Applying the Item Response Theory to English Classroom Examinations for Ethnic Students

Lanfen Ji¹, Xiaoqin Zhang², Dianjun Lu³, Dianxiang Lu⁴*

¹Foreign Language Department of Qinghai Normal University, Xining 810008, China
²Financial and Economic College of Qinghai University, Xining 810008, China
³Department of Mathematics of Qinghai Normal University, Xining 810008, China
⁴Medical College of Qinghai University, Xining 810001, China

Abstract

Tests and assessment are necessary for the presentation of the language learners’ competence in the process of learning. This paper discussed and is trying to find out the particularity and characteristics of English teaching at minority regions in consideration of the influencing factors: trilingual teaching, small classes, fewer students. The measurement of constructs of English classroom examination at minority regions is put forward. The techniques used in this paper are trying to keep pace with the advances in psychometric theory and methods. This paper presents an example of applying item response theory to English classroom examinations at minority regions and demonstrates graded response models.

Keywords: English Classroom Examination, Item Response Theory, Graded Response Models;

1. Introduction

Language learners need to be tested or assessed and will be presented with a set of the competence that the learners have in the language study. When students are assessed, both their competence and performance in the use of language are focused. Just like in other subjects, language tests are given with a variety of purposes in mind. However in a second language situation, two major purposes can be identified and these are (a) to determine how much the learner has learnt from the planned syllabus. This is called Achievement testing. (b) to determine the strength and weakness found in the language used by the students. This is known as diagnostic testing.

The purpose of writing this paper is trying to find out the similarities and dissimilarities of teaching English in regular college classrooms and in classrooms at minority regions. English learning characteristics of college students are: trilingual teaching, small classes, fewer students. How to test English classroom teaching effect is particularly challenging for faculty. The analyses of most classroom examinations suffer because of small sample sizes. In contrast to majority students, students at minority regions are multi-faceted, having more diverse educational experiences. Some groups of students are better prepared academically than other groups. Other students are not as prepared owing to not being able to graduate on time or they just were not sure what to do with their lives. Teachers who sample classroom examinations are challenged because students’ knowledge, skills, and attitudes vary. It is important for teachers to find out whether the variance is caused by differences in exami-
nation difficulty or student ability. The
diversity of student abilities decides test
equating using standard statistical meth-

The most notable development in the
last forty years is item response theory
(IRT). IRT is a family of techniques for
understanding the psychometric proper-
ties of measures and relationships be-
tween properties of the measures and the
individuals completing those measures.
These techniques have been used to ad-

The relationship between the estimate
of the latent trait, the response option
characteristics, and the probability of se-
lecting a particular option is presented
graphically with an option characteristic
curve (OCC). This item is rated on a six-
point scale with the anchors of ‘strongly
disapprove’ and ‘strongly approve’. The
estimate of the latent trait is along the x-
axis and the probability of selecting a par-
ticular response option at a given level of
the latent trait is along the y-axis. In IRT,
the latent trait is referred to as theta
(i.e., \( \theta \)). The values of theta are expressed
as standardized scores (i.e., z-scores).

2. Graded response models

In most studies on language learning, au-
tors usually aim at attitudinal, and per-
sonality research, attitudinal, and per-
sonality research, the measures utilize re-
sponse options that are ordered (e.g., Li-
kert-type response options) and the data
are used to make conclusions concerning
the level of the construct that is assessed
by the measure. In these cases, graded
response models can be used. In these
models, the IRT analysis examines the
relationships between item or option pa-
rameters, person parameters, and the se-
lection of a particular option. For cumula-
tive graded response models, it is as-
sumed that the value of the latent trait is
smaller for individuals who choose the
first response option than for individuals
who choose the second response option in
an ordered response set. This assumption
exemplifies the dominance response
process that underlies these models.

Samejima’s (1969) graded response
model is a particular cumulative graded
model that is often used in organizational
research (see van der Linden & Hamble-
ton, 1997, for a discussion of other
graded response models). In Samejima’s
(1969) graded response model (GRM),
two parameters associated with the items
are estimated. The first is an option diffi-
culty parameter. The difficulty parameter
is referred to as the ‘threshold’ param-
eter. This refers to the probability of an
individual with a given level of the latent
trait selecting a given option (e.g., disap-
prove) or any of the subsequent higher ordered options (e.g., neutral, approve, and strongly approve). Specifically, this parameter is the point on the theta scale where there is a 50% chance that a given option or a higher ordered option will be selected (i.e., \( P(\theta) = 0.50 \)). In other words, this parameter represents the thresholds between the response options. The second parameter is the discrimination parameter. This parameter represents how well an option discriminates between individuals at different levels of the latent trait. The larger the value, the better the option is at discriminating between individuals at different levels of the latent trait.

To estimate the OCCs, one first estimates the boundary response functions. Boundary response functions are the cumulative probability of selecting a response option equal to or higher than the current response option. Option difficulty parameters are estimated for \( m_i - 1 \) boundary response functions where \( m_i \) equals the number of response options. Each boundary response function has a difficulty parameter, but only one discrimination parameter is estimated for each item in Samejima’s (1969) GRM. Therefore, on a leadership measure with six response options, five difficulty parameters and only one discrimination parameter are estimated. The boundary response functions are used to estimate the OCC. Mathematically, the boundary response function is expressed as,

\[
P_{ik}^*(\theta) = e^{D_{ij}(\theta-b_k)}/1 + e^{D_{ij}(\theta-b_k)}
\]

where

\[
P_{ik}^*(\theta)
\]

is the probability of a respondent at a particular level of theta responding to option \( k \) or any of the other higher ordered options on item \( i \), \( b_{ik} \) is the option difficulty parameter, \( a_i \) is the discrimination parameter, \( D \) is a scaling constant equal to 1.702, and \( e \) represents an exponential function. Thus, the probability of selecting option \( k \) is a function of the level of the latent trait, the difficulty of the option, and the discrimination. In essence, the boundary response functions are estimated by utilizing a two-parameter logistic IRT model on the response option data (see Hambleton et al., 1991 for a discussion of logistic models).

From the boundary response functions, the probability of selecting a particular option and the OCCs are estimated. The probability of selecting a particular option \( P_{ik}^*(\theta) \), is determined by subtracting the boundary response functions for each option. Mathematically, this is represented as

\[
P_{ik}^*(\theta) = P_{i(k-1)}^*(\theta) - P_{ik}^*(\theta)
\]

where the probability of selecting an option is a function of the conditional probability of responding above the threshold parameter (i.e., \( b_{ik} \)) for option \( k-1 \) minus the conditional probability of responding above the threshold parameter for option \( k \). For example, on a four option item \( P_{ik}^*(\theta) \) is as follows,

\[
P_{i0} = 1 - P_{i1}^*(\theta)
\]

for option 1,

\[
P_{i1} = P_{i1}^*(\theta) - P_{i2}^*(\theta)
\]

for option 2,

\[
P_{i2} = P_{i2}^*(\theta) - P_{i3}^*(\theta)
\]

for option 3,

\[
P_{i3} = P_{i3}^*(\theta) - 0
\]

for option 4.

The probabilities that are computed for each response option at each level of \( \theta \) serve as the basis for the OCC.

3. ACKNOWLEDGMENTS

This paper was undertaken while the author was a visiting scholar at the University of California, Riverside. It was sponsored by Western Light Talent Culture Project supported by Chinese Academy of Sciences (Grant No.2009(236)).

4. REFERENCES


