and speculation). In the main phase of the interview, the candidate receives a topic to take note on that. This phase was designed to give candidate the initiative from the start, to encourage him become an active participant in the conversational exchange rather than providing him with minimal responses to a series of questions (Brown, 2003). For the purpose of the current study, the researchers have followed a more or less similar procedure conducting the interviews.

However, concerns have been voiced on numerous occasions that candidates may produce different performances, hence, receive different ratings in different interviews (Wilds, 1975; Shohamy, 1983; Bachman, 1988, 1990; Stansfield, 1990; McNamara, 1996). Stansfield and Kenyon (1992) for example claim that because the interviewer is free to select the topic and ask whatever he/she wishes to, the same candidate may give two different performances with two different interviewers. Lazaraton (1996) warns of the danger of such variation to test fairness:

...the achievement of consistent ratings is highly dependent on the achievement of consistent examiner conduct during the procedure, since we cannot ensure that all candidates are given the same number and kinds of opportunities to display their abilities unless oral examiners conduct themselves in similar, prescribed ways (p.19).

The pioneers in this field have increasingly shed light on the issue of interviewer variation, fairness and rating scales in oral proficiency interviews (Lazaraton, 1996; Bachman, 2001; Fulcher, 1996). They also have elaborated on the use of the prescribed topics and questions in order to achieve more valid and comparable results for all the participants (Stansfield and Kenyon, 1992). In the present study, the researchers selected a single predetermined topic and a relevant set of questions. The only difference between the two versions of the text was set on the length of the input (the prompt) that the two experimental groups received. The prompts had similar topic, but

one group received the long prompt while the other group received it in the short form. These kinds of context-based interviews are assumed to be more useful than the general-purpose tests (Douglas, 1997). The implication of this kind of interview is recommended in high-stakes decision making and low-stakes proficiency evaluation. Current research believe that inconsistencies in topics and text structures (descriptions, argumentative, narrative, and speculation), and questions to be asked will be threats to the validity and fairness of the interview test tasks (Brown et al., 2005; Bachman, 2001; Douglas, 1997). Therefore, when groups of candidates attend the interviews for the purposes of employment, academic admissions, and so many other high- and low-stakes decision-makings and different interviewees receive different topics and text structures to discuss; their performance cannot be comparable. This will violate the principle of test fairness. While some test takers who are less competent may benefit from a descriptive topic of discussion, more competent candidates may struggle to come by with the argumentative or informative context.

The other famous measure of oral language proficiency is the Test of Spoken English (TSE) by ETS (Educational Testing Service). The Test of Spoken English (TSE) is a test of speaking ability designed to evaluate the oral language proficiency of non-native speakers of English who are at or beyond the post secondary level of education (Douglas and Smith, 1997). The TSE was developed on the promise that language is a dynamic vehicle for communication driven by underlying competencies, which interact in various ways so that communication can be effective. The revised TSE is designed for use as a general screening test for teaching and research assistant selection and professional certification (Butler et al., 2000).

From the debut of TSE in 1979 until the revision of TSE in 1992, there has been some kinds of criticism upon a number of areas: task difficulty of speaking into a tape recorder (Bailey, 1987; Plakans and Abraham, 1990), the unnaturalness of having only one chance to hear and respond to questions (Plakans and Abraham, 1990), the lack of relationship of the content (e.g., describing a bicycle) to academic and professional work (Bailey, 1987; Byrd, 1987; Gallego et al., 1991), the lack of attention to the measurement of listening skills, interpersonal skills, and professional expertise (Plakans and Abraham, 1990), the difficulty of using a “native speaker” standard as a scoring template (Barrett, 1987; Godfrey and Hoekje, 1990), and examinees coming from different language background may have different mean scores (Barrett, 1987; Costantino, 1988). In the light of this discussion, then, a framework for revising the

TSE ought to include a thorough specification of the construct of speaking ability in non-test situations and a more thorough specification of the scoring rubric and level descriptors to include a wide range of competencies. The TSE has been revised so that it would better reflect current views of language acquisition and testing, specifically modern notions about communicative competence (Powers et al., 1999). During the revision procedure, three of the seven original sections of the test including read aloud, complete sentences, and answering questions about a single picture were deleted.

They were considered to be the least communicatively oriented portions of the test and, thus, less widely accepted as measuring oral proficiency. Recently, a narrative task based on a sequenced set of picture prompts is used in the TSE as a measure of spoken language ability (Elder, Iwashita, and McNamara, 2002). Using the same kind of task (picture narration) could eliminate one of the sources of error in oral proficiency testing to a large extent. However, Bachman (1990) identifies some sources of error in oral proficiency testing in inconsistency of tasks. He has highlighted the needs for tests to be consistent across examinees. He argues that, because each OPI is conducted by a different interviewer with different examinees, the potential for inconsistency is great. Lazaraton (1997) has explained that, “it is the question *types* and *tasks* that are standardized in OPI, rather than specific questions or specific topics.” Therefore, the potential for low levels of interrater reliability is high.

Considering the above mentioned justifications, in the present study, the researchers came to an agreement to use a single topic of discussion in the prompts (input), but in two different lengths. Further, a set of predetermined and consistent questions were asked. This is assumed to promote the interrater reliability and construct validity of the interview task designed. This consistency in topic and questions would make the interview results comparable across the three different groups of respondents.

The Study

This study investigated the length of the linguistic input, aptly called the prompt, on interviewees’ responses in terms of the target vocabulary captured from the written prompt (input) as the main objective. The other objective was to evaluate the quality of the produced speech with respect to the number of employed structures from the offered prompts by each specific group of interviewees. In this study, there were 45 Malaysian undergraduate TESL students who were randomly assigned into three groups, the two experimental groups that received the prompts with the same topic in

two different lengths and the control group that attended the interview impromptu. The researchers looked at the facilitative effects of the enriched and long prompt on the candidates' performances in oral proficiency interview from two different dimensions of task-based assessment and factors influencing task difficulty. Most of the researchers in the field of task-based assessment came to an agreement that four dimensions (immediacy, adequacy, planning time, and perspective) were the ones that lent themselves most readily to experimental manipulation (Elder, Iwashita, and McNamara, 2002). In the current study, three of these four dimensions (immediacy, adequacy, and planning time) were taken into consideration for further analyses.

The 'immediacy' dimension could be easily manipulated by asking the candidates to discuss the topic after the prompt had been removed, or to discuss the topic having some notes from the prompt in front of them. In the present research, the researchers agreed on the lack of immediacy of the prompts in order to investigate the amount of the information and the number of target vocabulary which could be recalled from the prompts. Moreover, it was hypothesized that letting the candidates rely on their own memory, world knowledge, and language knowledge would give a better image of test takers' language ability. In this case, the prompts were used just as a stimulus to activate their schemata on that real-life event, and then it is the test taker's communicative competence, strategic competence and language ability which will play the most important roles in interviewee's effective communication. Likewise, comparable groups of candidates could discuss the topic with and without preparation time (planning dimension).

In this research, the two experimental groups had 10 minutes to read the written prompt and prepare for the interview while the candidates in the control group discuss the topic impromptu, i.e., without preparation. The 'adequacy' dimension had been manipulated in this study by giving the long prompt (about 600 words, more cognitively demanding) to one of the experimental groups and the short prompt (about 300 words, less cognitively demanding) to the other treatment group. Manipulation of this dimension of the task was the main focus of the study in order to investigate the extent to which the enriched linguistic input (long prompt) affect the candidates' performances in an oral proficiency interview.

The content of the prompt was about "*September 11, 2001*". It was adopted from the Wikipedia Encyclopedia (http://en.wikipedia.org/wiki/September_11,_2001_attacks), on the assumption that the participants will be able to interact with the authentic text

more efficiently compared to an abstract or non-familiar text.

Bachman and Palmer (1996) suggest that authenticity has a potential effect on test takers' perceptions of the test and on their performance. Bachman (1991) believes in two different kinds of authenticity; situational authenticity that is, the perceived match between the characteristics of test tasks to Target Language Use (TLU) tasks, and the interactional authenticity, which is the interaction between the test taker and the test task.

The prompts used in the current research were authentic as they described an event which had happened in the real-life context (situational authenticity). As authenticity and in turn interactivity are not absolute notions they can be considered relative (Bachman, 1991). In this case, some test takers were able to interact with the text while some other failed. Therefore, it cannot be asserted that the text was fully interactive and/or authentic from the perspective of the test takers' performance.

Widdowson (1978) has called the feature of authenticity of a test its "functionality". He believes that the degree of functionality depends on the interaction between the test taker and the task material. In case of this study, the given test task was observed functional for some test takers and not functional for some of the participants. This aspect of authenticity or functionality has been supported by the achieved results as well as the interviewees' feedback. Most of the test takers found it easy to talk about such socially familiar event since they have already had some background on the topic, while some other interviewees were not able to fully interact with the text, supposedly, due to the lack of interest in such topics.

However, these might be the individual differences and interests which interfere with the interactivity dimension of the task. Lewkowics (2000) looks at authenticity from another angle. He believes that the texts which are extracted from 'real-life' sources are potentially more authentic than the texts written for the pedagogical purposes. In case of the current research, the content of the prompts were adopted from a frequently visited and revised live text, the online encyclopedia, in order to observe this aspect of authenticity. However, the issue of authenticity will remain controversial as the degree to which a text may stretch ties with reality is relative and text-type bound. It becomes even more complicated once the extension of functionality is assumed as individuals' degree of success of failure of negotiating meaning with the information elaborated on in the text. It must not be neglected that reality is constructed by the reader through his/her repeated exposure to particular text types and matching and re-matching his

perceptions with the text information structure. Therefore, it remains almost difficult, if not impossible, to ascertain reader has constructed authenticity while interacting with a short or long prompt employed for the purposes similar to the ones in this piece of research.

Methodology

Overview

Malone (2003), reviewing the studies in investigating aspects of oral language proficiency, argues that the studies which included quantitative analysis lacked qualitative analysis, whereas, the studies that included fewer numbers of examinees included more qualitative research. She continues to argue that both qualitative and quantitative data are necessary for the type of analysis needed to shed light on the reliability and validity of the OPI. In the current research, both quantitative and qualitative methods were used in a mixed-method approach for more valid results and discussions. The present study employed an experimental research design, specifically the randomized subjects, posttest-only control group design (Ary, Jacobs, and Razavieh, 2002, p. 305). Three randomized groups of respondents participated in this study, two experimental groups and one control group. In the two experimental groups, the participants took part in an immediate retrospective oral interview after reading the prompts. Quantitative data were analyzed using One-way Analysis Of Variance (ANOVA), while qualitative data were analyzed through content analysis (transcribed interview content analysis), and conversation analysis techniques.

Participants

A total of 45 Malaysian undergraduate TESL students participated in the study, the majority being 22-23 years old, on the verge of completing their Bachelor's degrees, all of whom had previously passed a course in English Public Speaking Skills. The students in the intact class were randomly divided into three groups, in which one of the experimental groups received the long prompt while the other treatment group received the short prompt and the participants in the control group attended the interview impromptu.

Materials

For the interview materials, two prompts sharing the same topic but different in length and number of target vocabulary were adopted from the Wikipedia encyclopedia and were tailored into long and short prompts for the purposes of the current research. The readability indices for the two prompts were investigated using Type Token Ratio and Flesch Reading ease (below 60) which is considered as the cutoff score for the university level students. It was found that they were approximately the same. For the next step, a set of questions were prepared based on the prompts. All of the candidates responded to the same questions in order to remove the effect of variability in tasks and interview questions. A checklist of the target vocabulary of the prompts was prepared in order to examine those target vocabulary used (borrowed) by the candidates from the prompts. The interviews were tape recorded for further probe and analysis. At the same time, field-notes were taken by the rater while the interviewers were conducting the interviews.

Statistical analysis

In the present research, the number of target vocabulary used by each participant had been marked in his/her checklist. The total number of the target vocabulary used across the interaction was adopted as the vocabulary performance score achieved by the respondent. This is in line with Brown and Yule (1983) who contend: in a task-based assessment the score achieved by the examinee on a particular task is expressed in terms of the positive number of points of required information which were expressed. Moreover, a one-way ANOVA was conducted to test the research hypotheses with the significance level set at $P < .05$.

Results and Discussion

This study investigates the effect of the length of linguistic input, aptly called the prompt, on interviewees' responses in terms of the target vocabulary used from the prompt and the quality of responses produced with respect to the sentence structures observed. The results of the study are presented in both qualitative description of the overall oral performance of the candidates and quantitative analysis of the target vocabulary used by the participants during the interview.

The first objective of the study focuses on the qualitative differences in the language produced by the three groups receiving long, short, and no prompt.

The first research objective was translated into the following research question:

-What are some qualitative differences in performances of the three groups of the participants with respect to the amount and length of the provided input?

The analysis of the transcribed interview contents revealed that the linguistic input provided prior to the testing context [interview] had an impact on candidates' performances. The results of the study for the first objective for the three groups of the interviewees are described in order

Control group

It was observed that the majority of the participants in this group reacted with hesitation, looking for ideas to grasp in order to initiate the conversation.

I: Interviewer, **R:** Respondent

Change of the floor

I: Who claimed the responsibility for the event? I mean for the hijacks and the crashes and..[interrupted]

R: That was...eem, just after those bombing in the Middle Eastern countries, so...[hesitates]... they have decided to attack America, they

I: Who claimed the responsibility?

Topic Nomination

R:.... these things are horrible, seem, [laughs] to me, that was not like it when they have no economy, they face joblessness that is a problem with their economy [new topic], they need jobs and..

I: But the event was drastic in its scope wasn't it?

R: Yes, almost everybody heard about it. We have had the experience here, the economy was affected....

Sometimes they felt embarrassed and trapped responding the questions. Among the communicative strategies used by the interviewees in this group, topic avoidance was the most common one. The interviewers repeatedly encountered the interviewees' interests to turn the floor to a more favorable one. However, this topic nomination was

managed and conquered by the interviewers. It was realized that in case of no prompt in hand, the interviewees started describing the event rather than reasoning and arguing the pros and cons. This finding of the study has strongly verified Douglas' (1997) belief that "language cannot be tested in vacuum, and there must be some stimulus whether in the form of written or listening input." This finding of the study is also in the line with the findings of a study conducted in Hong Kong on impromptu public speaking. Hsieh (2006) in Hong Kong investigated the problems which students regularly encounter in *English Impromptu Speech Contest*. She found that the students who did not have enough topical knowledge and vocabulary resources were unable to deal with an unfamiliar topic of discussion; therefore, they decided to give up at the early stages of the contest.

Long prompt group

The participants in this group received a long prompt (about 600 words) pertaining to the topic prior to the interview. Qualitative analysis of the data revealed that the most proficient speakers in this group took the most advantage of the prompt. It is worth pointing out that there are two types of knowledge, as Anderson (1976) believes: declarative knowledge that is knowledge of facts or about things, and procedural knowledge that is knowledge about how to do things. Here the content of the prompt gave the respondents (this particular group of participants) a full image of the fact (declarative knowledge enhanced). However, it is the test taker's language ability and communication strategies (communicative competence) which play the most crucial role in interviewee's output as an effective communicator (procedural knowledge). The long prompt group qualitatively performed better compared to the other experimental group whom received the short prompt. The researchers believe that it is the enriched content of the prompt which made the respondents of this group appear effective and fluent communicators. Their activated world knowledge along with the genuine information from the prompt helped them overcome the inefficiencies of their knowledge; therefore, it led to producing a coherent and organized chunk of speech without any communication stress.

About 46.6% of the respondents in this group neglected the exact content of the prompt including the intended target vocabulary. They just managed to realize the overall idea from the prompt to negotiate meaning in their own ways. About 43.3% of the participants in this group kept using the content of the prompt to a large extent, but

in dealing with the target vocabulary of the prompt they preferred to use their own synonyms. It was realized that only 10% of the interviewees were able (or interested in) to memorize everything and adopt the exact target vocabulary from the prompt. However, the researchers are encouraged to believe in the effect of the long prompt on the quality of the test takers' overall performance; even though, they did not use much of the target vocabulary; it helped them to conquer their communication stress and interview fright as in:

I: Interviewer, **R:** Respondent

[Effective communicators]

I: Were they in the Bush's administration responsible?

R: ..., yes that's what they have provided the social backgrounds for. The America was responsible in creating this catastrophe for the people of the US on top of that, they are, either in economy or in their policies, they must be held responsible.

I: were there huge economic effects observed?

R: what do you expect out of dust they have blown themselves in and around the world? The Pentagon, the CIA all the policies they have introduced ... economy is sure affected....

Short prompt group

This group of interviewees received a short prompt (about 300 words) pertaining to the topic prior to the interview. About 30% of these candidates used the exact target vocabulary from the prompt to a large extent and this led to the production of a smooth flow of speech with no hesitancy. Another 30% of these respondents were able to discuss the pros and cons and create an organized chunk of speech. About 40% of the interviewees in this group applied 15% of the target vocabulary from the prompt. They described the event independent of the exact content of the prompt. They grasped an overall idea from the prompt and reproduced it in their own words and structures.

Although the grammatical structure of the prompts was not the focus of this research, it was observed that they were quite independent of the time/tenses used in the prompts. The most common used tense among the interviewees' speech was Simple Past Tense and very rarely it was in Past Perfect Tense. However, it was realized that the interviewees preferred to ignore the variety of the tenses introduced in the prompts. It

could be interpreted that they just focused on the negotiation of their intended meaning rather than focusing on the structures and forms which they were provided with. However, the load of borrowed vocabulary was bigger than the load of grammatical structures retrieved from the prompts across the produced speech.

The qualitative analysis revealed that most of the short prompt group respondents were fluent speakers so that they could easily recall the target vocabulary from the prompt to a large extent. Therefore, they seemed to have no hesitation to look for the appropriate vocabulary to fill in their speech gaps. Unlike the short prompt group, the long prompt group, due to the load of the vocabulary and the limitation of time, were not able to easily recall and apply the exact target vocabulary from the prompt. It can be claimed that the performance of the respondents in the long prompt group was qualitatively better compared to the short prompt group participants due to the density of the information included in the long prompt. The long text must have provided a comprehensive discourse context which led to a better speech production. Therefore, the length of prompt has crucial effects on interviewees' performance. It provides them with enough linguistic resources to be employed in the oral proficiency interview.

Quantitative analysis of the data

Descriptive statistics and frequency analysis

The second objective of the study which was quantitatively pursued addresses the following comprehensive research question:

-Are there any significant differences among groups receiving long, short and no prompts in using the target vocabulary from the prompts?

Table 1 illustrates the frequency analysis of the interviewees' use of the target vocabulary by the two experimental groups and the control group.

Table 1: Frequency Analysis of the Target Vocabulary used by the three groups

Groups	Mean	Standard deviation	Minimum	Maximum

1.00	7.80	4.27	2.00	17.00
2.00	8.60	2.47	5.00	12.00
3.00	4.13	1.64	2.00	7.00
Total	6.84	3.53	2.00	17.00

Groups 1: long prompt; 2: short prompt; 3: no prompt (control group)

Here, for group 1 (long prompt group), the relatively higher mean of 7.8 signifies that the participants in this group were dependent on the content of the prompt to a large extent. Simultaneously, it is noteworthy to remind that the relatively large standard deviation (SD=4.27) could be due to the wide range of the target vocabulary used by some of the participants in this group from the prompt.

Group 2 (short prompt group) achieved the highest mean (8.6) among the three groups and a smaller standard deviation (SD=2.47) compared to the long prompt group. Results of the short prompt group verified the relatively high intensity of the rate of the borrowed target vocabulary by the respondents of this group from the prompt. The smaller standard deviation of this group compared to the long prompt group was due to the consistency and homogeneity of the number of target vocabulary used by the participants of this group. Based on the results of these two groups, it seems a plausible justification that the number of the target vocabulary borrowed from the prompt is a function of the length of the prompt along with the amount of time allocated to accomplish the task of reading the prompt.

The lowest observed mean (4.13) and standard deviation (1.64) belong to the control group, group 3 that received no prompt. This finding of the study once again verified Douglas' (1997) statement that language cannot be tested in vacuum and there must be some sort of input whether in the form of a written prompt or a listening stimulus. Most of the participants in this group felt trapped and embarrassed. They were not able to interact with the interviewer easily and smoothly, especially those who were not interested in the topic of the interaction. Nevertheless, it is significant to point out that, there were few respondents who were capable of handling the task easily due to their own topical knowledge. Although, they did not use many of the words which had been

considered as target in their speech, their overall performance was at an acceptable level of proficiency.

Inferential statistics

The inferential statistics test used for the second objective to compare means across the three groups of the interviewees was the one-way Analysis of Variance (ANOVA). Table 2 and Table 3 represent the summary ANOVA table and the post hoc (Multiple comparisons) table respectively.

Table 2: ANOVA Table

		Sum of Squares	df	Mean Square	F	Sig.
Target vocabulary	Between	170.178	2	85.089	9.411	.000
Groups used from prompt						
* length of prompt	Within	379.733	42	9.041		
Groups		549.911	44			
Total						

Table 3 illustrates the summary of the Post Hoc Comparison (Tukey’s HSD).

Table 3: Multiple Comparisons (Post Hoc)

(I)length of Prompt	(J)length of Prompt	Mean Difference (I-J)	Std. Error	Sig.	95% confidence Interval	
					Lower Bound	Upper Bound
1.00	2.00	-.80000	1.09795	.748	-3.4675	1.8675
	3.00	3.66667*	1.09795	.005	.9992	6.3341
2.00	1.00	.80000	1.09795	.748	-1.8675	3.4675
	3.00	4.46667*	1.09795	.001	1.7992	7.1341
3.00	1.00	-3.66667*	1.09795	.005	-6.3341	-.9992
	2.00	-4.46667*	1.09795	.001	-7.1341	-1.7992

* The mean difference is significant at the .05 level.

The results in summary ANOVA table indicate the rejection of the null hypothesis due to the large F-value (9.411) and small P-value (.0001). So far, it is found that there are

differences among the three groups in the number of target vocabulary used from the prompt. After a post hoc analysis was done, the pairs which were significantly different were determined. The results from Multiple Comparisons (post hoc) revealed that the two experimental groups were significantly different from the control group, but the difference between the treatment groups (long prompt and short prompt groups) was not statistically significant. Although, the short prompt group achieved a higher mean of the target vocabulary used from the prompt and they performed quantitatively better than the two other groups, the participants in the long prompt group were qualitatively more capable of handling the questions and taking turns during the interviews.

Conclusion

The overall results of the study revealed that giving the interviewees prompts pertaining to the topic prior to the interview is significantly helpful in producing cohesive and organized chunks of speech. The findings suggest that the longer prompt provides more fruitful '*floor*' for the participants to ground their discussions and responses on. The participants who received the long prompt enriched with more target vocabulary and information made higher '*ceiling*' for their speech production. In this case, they were able to answer the majority of the questions posed by the interviewers. Although the long prompt group participants were not able to recall exact target vocabulary included due to the limited time for preparation, they were capable of describing the event in their own words and structures based on the gist of the material which they extracted from the received prompt. However, the optimal length of prompt for the interview purposes as such remains open to further studies. The multiple dimensions of conducting oral proficiency interview, time management (practicality), offering efficient prompts (construct relevant variation), raters' inter/intra reliability, as well as authenticity and interactivity requires a wise manipulation of interview prompt.

Implications and suggestions for further research

Building on Douglas' (1997) statement that context-based and field-specific test may be more useful than general purpose test for making situation-specific judgments about language ability; the researchers argue that it would be a more efficient strategy to make the high-stakes interviews field-specific. We continue our argumentation that

providing the candidates a complete, context-specific prompts prior to the interview would lead to making more accurate decision on test taker's language ability. Moreover, it seems inaccurate that everybody attending interviews for a variety of purposes possesses a genuine general knowledge about everything. In some cases, the test taker has the potential communicative language ability, but due to the lack of topical knowledge may fail in an employment interview. Another application of giving prompts based on the findings of the current research is in the reduction of the test anxiety and communicative stress. As it has been supported by the findings of the study, the respondents of the control group who were not aware of the topic of discussion looked more embarrassed and made more false starts due to their hesitation. Changing the text structure from narratives to other formats (expository, description, problem-solution), accompanying pictorial devices with the written prompt (diagrams, scientific graphic presentations), and manipulating the preparation time for different prompt lengths would cast more light on the issue.

References

- Anderson, J. (1976). *Psycholinguistic experiments in foreign language testing*. Queensland: University of Queensland Press.
- Ary, D., Jacobs, L.C., and Razavieh, A. (2002). *Introduction to Research in Education*. Wadsworth: Thomson Learning.
- Bachman, L.F. (1988). Problems in examining the validity of the ACTFL oral proficiency interview. *Studies in Second Language Acquisition*, 10(2), 149-64.
- 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: 671-704.
- Bachman, L.F., and Palmer, A.S. (1996). *Language Testing in Practice*. Oxford: Oxford University Press.
- Bachman, L.F. (2001). Some construct validity issues in interpreting scores from performance assessments of language ability. In *New Perspectives and Issues in Educational Language Policy*, Cooper, R. L., E. Shohamy and J. Walters (eds.), 63-90.

- Bailey, K.M. (1987). Test of Spoken English. In J.C. Alderson, K.J. Krahnke, & C.W. Stansfield (eds.), *Reviews of English language proficiency tests*, (pp. 84-86). Washington DC.: TESOL.
- Brown, A. (2006). Personal communication through e-mail, April 24, 2006. Princeton. New Jersey: Educational Testing Service.
- Byrd, P. (1987). Being seduced by face validity: linguistic and administrative issues in videotaped simulation testing. In N.V.N. Chism & S.B. Warner (eds.), *Institutional responsibilities and responses in the employment and education of teaching assistants: Readings from a national conference*, (pp. 355-357). Columbus, Ohio: The Ohio State University Center for Teaching Excellence.
- Barrett, R.P. (1987). The SPEAK Test: some comments by a former user. *NAFSA Newsletter*, 38, 7.
- Barnwell, D. (1989). 'Naïve' native speakers and judgments of oral proficiency in Spanish. *Language Testing*, 6, 152-63.
- Brown, A., Iwashita, N., and McNamara, T. (2005). An Examination of Rater Orientations and Test Taker Performance on English for Academic Purposes Speaking Tasks. (*Monograph Series MS-29*). Princeton, NJ: Educational Testing Service.
- Brown, A. (2003). Interviewer Variation and the Co-construction of Speaking Proficiency. *Language Testing*, 20(1), 1-25.
- Brown, A. and Lumley, T. (1997): Interviewer variability in specific-purpose language performance tests. In Kohonen, V., Huhta, A., Kurki-Suonio, L., and Luoma, S. *Current developments and alternatives in language assessment: Proceedings of LTRC 96*, (pp. 137-50) University of Tampere.
- Brown, G. and Yule, G. (1983). *Teaching the Spoken Language: An approach based on the analysis of conversational English*. Cambridge: Cambridge University Press.
- Butler, F.A., Eignor, D., Jones, S., McNamara, T.F., and Suomi, B.K. (2000). *TOEFL 2000 Speaking Framework: A Working Paper*. Princeton. New Jersey: Educational Testing Service.
- Cafarella, C. (1994). Assessor accommodation in the V.C.E. Italian oral test. *Australian Review of Applied Linguistics*, 20, 21-41 .
- Chalhoub-Deville, M. and Fulcher, G. (2003). The Oral Proficiency Interview: A Research Agenda. *Foreign Language Annals*, 36(4).
- Chalhoub-Deville, M. (1995). Deriving oral assessment scales across different tasks and rater groups. *Language Testing*, 12, 16–33.
- Costantino, M. (1988). *Performance of international students on the SPEAK Test*.

- Unpublished manuscript. University of Pennsylvania, Philadelphia, PA.
- Douglas, D. (1997). *Testing Speaking Ability in Academic Contexts: Theoretical Considerations*. Princeton, New Jersey: Educational Testing Service.
- Douglas, D. and Smith, J. (1997). *Theoretical Underpinnings of the Test of Spoken English Revision Project*. Princeton, New Jersey: Educational Testing Service.
- Elder, K., Iwashita, N., & McNamara, T. (2002). Estimating difficulty of oral proficiency tasks: what does the test-taker have to offer? *Language Testing*, 19(4), 347-368.
- Fulcher, G. (1996). Invalidating Validity Claims for the ACTFL Oral Rating Scale. *System*, 24, 163-72.
- Gallego, J.C., Goodwin, J., & Turner, J. (1991). ITA oral assessment: The examinees perspective. In J.D. Nyquist, R.D. Abbott, D.H. Wulff, & J. Sprague (eds.), *Preparing the professoriate of tomorrow to teach: Selected readings in TA training*, (pp. 404-412).
- Godfrey, D. & Hoekje, D. (1990). *The use of native speaker norms in evaluating nonnative speaker oral proficiency*. Unpublished manuscript, West Chester University.
- Iwashita, N., Brown, A., McNamara, T.F., and O'Hagan, S. (2008). Assessed levels of second language speaking proficiency: How distinct? *Applied Linguistics*, 29, 24-49.
- Halleck, G.B. (1992). The oral proficiency interview: Discrete point test or a measure of communicative language ability? *Foreign Language Annals*, 25(3), 227-31.
- Hsieh, S.M. (2006). Problems in preparing for the English impromptu speech contest. *RELC Journal*, 37(2), 216-235.
- Huff, K., Powers, D.E., Kantor, R.N., Mollaun, P., Nissan, S., and Schedl, M. (2008). Prototyping a new test. In Chapelle, C., Enright, M., and Jamieson, J. (eds.), *Building a validity argument for the Test of English as a Foreign Language*. New York and London: Routledge.
- Lazaraton, A. (1996). Interlocutor support in oral proficiency interviews: The case of CASE. *Language Testing*, 13(2), 151-72.
- Lazaraton, A. (1997). Performance organization in oral proficiency interviews: The case of language ability assessment. *Research on Language and Social Interaction*, 30, 53-72.
- Malone M. E. (2003). Research on the Oral Proficiency Interview: Analysis, Synthesis, and Future Directions. *Foreign Language Annals*, 4, 491-497.

- McNamara, T.F. (1996). *Measuring Second Language Performance*. London and New York: Longman.
- Meredith, R.A. (1990). The oral proficiency interview in real life: Sharpening the scale. *Modern Language Journal*, 74(3), 288-96.
- Omaggio, A. C. (1986). *Teaching language in context*. Boston, MA: Heinle & Heinle.
- Plakans, B., and Abraham, R. (1990). The testing and evaluation of international teaching assistants. In Douglas, D. (ed.), *English Language Testing in U.S. Colleges and Universities*, (pp. 68-81). Washington, DC: National Association for Foreign Student Affairs (NAFSA).
- Plough, C.I., Briggs, S.L., van Bonn, S. (2010). A multi-method analysis of evaluation criteria used to assess the speaking proficiency of graduate student instructors. *Language Testing*, 27(2), 235-260.
- Powers, D.E., Schedl, M.A., Leung, S.W., and Butler, F.A. (1999). Validating the revised Test of Spoken English against a criterion of communicative success. *Language Testing*, 16(4), 399-425.
- Reed, D.J. (1992). The relationship between criterion-based levels of oral proficiency and norm-referenced scores of general proficiency in English as a second language. *System*, 20(3), 329-45.
- Ross, S. (1992). Accommodative questions in oral proficiency interviews. *Language Testing*, 9(2), 173-86.
- Ross, S. (1996). Formulae and inter-interviewer variation in oral proficiency interview discourse. *Prospect*, 11(3), 3-16.
- Shohamy, E. (1983). The Stability of the Oral Proficiency Trait on the Oral Interview Speaking Tests. *Language Learning*, 33(4), 527-539.
- Shohamy, E. (1994). The Validity of Direct versus Semi-Direct Oral Tests. *Language Testing*, 11(2), 99-123.
- Spolsky, Bernard. (2001). Prospects for the survival of the Navajo language: a reconsideration. In Tuttle, S., and Holton, G. (eds.), *Proceedings of the 2001 Athabaskan Languages Conference: University of California Los Angeles*, (pp. 32-53). Fairbanks Alaska: Alaska Native Language Center Working Papers Number 1.
- Stansfield, C.W. (1990). An evaluation of Simulated Oral Proficiency Interviews as measures of oral proficiency. In Alatis, J. E. (ed.), *Georgetown University roundtable on languages and linguistics*, (pp. 228-234). Washington, DC: Georgetown University Press.
- Stansfield, C. W., and Kenyon, D. M. (1992). Research on the comparability of the Oral Proficiency Interview and the Simulated Oral Proficiency Interview. *System*, 20(3), 347-364.

- Swender, E. (1999). *ACTFL Oral Proficiency Interview Tester Training Manual*. Yonkers, NY: American Council on the Teaching of Foreign Languages.
- Thompson, I. (1995). Inter-rater reliability of the ACTFL oral proficiency interview in five European languages: Data from English, French, German, Russian, and Spanish. *Foreign Language Annals*, 28, 407-22.
- Upshur, J., & Turner, C. (1995). Constructing rating scales for second language tests. *English Language Teaching Journal*, 49, 3–12.
- Upshur, J., & Turner, C. (1999). Systematic effects in the rating of second-language speaking ability: test method and learner discourse. *Language Testing*, 16, 82–111.
- Weir, C., J. (1990). *Communicative Language Testing*. NJ. Printice Hall Regents.
- Widdowson, H. (1978). *Teaching Language as Communication*. Oxford: Oxford University Press.
- Wilds, C. P. (1975). The oral interview test. In Jones, R. L. and Spolsky, B. (eds.), *Testing Language Proficiency*, Arlington Center for Applied Linguistics.
- Young, R. (1995). Conversational styles in language proficiency interviews. *Language Learning*, 45, 3-42.
- Young, R. & Halleck, G. B. (1998). “Let them eat cake!” or how to avoid losing your head in cross-cultural conversations. In Young, R., and He, A. W. (eds.), *Talking and testing: Discourse approaches to the assessment of oral proficiency*, (pp. 355–382). Amsterdam/Philadelphia: Benjamins.

Title

Accurate Recall Concerns in Language Classrooms: The Role of Clause Order in Memory Constraints

Authors

Amin Marzban (Ph.D. Candidate)

University of Isfahan, Isfahan, Iran

Mohammad Ali Ayatollahi (Ph.D. Candidate)

University of Isfahan, Isfahan, Iran

Bio Data

Amin Marzban, Ph.D candidate at the University of Isfahan, and is a member of teaching staff at Islamic Azad University. He has published articles in international journals including Asian EFL journal and Linguistics Journal. His research interests include psycholinguistics and teaching methodology.

Mohammad Ali Ayatollahi, Ph.D candidate at the University of Isfahan and a member of teaching staff at Islamic Azad University in Sepidan.

Abstract

This study explored whether the nonstandard complex sentences can constrain sentence processing. For this purpose, an experiment was designed to measure the repetition time and facility of recall for two types of structures i.e. complex sentences in which the matrix clause preceded the embedded one (standard clause order) and those in which the matrix clause followed the embedded clause (nonstandard clause order). The purpose was

to see if this type of syntactic alteration would exert additional processing load in a recall task. Upon exposure to these pairs, participants were asked to reproduce the heard material with standard clause order. It meant that in clauses with nonstandard order they had to rearrange them into standard ones while for the other group they had to merely repeat the heard material. Analysis of the results of a recall test revealed that when recalling the sentences with nonstandard clause order participants scored lower than they did when recalling sentences with standard clause order. T-test analysis revealed significant differences in the recall time needed and the accuracy of the repetitions. The findings have implications for both material developers and language teachers who make use of repetition, role plays and other types of oral practice in their courses.

Keywords: Working memory, repetition, processing load, time limit and recall tasks.

Introduction

Background

Memory theoreticians and researchers are still baffled by the fact that despite the vast capacity of mind to store and retrieve information, the amount that one can bring to mind at once is quite limited, suggesting a time limit operating on memory. Early research on the nature of such a limit focused on the notion of memory as a dual system. This view saw such a limit as an inherent property of mind rather than of the input. More recently, however, researchers have shifted their attention to the nature of the input and sentence processing mechanisms. Research on (extra-)linguistic factors affecting human sentence processing mechanisms suggests a number of influencing factors including negation (Clark, 1974), metaphors (Ortony *et. al.*, 1978), ambiguity (Frazier and Rayner, 1982), ellipsis (Garnham and Oakhill, 1987), anaphor resolution (Sepassi and Marzban, 2003, Sepassi, 2002), and sentence complexity (Bates, *et.al.* 1999) influence sentence processing. Working from both a linguistic (clause order) perspective, this study attempted to investigate the effect of clause order and processing load on sentence processing.

Statement of the problem

Recall tasks mainly require the parsers to process the heard material and produce some kind of output. They are notoriously dependent on working memory and the constraints associated with it (Baddeley and Gathercole, 1993; Carrol, 1999). Theories of sentence processing divide over the importance of the role of extralinguistic factors such as memory capacity and pragmatic expectations in sentence processing. While some theoreticians (e.g. Frazier and Rayner, 1982; Gorrell, 1995) consider the role of working memory capacity very important, others (e.g. Gibson, 1991; Pritchett, 1992; Weinberg, 1992; Philips, 1995, 1996) argue that processing is a function of the rapid incremental satisfaction of grammatical constraints such as the theta criterion rather than extralinguistic factors. There are also researchers who shun the question altogether. Instead, they attribute sentence processing difficulties to frequency in the hearer's linguistic environment. Thus, any difficulty in sentence processing could be attributed to low frequency of the concerned phenomenon (e.g. Trueswell *et. al.*, 1994; McDonald *et. al.*, 1992).

Significance of the study

Most research techniques employed in this area require interpretation or at least the parsing of sentences. Such techniques are not capable of distinguishing issues related to working memory constraints, which is a non-verbal phenomenon from those related sentence processing, which is an essentially verbal phenomenon. Therefore, we need a research technique that while tapping working memory does not involve actual sentence processing. We believe that simple recall tasks are a good candidate for this purpose.

If we accept that the difficulties that speakers experience in processing some types of sentences are due to inherently syntactic phenomenon such as theta role assignment rather than extralinguistic factors such as memory constraints, then there would be no place for the occurrence of such problems in task that do not actually require sentence interpretation such as simple recall tasks. Conversely, if we observed that language speakers took more time to respond to a recall task in sentences with exactly the same propositional content with the only difference being in their clause order, we would be in a position to argue that extra time needed to accomplish such tasks was due to the extra processing load and working memory constraints.

Theoretical Background

Psycholinguistic evidence suggests that memory constraints are an important bottleneck in language processing (Baddeley and Gathercole, 1993). In their classic experiment Peterson and Peterson (1959) found that when unfamiliar information such as a nonsense syllable is presented to human subjects who are prevented from rehearsing the information, they will retain the material almost perfectly for about three seconds. In the following fifteen seconds, however, their retention drops to almost zero.

Since Atkinson and Shiffrin's (1968) introduction of the notion of working memory and working memory the idea has received wide attention from psycholinguists. As was previously mentioned, working memory continues for some time after the stimulation has ceased. Exactly how long the activity goes on or if it is individual specific is not yet known though many agree upon a period of less than one minute for its availability (Carroll, 1999). However, the perfect recall is said to be at hand during the first three seconds (Witting and Williams, 1984).

Some experiments using recall intervals of 3,6,9,12,15, and 18 seconds have shown that the best retention is at a recall interval of three seconds, next 6 seconds, and so on. The worst recall occurred after 18 seconds (Klatzky, 1980; Bransford, 1979, Peterson and Peterson, 1959). Others, however, suggest that the best retention occurs within the first two seconds and it decays continually up to 15 or 20 seconds (Cowen, 1988). Such observations confirm the claim that working memory decays as a function of time faster than long-term memory (Stevick, 1996, Baddeley, 1986).

Lombardi and Potter (1992) maintain that memory for recalling entire complex and compound sentences is often quite accurate at least for short retention intervals (Lombardi & Potter 1992). The important question here is what would happen to this accurate memory concern if some kind of syntactic alteration is involved? Would it affect the recall of the heard information? Ferreira & Dell (2000) and Sepassi & Marzban (2005) showed how resorting to a real processing experience may affect the time needed for the recall of the information for at least a few seconds. There seems to be a longer period needed to retain the heard input as long as processing is involved.

While processing might enhance retention, it may in return linger the accurate recall time a little. There has been little organized work done on this issue particularly in an

EFL context like Iran where many learners complain about inappropriate teacher expectation in fulfilling oral tasks. This very issue was addressed in this study.

One question for the researchers who investigated sentence memory concerns was whether or not negative sentences took longer to be processed. Clark (1974) showed that negative sentences generally took longer to be processed. Later, Clark and Lucy (1975) examined whether or not this was true with both the literal and non literal meaning of utterances. They found that this was indeed the case. That is, sentences with non literal meaning e.g. requests worded as negative questions took longer to be processed.

Similarly, Ortony *et. al.* (1978) investigated whether or not sentences containing metaphors took longer to be processed. They found that although metaphors generally took longer to be processed, with 45-word as a preceding context they took almost the same time as sentences with literal meaning.

Another concern of the researchers who worked on sentence memory concerns was whether or not listeners retain the meaning but not necessarily the form. This, indeed, turned out to be the case. A classic example is the study carried out by Sachs (1997). Sachs wanted to know whether the passage of time had any effect on whether the listeners retained only the meaning or both meaning and form. She had the participants in her study listen to passages on the tape. At various intervals she stopped the tape and presented the listeners with four choices: the original sentence, two paraphrases of the original sentence, and sentence with a different meaning and wording. She also changed the interval between the test cue and response time. The results showed that with time the retention of form declined, so that after about 12.5 seconds (40 syllables) the memory for form substantially declined. However, the memory for meaning remained generally unaffected by time. Ni *et.al.* (1996) found that subject relative clauses were easier to processes and parse than object relative clauses.

In the field of education Ausubel (1968) was among the first to notice the shortcomings of memory dependent learning activities. He proposed a helpful strategy he called *advance organizers* to help ease the burden of processing entirely new information. Advance organizers are pre-task activities that help relate new information with old information. They often take the form of such schema building activities as brief introductions, concept maps, discussion questions, etc that usually are placed at the beginning of any classroom activity.

This study sought to investigate the perfect recall interval for complex structures when a syntactic alteration is involved. The recall interval at play within this piece of research was the one maintained by Klatzky (1980). The syntactic alteration at play here is changing the position of main and subordinate clause in complex sentences. According to Caplan and Waters (1995) when the canonical order of thematic roles is not standard, complexity increases provided the number of propositions is the same. They argue "an appropriate measure of syntactic complexity is the comparison of sentences with canonical and non-canonical orders of thematic roles that are matched for length and number of propositions" (p. 640).

The following are the directional hypotheses tested in this study.

H1: $X_{SCO} > X_{NCO}$

H2: $X_{SCO} < X_{NCO}$

The null hypothesis would be:

$H_0: X_{SCO} = X_{NCO}$

Key:

X_{SCO} : Mean of structures with Standard Clause Order

X_{NCO} : Mean of structures with Nonstandard Clause Order

Method

Participants

The participants in the study consisted of 40 junior students of English at Islamic Azad University in Shiraz. Each of the participants was randomly chosen from among male and female students. The justification for choosing such participants-- particular target population—was that they had all been exposed to a sufficient corpus of both formal and informal English during their three years of study within the university context. This background knowledge would act as a built-in ability--bias-- to perform the tasks more professionally and/or prefer one strategy to another.

Stimuli

The stimuli consisted of 2 pairs of 20 complex structures serving as a recall task. Structurally and semantically, all the structures had the same type of words included, the only difference being the position of main and subordinate clauses. Within the first

set of sentences, the main clauses appeared first and the reverse took place in the second set. This was done to approach a real processing activity.

Furthermore, in order to ensure that processing load is the same across all the sentences, clauses starting with the pronoun <you> were avoided on the test. The reason was that clauses starting with the pronoun <you> and without <that> complementizer exert a temporary processing load on the hearers. This is due to the fact that at the moment the hearer hears the pronoun <you> he/she has few clues (except for the meaning of the preceding verb) as to whether the pronoun <you> is the beginning of a clause or the direct object of the main verb. In other words, part of the processing load would be due to the effort to disambiguate between the two momentarily possible interpretations (Sepassi and Marzban, 2005).

Table 3.1. Position of task structures

	First Clause	Second Clause
Pair I	I concealed	He overheard the secrets
Pair II	He overheard the secrets	I concealed

Procedure

The 20 pairs of the structures of the task, in both types of structures, were tape recorded and played back through headphones to the participants, on an individual basis. For the first set, participants were required to listen to what was played and repeat the heard structure with no pause. Then, the same group of participants was instructed to put the two clauses of each structure of the second set in the correct order (i.e. to change the order of presentation) after hearing the material, and to orally report the reconstructed sentences to the interlocutors sitting in front of them. The counter balance design was used to take care of order effect.

This scheme (i.e. the use of headphones and an interlocutor – the interviewer) provided for an air of authenticity for the task. This claim was substantiated on the grounds that a real communicational setting would be constructed when the participants report the reconstructed sentences to an actual interlocutor. The criterion for scoring was whether or not the participants were able to repeat back the sentences they heard in

their canonical i.e. standard order. The participants received 5 marks for each sentence if they were able to reproduce the sentences in the first recall interval i.e., up to 3 seconds. They received 4 marks if were able to reproduce the sentence in the second interval i.e. 3 to 6 seconds and so forth. Therefore, a student with the perfect recall would score 100. However, later the grades were scaled down to a 10 point scale for ease of interpretation. Therefore, a participant with perfect recall would receive a grade of 10.

Data Analysis Procedure

The difference between the means for the accurate recall of the heard material in two streams of input (i.e. structure types 1 and 2), where in one set mere repetition and in the other some kind of syntactic alteration was added to the activity, could be the answer to the research question. In order to obtain the results paired t-test was run on between type means of the structures for the correct retention of the material.

Results and Discussion

The result of a paired t-test carried out on the recall task suggested a significant difference in the recall of nonstandard sentences, which involved extra- processing load and standard sentences, which involved no syntactic alteration. Thus the null hypothesis was rejected at the significant level of $p < 0.01$.

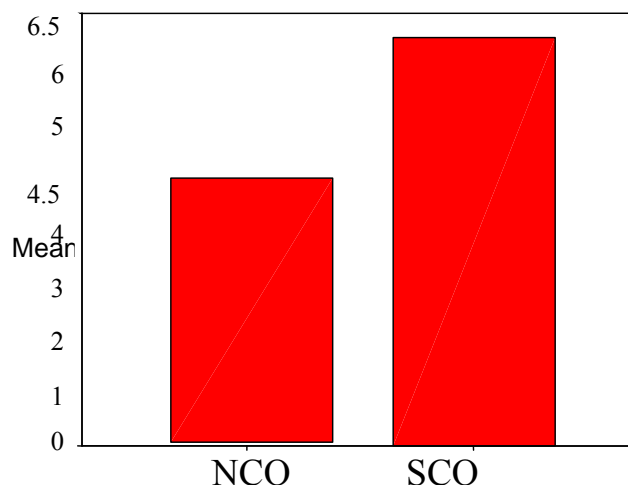
Table 3.1. t-test analysis of standard sentence and nonstandard sentence groups

Variable	No of Pairs	Mean	SD	t-value	df	Sig.
SCO	20	6.4634	1.485	6.55	19	.000
NCO		4.6341	1.655			

*** Significant at confidence level $p < 0.01$**

Key: SCO = Standard Clause Order

NCO = Nonstandard Clause Order



Conclusion

This paper reported an experiment designed to test whether syntactic alternations between canonical (standard) and non-canonical (nonstandard) clause order in complex sentences had any effects on sentence processing. It also sought to find out which of the three theoretical positions outlined in the introduction could be best supported by the experiment carried out in this study. The results showed that sentences with canonical clause order were more facilitative of recall than sentences with non-canonical clause order.

Could the lower recall facility of nonstandard sentences be attributed to the nature of input as Gibson, 1991; Pritchett, 1992; Weinberg, 1992; and Philips, 1995, 1996 seem to be implying? Or does it have to do with working memory constraints as Frazier and Rayner, 1982; and Gorrell, 1995 claim? Assuming that simple recall tasks that do not require the respondent to interpret the sentences tap only working memory but not sentences and syntactic processing, the experiment carried out and analyzed in this study seems to suggest that working memory constraints play a role in the efficiency of recall tasks. This finding is in line with Frazier and Rayner, 1982; and Gorrell, 1995 findings. The relative difficulty that the participants in this study experienced in the processing of non-canonical complex sentences suggests the recall tasks when combined syntactic alterations requires the allocation of more processing resources.

References

- Atkinson, R. C. & Shiffrin, R. M. (1968). Human memory: a proposed system and its control processes. In W. K. Spence J.T. Spence (Eds.) *The psychology of learning and motivation: Advances in research and Theory*. New York: Academic Press.
- Ausable, D. (1968). *Educational psychology: A cognitive view*. New York: Holt, Reinhart and Winston.
- Baddeley, A. D.(1986). *Working memory*. Oxford, England: Clarendon Press.
- Baddeley, A. D., and S. E. Gathercole. (1993). *Working Memory and Language*. Hillsdale, NJ: Erlbaum.
- Bates, E., Devescovi, A. & D' Amico, S. (1999). Processing complex sentences: A cross linguistic study. *Language and Cognitive Processes*, 14, 69-123.
- Bransford, J. (1979). *Human cognition: Learning, understanding, and remembering*. Belmont, CA: Wadsworth.
- Caplan, D. & Waters, G. (1995). Aphasic disorders of syntactic comprehension and working memory capacity. *Cognitive Neuropsychology*, 12(6), 637-649.
- Carrol, D. (1999). *Psychology of Language*. New York: Brooks/Cole Publishing Company.
- Clark, H. H. (1974). Semantics and comprehension. In T. A. Sebeok (Ed.), *Current trends in linguistics*, Vol. 12. (pp 1292-1428). The Hague: Mouton.
- Clark, H. & Lucy, P. (1974). Understanding what is meant from what is said: A study in conversationally conveyed requests. *Journal of Verbal Learning and Verbal Behavior*, 14, 56-72.
- Cowen, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information processing system. *Psychological Bulletin*, 104, 163-191.
- Ferreira, V.S. & Dell, G. S. (2000) Effect of ambiguity and lexical availability on Syntactic and Lexical Production. *Cognitive Psychology*, 40, 296-340.
- Frazier, Lyn and K. Rayner (1982) Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, 14 (2)178-210.
- Garnham, A., & Oakhill, J. (1987). Interpreting elliptical verb phrases. *Quarterly Journal of Experimental Psychology*, 39A, 611-625.

- Gibson, E. (1991). *A Computational Theory of Human Language Processing: Memory Limitations and Processing Breakdown*. Unpublished doctoral dissertation, Carnegie-Mellon.
- Gorrell, P. (1995) *Syntax and Parsing*. Cambridge: Cambridge University Press.
- Klatzky, R. L. (1980). *Human memory structures and processes*. Sanfransisco: Freemsn Publishers.
- Lombardi, L. & Potter, M.C. (1992). The regeneration of syntax in short term memory. *Journal of Memory and language, 31*, 713-733.
- MacDonald, M. C., Just, M. A., & Carpenter, P. A. (1992). Working memory constraints on the processing of syntactic ambiguity. *Cognitive Psychology, 24*, 56-98.
- Ni, W. Shankweiler, D. & Crain, S. (1996). Individual differences in working memory and eye-movement patterns in reading relative clause structures. *University of Connecticut Working Papers in Linguistics 6*.
- Ortony, A., Schallert, D. L., Reynold, R. E. & Amos, S. J. (1987). Interpreting metaphors and idioms: some effects of context on comprehension. *Journal of Verbal Learning and Verbal Behavior, 17*,456-477.
- Peterson, R. L. & Peterson, M. J. (1959). Short-term retention of individual items. *Journal of Experimental Psychology, 58*(2), 193-198.
- Phillips, C. (1995). Right Association in parsing and grammar. In C. Schutze, J. Ganger, and K. Broiher, (Eds.), *Papers in Language Processing and Acquisition*, (pp. 37-93).
- Phillips, C. (1996). *Order and Structure*. Unpublished doctoral dissertation, Massachusetts Institute of Technology - Massachusetts.
- Pritchett, B. (1992) *Grammatical Competence and Parsing Performance*. Chicago: University of Chicago Press.
- Lombardi, L., & Potter, M.C. (1992). The regeneration of syntax in short term memory. *Journal of Memory and language, 31*, 713-733.
- Sachs, J. (1997). Communication development in infancy. In J. B. Geason(Ed.). *The development of language 4th ed.* (pp. 40-68).Boston: Allyn & Bacon.
- Sepassi, F. (2002). *Intelligence-related constrains on information processing to serve as a platform to enhance the description adequacy of the competition model*. Unpublished PhD dissertation, Shiraz University, Iran.
- Sepassi, F and Marzban, A. (2005). On the effects of economization and disambiguation in the production of EFL learners. *Asian EFL Journal, 2*, Article 2. Retrieved August 20, 2008, from www.asian-efl-journal.com.

Stevick, E.W. (1996). *Memory, meaning and method* (second edition). Boston, MA: Heinle and Heinle.

Witting, A. F. & Williams, G. (1984). *Psychology: An introduction*. New York: McGraw-Hill Book Co.

Weinberg (1992). Parameters in the theory of sentence Processing: Minimal Commitment Theory Goes East. *Journal of Psycholinguistic Research*, 22, 339-364.

Appendix

Example of complex structures used

Main Clause	Subordinate Clause
I heard	he warned the president about the economy
I concealed	she overheard the military secrets
I worried	he trusted too many people.
I dreamed	she forgot to go to the big concert
I guessed	he remembered the questions on the TOEFL
I observed	she confirmed the hotel reservation
I revealed	he protested against government
I replied	she charged too much for the computer
I thought	he learned 5 languages
I suspected	she felt uncomfortable