University of Tabriz

Faculty of Persian Literature and Foreign Languages

English Language Department

Dissertation submitted to the Faculty of Persian Literature and Foreign Languages in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

In

English Language Teaching

Title

Rater effects in self-assessment, peer-assessment, and teacher assessment: A multi-faceted Rasch measurement approach

Supervisor: Farahman Farrokhi (PhD)

Advisor: Parviz Azhide (PhD)

Researcher: Rajab Esfandiari

February 2012

In the name of God

University of Tabriz

Faculty of Persian Literature and Foreign Languages

English Language Department

Dissertation submitted to the Faculty of Persian Literature and Foreign Languages in fulfillment of the requirements for the degree of

Doctor of Philosophy

In

English Language Teaching

Title

Rater effects in self-assessment, peer-assessment, and teacher assessment: A multi-faceted Rasch measurement approach

Supervisor: Farahman Farrokhi (PhD)

Advisor: Parviz Azhide (PhD)

Researcher: Rajab Esfandiari

February 2012

University of Tabriz

Faculty of Persian Literature and Foreign Languages

English Language Department

We hereby recommend that the dissertation by Rajab Esfandiari

entitled

Rater effects in self-assessment, peer-assessment, and teacher assessment:

A multi-faceted Rasch measurement approach

be accepted in partial fulfillment of the requirements for the degree of

PhD in ELT

Supervisor: Farahman Farrokhi (PhD)

Advisor: Parviz Azhide (PhD)

Examiners:

Internal examiner:

Ali Akbar Ansarin (PhD)

External examiners:

Javad Gholami (PhD)

Biook Behnam (PhD)

To my parents with love

Acknowledgements

When I started a PhD in English Language Teaching at the University of Tabriz some four years ago, I would never have fancied ending up doing language testing. By a strange quirk of fate, many-facet Rasch measurement differs markedly from planned focus on form which I chose to work on for my MA thesis. The printed pages of this dissertation, therefore, hold far more than the culmination of years of study. I would like to take this opportunity to thank those people who, one way or the other, helped me hatch up the idea of Rasch measurement and without whose help this dissertation would have been a complete fiasco.

First and foremost, I would like to offer my sincerest gratitude and express my deepest appreciation to my respectable supervising professor, Dr Farrokhi, whose highly perceptive comments, insightful observations, resourceful guidelines, illuminating insights, enduring inspirations and constructive suggestions are indelibly imprinted in the individual pages of this dissertation. He was so enthusiastic about research issues and, continually and convincingly, conveyed a spirit of adventure in regard to research. Despite his very hectic schedule and immense managerial responsibility, he was very approachable and always accepted me into his office very willingly and kindly, never failing to offer me a hot, freshly brewed cup of tea. His invaluable pieces of advice, persistent help and unfailing support, words of encouragement and life-long academic experience are highly valued and fully appreciated.

I would also like to extend my thanks to my advisor, Dr Azhideh, who made available his support in a number of ways and spared me his precious time to discuss and share with me his immensely hands-on statistical experience which improved the quality of statistical arguments made in this dissertation. His timely provision of comments, valid recommendations and helpful suggestions, and constant assistance enriched the present dissertation to varying degrees. I had the privilege of being his PhD student in Language Testing course during which period he introduced me with new areas of language testing and methodological innovations employed to unpack issues in language testing including Rasch measurement.

Special acknowledgement must be made of the internal examiner, Dr Ansarin, and external examiners, Dr Gholami and Dr Behnam, whose meticulous reading and fair comments contributed substantially to this dissertation.

A deep sense of gratitude and special thanks are due to professor Myford, of the University of Illinois at Chicago, who assisted me with the many-facet Rasch measurement (MFRM) from inception to completion of this dissertation. She always encouraged me to move on, felt duty bound to keep me working and moving forward, and was so curious to be well-informed of my work progress. She was so kind, generous, supportive, patient, forgiving and caring, always being there to lend her toddler a hand, who had just heard MFRM and could barely articulate it. I do appreciate her care and attention, and I will always remember her in my life.

I am deeply indebted to Dr Schaefer, of the Ochanomizu University, who was my supervising professor when I was a visiting scholar at Ochanomizu University in Tokyo, Japan for a period of six months. He was very kind to let me share his experience in bias analysis, lend me his files of MFRM courses he had registered for on line, and encourage me to ask him for many other books and papers which he provided me with very generously.

I would be remiss if I failed to extend my heartiest appreciations to Dr Ansarin, the current head of the English Language Department, who taught me courses in ESP and Psycholinguistics in MA and PhD programmes. Tremendous complements should also go to Dr Sabouri, the former head of the English Language Department. Thanks are also due to

Dr Andrade, of the University at Albany; Dr Linacre, the Research Director of Winsteps.com; and Dr Baghaei, of the Islamic Azad University, Mashhad Branch.

Special thanks should be given to two of my best friends who have the attitude and substance of geniuses and who have always been readily accessible to help me with their computer knowledge: Dr Mahmoodi, of the University of Tabriz; Dr Vaez Dalili, of the University of Isfahan; and Mr Imam Ali Mirzaei. They have been my constant companions over the years, and I have always enjoyed their company.

Last but not least, I feel obliged to acknowledge my well-behaved students at Imam Khomeini International University in Qazvin, although words alone cannot express the thanks I owe to them.

Abstract

Surname: Esfandiari **First name:** Rajab

Title of Dissertation: Rater effects in self-assessment, peer-assessment, and teacher assessment: A Multi-faceted-Rasch Measurement approach

Supervisor: Farahman Farrokhi (PhD)

Advisor: Parviz Azhide (PhD)

Degree: Doctor of Philosophy (PhD)

Major: English

Field: English Language Teaching

University: University of Tabriz

Faculty: Faculty of Persian literature and foreign languages

Dissertation pages: 179

Graduation Date: 17/11/1390

Key words: Many-facet Rasch measurement; self-assessment; peer-assessment; analytic rating scale; rater effects; rater-mediated second language assessment

Abstract

In performance assessment in general and in second language performance assessment in particular, raters are expected to assign ratees ratings, using a rating scale. In the process of rating, raters may commit some errors. These unwanted, rater-dependent sources of variability, which are unrelated to the students' abilities and manifested in various ways, could endanger the fairness and validity of decisions made based on the assigned ratings. Rater errors contribute to construct-irrelevant variance, and, if they are not detected and treated appropriately, they may result in obscuring an examinee's score and threaten the validity and fairness of second language performance assessment. As a result, they deserve further new research and investigation in second language performance assessment. Rater effects—severity/leniency effect, bias effect, central tendency effect and halo effect—have been more or less researched either in L1 or in L2, but rarely has any single study striven to address these effects in self-assessment, peer-assessment, and teacher assessment. The present study is an endeavor to employ a multi-faceted Rasch model (MFRM), a relatively newly developed measurement model, to detect these errors in three types of assessment: self-assessment, peer-assessment, and teacher assessment. To that end, 194 assessors—188 self-assessors and peer-assessors and six teacher assessors—were employed to assess 188 essays written by Iranian English majors at two-state run universities in Iran, using a 6-point analytic rating scale. The data were collected and analyzed, using Facets 3.68.1. to answer the research questions. The results of the MFRM analysis showed that of the three assessor types, teacher assessors were the most severe while self-assessors were the most lenient, although there was a great deal of variability in the levels of severity that assessors within each type exercised. MFRM also revealed differing patterns of severity and leniency among the three assessment types. For example, self-assessors and teacher assessors showed the opposite pattern of severity/leniency as peer-assessors toward the highest and lowest ability students. The results of further Facets analysis showed that the three types of assessor did not exhibit any sign of centrality either at group level or at individual level. Finally, Facets analysis showed that, at group level, the assessors did not exhibit any sign of halo effect, but, at individual level, all assessor types displayed considerable halo effect. Further analysis revealed that assessor types were unanimous about halo effect on four items, and that self-assessor showed more of a halo effect compared to the other two assessor types. The present study has possible implications for rater training, concurrent validity of peer ratings, and construction of rating scales. Most importantly, though cautiously, peer-assessors could be employed as an alternative to teachers for rating purposes.

List of Tables

List of Figures

List of Abbreviations

MFRM	Many-facet Rasch measurement
Df	Degree of freedom
L2	Second language
L1	First language

For correspondence with the researcher:

E-mail: raesfandiari@gmail.com

Refereed journal publications based on this dissertation:

1. Farrokhi, F., & Esfandiari, R. (2011). A Many-Facet Rasch Measurement Model to Detect Halo Effect in Three Types of Raters. *Theory and Practice in Language Studies*, *1*(11), 1531-1541.

2. Farrokhi, F., Esfandiari, R., & Schaefer, E. (to be published in September 2012). A Many-Facet Rasch Measurement of Differential Rater Severity/Leniency in Self-assessment, Peer-assessment, and Teacher assessment. *Journal of Applied and Basic Scientific Research*.

3. Farrokhi, F., Esfandiari, R., & Vaez Dalili, M. (2011). Applying the Many-Facet Rasch Model to Detect Centrality in Self-assessment, Peer-assessment and Teacher assessment. *World Applied Science Journal*, *15* (11), 76-83.

4. Farrokhi, F., Esfandiari, R., & Schaefer, E. (to be published in May 2012). A Many-Facet Rasch Measurement of Differential Rater Severity/Leniency in three types of assessment. *JALT Journal*.

5. Esfandiari, R., Myford, C., & Farrokhi, F. (under review). Severity Differences among Self-assessors, Peer-assessors, and Teacher Assessors Rating EFL Essays. *Assessing Writing*.

National and international conference presentations based on this dissertation:

1. Esfandiari, R., & Farrokhi, F. (2011). A Many-Facet Rasch Measurement Approach to Rater Effects. Paper Presented at the Ninth Annual International TELLSI Conference, Illam, 2011.

2. Esfandiari, R., Schaefer, E., & Farrokhi, F. (2011). Detecting Rater Effects in Three Types of Assessment. Paper Presented at the Tenth Annual JALT PAN-SIG Conference, Matsumoto, Nagano, 2011.

Table of contents

CHAPTER One:

INTRODUCTION