UNIVERSITI TEKNOLOGI MARA

COMPARING DIFFERENT ITEM THRESHOLDS FOR RASCH RATING SCALE MODEL

SHOLEH ATAEI

Thesis submitted in fulfillment
of the requirements for the degree of
Doctor of Philosophy

Faculty of Computer and Mathematical Sciences

March 2018

ABSTRACT

Measurement issues are old issues in educational testing, and some of them are still unsolved. The Rasch rating scales model is a unidimensional measurement model for analyzing data on human attitudes. This mathematical model describes the two parameters: items and persons as agents of response. This study investigates the measurement of rating responses (attitude items) for the Rasch rating scales using item threshold models. If measure is merely based on the graphical distribution of the Wright map without considering the item thresholds, the item measure does not reflect the difficulty of achieving a particular category in the rating scale. Although previous studies have described and interpreted the rating scale category functioning and Wright map for the rating scales model, only few have described the importance of thresholds in calibrating the measurement variables of the rating scales. The mathematical and practical concepts of the measurement must be examined further to address these issues. Accordingly, this study investigates and estimates the consistency of the rating scale model categories and the model calibration between the parameters of measurement/latent variables in a particular category based on the Rasch mathematical functions and real data. The data are gathered by conducting a questionnaire survey among 342 engineering students taking up Calculus III at the Universiti Teknologi MARA in Shah Alam. The students were asked to answer 20 revised Aiken attitude items using five response options ranging from strongly disagree to strongly agree. A questionnaire was developed and subjected to a pilot test, and its content validity, construct validity, and predictive validity all complied with the validity measurement guidelines. The data are investigated using three item types of thresholds, namely, the Rasch–Andrich threshold, Thurstone threshold, and Rasch-half-point threshold, based on mathematical functions and graphics. They display category curves and maps including conditional probability, cumulative probability, expected score curve, and Wright map based on three types of item thresholds. The analysis focuses on the conditions of the items in the threshold models to verify the quality of responses in each category and distribution of each response. The findings reflect the quality of responses at each category scale based on the estimated location of agreement–disagreement points on the latent variable. The Wright map illustrates such quality by calibrating the distribution of students and items based on the logit scales and by comparing the three item thresholds. The item thresholds model provides three different methodologies for selecting the point. Although the difference in their results is usually small but each one radically changes the interpretation. Therefore, it would be helpful for users to know which method is suitable for their targets. This research significantly contributes to the fundamental measurement of human attitudes by showing that these measurement models can demarcate the category intervals based on the rating scale responses at three types of item thresholds.

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to many people for their Support in helping me accomplish this study.

Firstly, I would like to thank my supervisor Assoc. Prof. Dr. Zamalia Mahmud for her constant encouragement, invaluable suggestions and expert guidance throughout the study. I would also like to thank my co-supervisors Assoc. Prof. Dr. Mohd Nasir Khalid and some lectures from computer and mathematical faculty for collecting data in this study.

I would like to deeply thank my family: Mom and my sister for all the love and moral support to make my dreams come true.

TABLE OF CONTENTS

CHAPTER ONE
INTRODUCTION

1.1 OVERVIEW

This chapter discusses the background of the study in relation to the Rasch Model. An overview of the measuring attitude rating scale model using Rasch measurement is also presented with the statement of the problem and the purpose of this research. The research objectives, significance, limitations, and some definitions or new terms of the studies are described as well.

1.2 BACKGROUND

Most researchers are basically concerned about the quality of test items and how examinees respond to it. In science, measurement is used as a tool to be utilised for a better evaluation of variables .Undoubtedly, measurement is a science of the new century, and as this science is more applicable, it will have a better place among other sciences and among those who need to develop psychological measurement.

The development of science relies on measurement (Green & Frantom, 2002). According to Wu (2007), there are many challenges about measurement.

Researchers have done some studies about teaching and learning (Ayob & Yassin, 2017; Clements & Sarama, 2014; Kelly et al., 2014; Van den Heuvel-Panhuizen & Drijvers, 2014). Burns (1995) conceived learning as a stable change in behaviour that consists of observation and intervention process like thinking and attitudes.

Assessment is a systematic collection process to validate of information in education program for developing students learning (Freeman et al., 2014; Mayer & Alexander, 2016; Schoenfeld, 2009). Assessment is a process that focuses on learning of students, hence it reflects to students' performance. Most importantly, assessment can show when students learn and how their attitudes are (Biggs, 2012; Brown et al., 1994).