



**UNIVERSITI PUTRA MALAYSIA**

**DEVELOPING LEARNERS' SELF-ASSESSMENT ACCURACY IN READING SKILLS  
AND STRATEGIES**

**FARAMARZ AZIZMALAYERI**

**FPP 2008 35**



**DEVELOPING LEARNERS' SELF-ASSESSMENT  
ACCURACY IN READING SKILLS AND  
STRATEGIES**

**FARAMARZ AZIZMALAYERI**

**DOCTOR OF PHILOSOPHY  
UNIVERSITI PUTRA MALAYSIA**

**2008**



**DEVELOPING LEARNERS' SELF-ASSESSMENT ACCURACY IN  
READING SKILLS AND STRATEGIES**

**By**

**FARAMARZ AZIZMALAYERI**

**Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia,  
in Fulfillment of the Requirements for the Degree of Doctor of Philosophy**

**September 2008**



To my late father,  
my loving mother,  
my supportive wife Masoumeh, and  
my cute son Farbod



Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfillment  
of the requirement for the degree of Doctor of Philosophy

**DEVELOPING LEARNERS' SELF-ASSESSMENT ACCURACY IN  
READING SKILLS AND STRATEGIES**

By

**FARAMARZ AZIZMALAYERI**

**September 2008**

**Chairman: Arshad Abd. Samad, PhD**

**Faculty: Educational Studies**

Self-assessment is one example of what is currently called alternative assessment. Alternative assessment procedures are often developed in an attempt to make testing and assessment more responsive and accountable to individual learners, to promote learning, and to enhance ethic and equality in education (McNamara, 1998). Attempts have been made to ensure validity and to establish credibility for self assessment as an alternative assessment. The currently in-use self assessment instrument maintains 'can-do statements' which are developed based on the ACTFL. The new guidelines by the Council of Europe (2001) are intended to be used as



references for languages called The Common European Framework for References (CEFR). The utilized self assessment instrument in this research is CEFR-based officially developed.

The purpose of this study is to empirically investigate the effects of training students taking self assessment checklist and reflecting to the reading proficiency benchmarks. The learners' accuracy with self assessment is investigated across an intensive period of training. The TESL students have been trained for a whole semester taking the CEFR-based self assessment of reading skills. There are four specific objectives for this study: 1) to investigate the accuracy of self assessment results, 2) to determine the optimal training period, 3) to estimate the maximum correlation between the CEFR-based self assessment instrument and an established measure of reading proficiency (TOEFL), and 4) to probe into the quality of improvement and change in learners' perceptions, understanding and internalization of the CEFR benchmarks. In order to meet the above objectives, four research questions were asked. The research design utilized is mixed-method quasi-experimental design in which qualitative findings triangulate the quantitative results. A class of semester 7 TESL students, as a whole intact group, attended a semester training period. They were tested for their current reading proficiency using the TOEFL reading test battery. Meanwhile, they attempted the CEFR-based self assessment checklist of reading skills three times across the training period. These two measures were used for the purpose of the statistical analyses. Repeated Measures Two-Way ANOVA and Multiple Regression were used to test the hypotheses.



It was found out that the provided training has been effective as the student's results were more accurate in the second and the third administrations of the self assessment compared to the first attempt (large Effect Size,  $\eta^2 = .185$ ,  $p < .05$ ).

The results of the Stepwise Multiple Regression Analysis revealed that among the three administrations of the self assessment instrument, the second and the third administrations explained 79% of the observed variation in the dependent variable together; therefore, the adopted model included the second and the third administrations. The Correlation Coefficients in the Correlation Matrix depicts that the third administration of the self assessment instrument is highly correlated with the external proficiency measure (TOEFL) ( $r = .88$ ,  $p < .05$ ). The qualitative inquiry revealed that students are apt to achieve self assessment autonomy and acquire the ability to self assess via intensive training. Therefore, as the result of training students taking self assessment of reading, their results are more accurate and dependable.



Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Doktor Falsafah

MEMBINA KETEPATAN PENILAIAN KENDIRI PELAJAR DALAM  
KEMAHIRAN DAN STRATEGI MEMBACA

Oleh

**FARAMARZ AZIZMALAYERI**

**September 2008**

**Pengerusi: Arshad Abd. Samad, PhD**

**Fakulti: Pengajian Pendidikan**

Penilaian sendiri adalah satu contoh pentaksiran alternatif. Prosedur pentaksiran alternatif biasanya dibina sebagai percubaan untuk menjadikan pengujian dan pentaksiran lebih responsif dan bertanggungjawab kepada pelajar secara individu dalam menggalakkan pembelajaran dan, untuk meningkatkan etika dan kesamaan dalam pendidikan (McNamara, 1998). Usaha telah dilakukan untuk memastikan kesahihan dan kredibiliti pentaksiran sendiri sebagai pentaksiran alternatif. Instrumen pentaksiran sendiri yang sedang digunakan mengekalkan “pernyataan boleh melakukan” yang dibina berasaskan ACTFL. Garis panduan baru oleh *The Council of Europe* (2001) dihasratkan menjadi rujukan untuk bahasa yang dipanggil sebagai *The Common European Framework for References* (CEFR). Instrumen yang digunakan dalam penyelidikan ini, secara rasminya dibina berasaskan CEFR. Tujuan kajian ini adalah untuk meneliti kesan latihan menggunakan senarai semak dan melakukan refleksi terhadap penanda aras kefasihan membaca secara empirikal melalui pentaksiran sendiri. Ketepatan pentaksiran pelajar diukur sepanjang satu tempoh latihan yang intensif. Pelajar TESL berkenaan telah dilatih menggunakan pentaksiran sendiri yang berasaskan CEFR selama satu semester dalam kemahiran membaca.





Kajian ini mempunyai empat objektif khusus, iaitu untuk: 1) menyiasat ketepatan keputusan pentaksiran sendiri, 2) menentukan tempoh latihan yang optimal, 3) menganggar korelasi maksimum antara alat penilaian sendiri berasaskan CEFR dengan pengukuran kefasihan membaca mapan (TOEFL), dan 4) mencungkil kualiti penambahbaikan dan perubahan dalam persepsi, kefahaman, dan pencernaan penanda aras CEFR dalam kalangan pelajar. Bagi memenuhi objektif di atas, empat soalan kajian telah dibina. Untuk itu, kajian ini menggunakan reka bentuk campuran (mixed method) melalui gabungan kaedah eksperimen quasi dengan kaedah kualitatif bagi membolehkan dapatan kuantitatif dipadankan dengan dapatan kualitatif. Satu kelas pelajar TESL Semester 7 telah diberi latihan berkenaan. Mereka diuji tentang kefasihan membaca menggunakan soalan ujian TOEFL. Pada masa yang sama mereka mencuba senarai semak pentaksiran sendiri CEFR sepanjang tempoh latihan. Kedua-dua ukuran ini digunakan untuk tujuan analisis statistik. Hipotesis diuji menggunakan Ukuran Berulang ANOVA Dua Hala dan regresi berganda. Kajian mendapati latihan yang diberi adalah efektif kerana prestasi pentaksiran sendiri kedua dan yang ketiga adalah lebih tepat daripada yang pertama (large Effect Size,  $n^2=0.185$ ,  $p<.05$ ). Keputusan analisis regresi berganda Stepwise menunjukkan, antara ketiga-tiga kali pentaksiran sendiri berkenaan, pentaksiran pada kali kedua dan ketiga lebih tepat dan ini menjelaskan kewujudan 79% variasi di kalangan keseluruhan pembolehubah bersandar; model yang digunakan memuatkan pentadbiran ujian kedua dan ketiga. Koefisien hubungan dalam matrik hubungan menunjukkan bahawa pentadbiran pentaksiran ketiga mempunyai korelasi yang tinggi dengan pengukuran kefasihan luaran (TOEFL) ( $r=.88$ ,  $<.05$ ). Kajian secara kualitatif pula menunjukkan bahawa pelajar berpotensi untuk membina autonomi dalam pentaksiran sendiri melalui latihan intensif. Oleh itu, latihan menduduki pentaksiran sendiri tentang membaca menghasilkan keputusan yang lebih tepat dan boleh dipertanggungjawabkan.

## ACKNOWLEDGEMENTS

I would like to express my sincere thanks and appreciation to a number of people for their generous assistance and encouragement during this project. I owe special thanks and appreciation to my supervisor, mentor, and the Deputy Dean of the Faculty, Associate Professor Dr. Arshad Abd. Samad for his awe-inspiring contributions, insightful input, intellectual support, comments and cheering, which made this research possible. Without his help, encouragement, and patience, this research would still be in the evolutionary stage. Dr. Arshad had always believed in me, when he trusted me to replace him for numerous course credits, when he believed in my methodology, and in many times when he wanted me to help out. Thank you Dr. Arshad.

I also wish to thank my committee member and the Dean of the Faculty, Professor Ab. Rahim Bakar who offered valuable intellectual input. Without his help, I would not have come so far. Prof. Rahim has always been a close friend to talk to. Thank you Prof. Rahim.

Many thanks are also to my committee member and the Head of the Department Dr. Roselan Bin Baki for his guidance, passion, and extensive contributions on my dissertation. Dr. Roselan was always interested in reading my manuscripts. His most welcoming smiles were always offered every time we had meetings. Dr. Roselan I appreciate that, thank you.



I also appreciate Associate Professor Dr. Jayakaran Mukudan the chair of the examination committee for his support and encouragement which I have always had, from when he read the first proposal and provided insightful comments to the viva session. Thank you Dr. Jaya.

I appreciate that I have learned from these scholars more than from my readings. Whatever is presented in this dissertation comes from this learning than the books, as after all, the journey matters not the destination. I acknowledge that all of the mistakes remaining in the dissertation are of my own.

I am deeply grateful to my mother, my mother-in-law, my sister-in-law in the US and to the rest of my family who kept encouraging me during my study. My thanks to them who believed in me, expressed their patience and provided me with moral support for the past three years.

Finally, I am endlessly thankful to my nice wife Masoumeh who is also my best friend. She cheered and supported me daily, with her patience and extra generosity with her time to edit my stylistic errors and correct my mistakes. She continuously encouraged me through these years of my study and research. My degree and this thesis would not have been accomplished without her.

The last but not the least, my special thanks to my cute son Farbod for the time we should have spent together. Thank you son, I love you.



I certify that an Examination Committee has met on 12<sup>th</sup> September 2008 to conduct the final examination of Famarz Azizmalayeri on his Doctor of Philosophy thesis entitled “Developing Learners’ Self-assessment Accuracy in Reading Skills and Strategies” in accordance with Universiti Pertanian Malaysia (Higher Degree) Act 1980 and Universiti Pertanian Malaysia (Higher Degree) Regulations 1981. The Committee recommends that the student be awarded the Doctor of Philosophy.

Members of the Examination Committee are as follows:

**Jayakaran Mukudan, PhD**

Associate Professor  
Faculty of Educational Studies  
Universiti Putra Malaysia  
(Chairman)

**Malachi Edwin Vethamani, PhD**

Associate Professor  
Faculty of Educational Studies  
Universiti Putra Malaysia  
(Internal Examiner)

**Mohamad Sahandri Gani Hamzah, PhD**

Associate Professor  
Faculty of Educational Studies  
Universiti Putra Malaysia  
(Internal Examiner)

**Mohamad Amin Embi, PhD**

Professor  
Faculty of Education  
Universiti Kebangsaan Malaysia  
(External Examiner)

---

**HASANAH MOHD. GHAZALI, PhD**

Professor and Dean  
School of Graduate Studies  
Universiti Putra Malaysia

Date: 27 November 2008



This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfillment of the requirement for the degree of Doctor of Philosophy. The members of the Supervisory Committee were as follows:

**Arshad Abd. Samad, PhD**  
Associate Professor  
Faculty of Educational Studies  
Universiti Putra Malaysia  
(Chairman)

**Ab Rahim Bakar, PhD**  
Professor  
Faculty of Educational Studies  
Universiti Putra Malaysia  
(Member)

**Roselan Bin Baki, PhD**  
Lecturer  
Faculty of Educational Studies  
Universiti Putra Malaysia  
(Member)

---

**HASANAH MOHD GHAZALI, PhD**  
Professor and Dean  
School of Graduate Studies  
Universiti Putra Malaysia

Date: 18 December 2008



## **DECLARATION**

I declare that the thesis is my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously, and is not concurrently, submitted for any other degree at Universiti Putra Malaysia or at any other institution.

---

**FARAMARZ AZIZMALAYERI**

Date: September 2008



## TABLE OF CONTENTS

	<b>Page</b>
<b>ABSTRACT</b>	iii
<b>ABSTRAK</b>	vi
<b>ACKNOWLEDGEMENTS</b>	viii
<b>APPROVAL</b>	x
<b>DECLARATION</b>	xii
<b>LIST OF TABLES</b>	xvi
<b>LIST OF FIGURES</b>	xvii
<b>LIST OF ABBREVIATIONS</b>	xviii
<b>CHAPTER</b>	
1 <b>INTRODUCTION</b>	1
2 <b>LITERATURE REVIEW</b>	24
Introduction	24
Learner Autonomy	27
Autonomy and Self-assessment	30
Self-Assessment and Learner-Awareness	33
Learner Training	37
Self-Assessment as Norm-Referenced Assessment	39
The ACTFL Guidelines as the Criterion	40
Alternative Assessment	44
Self-Assessment, Theory and Practice	50
Self-Assessment before CEFR	53
CEFR-Based Self-Assessment	54
The Common European Frameworks	57
Pros and Cons for CEFR	60
Carroll's Model	71
Bachman's (1990) Model of Communicative Language Ability	73
Assessing Reading Comprehension	76
The Factorial Approach	77
Informed Tuition: the sub-skill approach to reading comprehension	78



	Think-aloud techniques in reading comprehension	81
3	<b>METHODOLOGY</b>	86
	Introduction	86
	Procedure	86
	Phase one: Preliminary data collection	89
	Phase two: Actual data collection	94
	Phase three: Qualitative data collection	96
	Research Design	101
	The Quasi-Experimental Design	104
	Internal Threats	106
	Threats to External Validity	108
	Review of Research Questions and their Related Statistical Analyses	111
	Research Question One	112
	Research questions for the Repeated Measure Two-Way ANOVA	115
	Research Question Two	116
	Main regression research question	117
	Research Question Three	119
	Qualitative Methodology Reviewed	121
	Research Question Four	122
	Model Explanation	127
	Location and Duration of the Study	129
	Sampling	129
	Instruments	130
	The TOEFL, Test of General Reading Proficiency	131
	The CEFR-Based Self-Assessment Checklist of the Reading Skills and Strategies	131
	Quantification of the Checklist	132
	Description of the Treatment	133
	Data Analysis	138





4	<b>RESULTS AND DISCUSSION</b>	141
	Introduction	141
	Demographic report of the participants	141
	Research Objectives and Research Questions	146
	Research Objective and Research Question one	146
	Discussion for Research Question one	159
	Research Objective and Research Question two	166
	Discussion for Research Question two	169
	Research Objective and Research Question three	172
	Discussion for Research Question three	175
	The Qualitative Research Question	177
	The Qualitative Analysis	178
	Summary	193
5	<b>SUMMARY, CONCLUSION AND RECOMMENDATIONS FOR FUTURE RESEARCH</b>	194
	Introduction	194
	Summary of the Major Findings and Conclusions	195
	Final Remarks	208
	Suggestions and Recommendations for Future Studies	210
	<b>REFERENCES</b>	216
	<b>APPENDICES</b>	228
	<b>BIODATA OF STUDENT</b>	264



## LIST OF TABLES

<b>Table</b>	<b>Page</b>
4.1 Frequency Distribution of Gender	142
4.2 Frequency distribution for race	143
4.3 Reported SPM results	144
4.4 Reported MUET Overall Results	145
4.5 Descriptive Statistics for Trial (3 levels) and Treatment (3levels)	148
4.6 The Multivariate Tests for the Trial Main Effect [within-subject]	150
4.7 Assumption of Sphericity	151
4.8 Treatment Main Effects (between-subjects effect)	151
4.9 Pairwise Comparison for the group differences in Treatment	152
4.10 Trial Pairwise Comparison (Training)	154
4.11 Treatment Pairwise Comparison (achievement)	155
4.12 Comparison between Trial and Treatment	156
4.13 Model Summary table	167
4.14 Regression ANOVA Table	168
4.15 Coefficient Table	168
4.16 Correlation Coefficients	173
4.17 Generated proficiency descriptors (vague descriptors)	180
4.18 Generated descriptors (unambiguous)	181
4.19 Summarized proficiency bands	182



## LIST OF FIGURES

<b>Figure</b>	<b>Page</b>
2.1 The ACTFL Proficiency Pyramid	42
2.2 The CEFR Levels	55
2.3 Bachman's (1990) Model of Target Language Use (Components)	74
2.4 Bachman and Palmer's (1996) Model of Target Language Use	75
2.5 Theoretical framework	85
3.1 General Model for Non-Randomized, Pretest-Posttest Design	105
3.2 Repeated Measures Two-Way ANOVA (3X3)	113
3.3 Multiple Regression Relationships	116
3.4 General Model of the Qualitative Design	126
3.5 The Schematic Diagram of the Intervention/Treatment	137
3.6 The Research Framework	140
4.1 The Estimated Marginal Means for Trial	158



## **LIST OF ABBRIVATIONS**

**ACTFL:** American Council on the Teaching of Foreign Languages

**CEFR:** Common European Framework of References

**TOEFL:** Test Of English as a Foreign Language

**TOEIC:** Test Of English used for International Communication



## CHAPTER 1

### INTRODUCTION

#### **Background of the Study**

Self-directed language learning and learner autonomy have become increasingly popular in recent years. In accordance with these developments, language assessment and evaluation have also experienced drastic changes. Many advances are made with the theme of sharing the responsibility of learners' learning with themselves in language education. Assessment practices now seem to have started assuming learners' responsibility for the assessment of their learning processes and products (Little, 1991; Benson and Voller, 1997; Benson, 2001; Luoma and Tarnanen, 2003; Gardner, 1999). Teaching methodology, testing and assessment innovations, and many theoretical improvements have been introduced and are re-examined approaching the new student-centered paradigm (Nunan, 1998). This was called the Alternative Paradigm (Hamayan, 1995). The Alternative Paradigm refers to the procedures and techniques which can be used within the context of instruction and can be easily incorporated into the daily activities of school or classroom (Hamayan, 1995, p. 213). It is particularly useful with English as a Second Language students because it employs strategies that ask students to show what they can do, in contrast to the traditional testing in which "students are evaluated on what they integrate and produce rather than on what they are able to recall and produce" (Huerta-Macias, 1995, p. 9). The alternative paradigm relies on the student's performance while attending to the goals and achieving the objectives of the test. It



is claimed that alternative assessment generally meets some major criteria (Huerta-Macias (1995, p. 90), as follow:

- 1) Focuses on documenting individual student growth over time, rather than comparing students with one another in a criterion-based referenced system of values.
- 2) Emphasizes students' strength, on what they know not on what they do not, and
- 3) Considers learning styles, language proficiency, cultural and educational backgrounds.

Huerta-Macias (1995, p. 90) states that the main goal for alternative assessment is to “gather evidence about how students are approaching, processing, and completing real life tasks in a particular domain.” In order to assess language abilities in an alternative paradigm, one must ensure of assessing language skills and sub-skills in a well-supported assessment system of language performance. This performance-based assessment is believed to have attended the concepts of process-based views of language learning and language teaching. Baron (1991, p. 190) states that “when students internalize a definition of what quality means and can learn to recognize it, they have developed a very valuable critical ability. They can talk with their teacher about the quality of their work and take steps to acquire the knowledge and skills required to improve it”.



## **Alternative Assessment and Learner Autonomy**

The ongoing nature of alternative assessment involves students and teachers in making judgments about the students' progress in language using non-conventional strategies (Hancock, 1994, p. 7). Therefore, learner's autonomy and critical learning theories play significant roles in justifications for utilizing alternative assessment. Recent discussions have apparently emphasized that students tend to reflect and provide feedback on their personal satisfaction with their learning (Gardner and Miller, 1999). Learner's self-reflection is the very primary step of autonomy. Hence, autonomy is defined as "situations in which the learner is totally responsible for the decisions concerned with his leaning and the implementations of those decisions" (Gardner and Miller, 1999, p. 6). Autonomous learner is, by definition, "an active participant in the social processes of classroom learning, and an active participant in interpreting the new information in terms of what he already and uniquely knows" (Dam, *et a.l.*, 1990, p. 102). In order for achieving a self-generated estimation of one's own general or specific cumulative language proficiency, language learners are recommended to choose among the many multi-purpose instruments which have been experimentally tested so far. Among these instruments, some are made popular due to practitioners' and researchers' results achieved in a number of qualitative and quantitative studies (Bachman and Palmer, 1989; Oscarson, 1997; Ross, 1998). So far, portfolios and self-assessments are explored experimentally and practiced globally. Norris (2004, p. 1) reviews these assessment genres in: "recent discussion of so-called 'alternative assessment' has highlighted the potentials, usefulness of a variety of innovative testing procedures, including portfolios, self and peer

assessment, conferencing, diaries, and learning logs and teacher checklists and observations”.

Self-assessment appeared to come of age in the 1980s with the publication of a Council of Europe text on the topic by Oscarson (Todd, 2002). It is now widely accepted that self-assessment is a successful attempt for assessing learning process and locating personal profile matched or miss-matched stance (McNamara, 2000). Since its early introduction, self-assessment has been attempted frequently around the world. Among others, the *Council of Europe* has remarkably conducted researches trying to test the usefulness of self-assessment in language learning and assessment (North, 2000). The result was the new self-assessment checklist based on the confirmed CEFR guidelines which will be discussed later in this chapter.

### **The Common European Framework of Reference for Languages (CEFR)**

*The Common European Framework of Reference (CEFR)* was introduced after almost a period of almost forty years of researches and inquiries into the establishing of a “Common Reference” of criterion for practicing teaching, assessing, and evaluating languages in the European context. The foundation relies on the early conceptualizations and the speculations of the Notional-Functional syllabus (Wilkins, 1976), and the development in years after the first introduction in the theory of The Threshold Level (van Ek, 1976; Trim, 1999). The Council of Europe in 2001 released the revised edition of these references which contained six levels, A1, A2, B1, B2, C1, and C2. Attempts are made to describe language proficiency





through a group of scales composed of ascending levels of descriptors couched in terms of outcomes (Weir, 2005). The major implications and applications for the CEFR references are cited as follows (The Council of Europe, 2001, p. 16):

- 1) For the specification of the contents of the tests and examinations,
- 2) For stating the criteria to determine the attainment of a learning objective, and
- 3) For describing the level of proficiency in existing tests and examinations, then enabling comparisons to be made across different systems of qualifications.

The four language skills are addressed across the six main levels. The reading skill is conceptualized in terms of attainments and abilities from the early reading ability to the highest intended level of reading proficiency. The descriptor for A1 as the primary level of the reading ability reads as: “can understand familiar names, words and very simple sentence; for example on notes or posters or in catalogues.” (Council of Europe, 2001). And for the highest level of reading proficiency, it states: “can read with ease virtually all forms of the written language, including abstracts, structurally or linguistically complex texts....”

The self-assessment grid, in the CEFR self-assessment checklist, consists of all can-do statements benchmarking the proficiency descriptors from the CEFR. As Little (2005, p. 324) states: “the CEFR scales do not claim to model progression...they present a hierarchy of communicative tasks whose successful performance depends on underlying linguistic competence.”

