Article History
Received: 05.12.2022
Received in revised form: 01.03.2023
Accepted: 07.03.2023
Article Type: Research Article

International Journal of Contemporary Educational Research (IJCER)

www.ijcer.net

A Mixture Rasch Model Analysis of Data from a Survey of Novice Teacher Core Competencies

Türker Toker[1], Kent Seidel[2]
[1]Usak University, 0000-0002-3038-7096
[2] University of Colorado Denver, 0000-0002-7469-2105

To cite this article:

Toker, T. & Seidel, K. (2023). A mixture rasch model analysis of data from a survey of novice teacher core competencies. *International Journal of Contemporary Educational Research, 10*(1), 147-156. https://doi.org/10.33200/ijcer.

International Journal of Contemporary Educational Research

Volume 10, Number 1, March 2023, Page 147-156 ISSN: 2148-3868

A Mixture Rasch Model Analysis of Data from a Survey of Novice Teacher Core Competencies

Türker Toker*[1], Kent Seidel[2]

[1]Usak University, [2]University of Colorado Denver

Abstract

Although the Rasch model is used to measure latent traits like attitude or ability where there are multiple latent structures within the dataset it is best to use a technique called the Mixture Rasch Model (MRM) which is a combination of a Rasch model and a latent class analysis (LCA). This study used data from a survey for teachers, teacher candidates, and teacher education program faculty with a sample of 296 candidates, 648 graduates, and 501 program faculty. Survey items based on these competencies were administered to in one Western state to ascertain how well the program candidates attended prepared them for the teaching profession. The 40 items common to surveys of the three groups were submitted to mixture Rasch analysis to determine whether distinct patterns of item response were discernible. Analyses yielded two classes which brings the construct validity of the survey into question. Results showed that the Mixture Rasch Model is and can be useful to determine sub-groups for survey researchers. This research presents a demonstration of usefulness of the Mixture Rasch Model for the analysis of survey data.

Keywords: Mixture rasch model, Validity, Survey research, Teacher effectiveness

Introduction

The mixture Rasch model was first introduced by Rost in 1990. The model was proposed to create a combination of the Rasch model with latent class analysis. Main assumption is that Rasch model holds for all participants within a latent class, but it allows for different sets of item parameters between the latent classes (Rost, 1990). As a result of this, the model might be applied to validate responses to a test or questionnaire. Since the Rasch model has some strict item and homogeneity assumptions, the MRM becomes useful when some item and population homogeneity assumptions are relaxed. If it is known that there are heterogeneous structures in the population, a single population assumed statistical model might not produce valid results (Sen, 2016).

Basically, mixture Rasch models are a combination of two models: a Rasch model a latent class analysis model (Kaiser & Keller, 2001). One main advantage of this useful combination is unlike the quantitative information provided by Rasch models, mixture Rasch models supply information about quantitative and qualitative structures within the dataset. If there are multiple latent structures within the dataset traditional IRT models may produce biased results. However, the mixture Rasch model overcomes this issue and becomes handful. Mixture Rasch models can detect participant heterogeneity and the related item structures, the size of latent classes, and the latent score distribution (Baghaei & Carstensen, 2013). MRM not only detects the subclasses based on qualitative information but also invastigates quantitative information about the sample and the scale used for data collection (Sen, 2016). Another advantage of MRM is when researchers think that participants use different strategies or there are instructional differences, curriculum etc. or a model including additional factors of quantitative differences within strategies MRM becomes more useful (Toker & Green, 2021).

MRM is primarily utilized to investigate individual variations in strategy usage and knowledge disparities, as well as to explore unidimensionality and validation (Baghaei & Carstensen, 2013). When data fails to conform to the typical unidimensional Rasch model, MRM identifies the subgroups or latent classes within the entire sample that adhere to the Rasch model. These subgroups are, in essence, the latent classes that differ qualitatively. MRM identifies subgroups with DIF, rather than conducting DIF across manifest variables such as sex or native language (Pishghadam et. al., 2017). Once latent classes are identified, the content of the items must be scrutinized to determine the nature of the qualitative distinctions between the classes that caused the DIF. The subsequent step

* Corresponding Author: *Turker Toker, turker.toker@usak.edu.tr*

is to examine the nature of the classes and establish how they vary. In other words, class membership should be associated with some qualitative differences in test-takers. Such differences may indicate vital factors in the learning process and may contribute to substantive theories (Baghaei & Carstensen, 2013).

Additionally, if a model with two or more latent classes provides a better fit than a model with only one latent class, the measurement invariance assumption is violated and a single Rasch model is not a fit. When there are more than one latent classes in the dataset, separate Rasch models with separate sets of item difficulties are required. These different item difficulties are considered latent sub-groups where they are not determined by covariates (Frick, Strobl, & Zeileis, 2015).

The Mixture Rasch Model

Mislevy and Verhelst (1990) and Rost (1990) used the dichotomous MRM as a first application of Mixture Item Response Theory (MixIRT) model to explore unobserved heterogeneity in a population by revealing different latent subgroups of individuals based on their observed response patterns. Dallas and Wilse (2014) used the Mixture Rasch Model to demonstrate all stages of model estimation and selection, description of model results, and follow-up analyses using real survey data. It was expected that the MRM would identify classes of individuals who have different response patterns on the survey. Mixture IRT models possess an advantage over IRT models in that they offer not only quantitative data but also qualitative data pertaining to both the items and the test-takers, in contrast to IRT models which only provide quantitative information (Sen & Toker, 2021). Following mathematical formulas used by Rost (1990) to explain the proposed model.

If ρ_{vig} indicates person *v* answering "yes" or correctly answering item *i* and this person belongs to latent class *g* then one can say that subjects' response probabilities can be shown by the dichotomous Rasch model

$$p_{vig} = \frac{\exp(\tau_{vg} + \sigma_{ig})}{[1+\exp(\tau_{vg} + \sigma_{ig})]}, \tag{1}$$

where τ_{vg} is the participant's ability and $\sigma_{ig}$is the item easiness parameter. Within each latent class g an indeterminancy constraint $\sum_i \sigma_{ig} = 0$ must hold. Furthermore, if the researcher thinks latent classes are mutually exclusive and exhaustive, structure of the latent class is as follows:

$$p_{vi} = \sum_g \pi_g \, p_{vig}$$

$$\sum\nolimits_g \pi_g \, \frac{\exp(\tau_{vg} + \sigma_{ig})}{[1+\exp(\tau_{vg} + \sigma_{ig})]} \tag{2}$$

where p_{vi} is the unconditional response probability and π_g is the class size parameter or "mixing proportion with constraints between 0 and 1 and $\sum_g \pi_g = 1$.

To define the entire model since simpler models do not explain and specify how to deal with the person parameter, τ_{vg} , it is crucial to control the person parameters using a Rasch-like model structure. To get the likelihood function, it is also important to get the pattern of probabilities *p*(x) which is x = (x_1, x_2, ... x_i …. x_k) where x_i = 0 or 1. So the formula for the pattern probability can be:

$$p\ (\mathrm{x}) = \sum\nolimits_{\mathrm{g}} \pi_{\mathrm{g}}\ \mathrm{p}(\mathrm{x} \mid \mathrm{g}) \tag{3}$$

where *p* (x | g) is the product of response probabilities defined by Equation 1 over all items. In the Rasch model the number of correct item responses is used to estimate τ. So, all persons with the same score *r* have the same τ score. As a result of this, the pattern probability *p* (x | g) can be rewritten with the score *r* associated with a given pattern as follows

$$p\ (\mathrm{x} \mid \mathrm{g}) = \ p\ (\mathrm{x} \mid \mathrm{g, r}) \bullet p\ (\ \mid \mathrm{g})\ . \tag{4}$$

This factorization is quite important and useful since only the first factor depends on the item parameters σ_{ig},

$$p\ (\mathrm{x} \mid \mathrm{g}, \mathrm{r}) = \exp(\textstyle\sum_i x_i \sigma_{ig})\ /\ \Phi_r\,[\exp(\sigma)]\ . \tag{5}$$

In this formula, Φ values are the symmetric functions of order r of the delogarithmized item parameter values. Moreover, only the second factor depends on the ability distribution in class g. The MRM is also a "distribution free" model just like the simple Rasch model.

A combination of all these elements defines the likelihood function of the model as follows;

$$L = \textstyle\prod_x \{\sum_g \pi_g \pi_{rg} \exp(\sum_i x\ \mathrm{i}\sigma_{ig})\ /\ \Phi\mathrm{r}\ [\exp(\sigma)]\}^{\ n(\mathrm{x})}\ , \tag{6}$$

where $n(\mathrm{x})$ denotes the observed number of response patterns x, and the score probabilities $\pi_{rg} = p\ (\mathrm{r} \mid \mathrm{g})$ have been rewritten by using Greek letters for renaming the model parameters.

Using the model formulas, the number of independent model parameters is constructed as follows:

> *h - 1* class size parameters π_g, where *h* is the number of classes,
> (*k* - 1) *h* class-specific item parameters, where k is the number of items measured and must be lowered by 1 because of the norming constraint, and
> 2 + *h* (*k* - 2) class-specific score probabilities, because one parameter in each class depends on the sample size and the class size, and the two parameters for the 0 and 1 vectors are class independent

As it can be seen from Equation 1, the Rasch model is a one-class solution of the proposed model. Additionally, the same Equation is a special case of a simple latent class analysis.

Professional Preparation Survey

Research has shown that teachers play a critical role in enhancing student achievement (Goe, 2007; Hanushek & Rivkin, 2010; National Research Council, 2010). This finding has been made possible through better asse ssments, P-12 standards, data systems, and statistical analyses such as growth and value-added modeling. It has been established that effective teachers improve student achievement, while ineffective teachers negatively impact students, sometimes for several years (Sanders & Horn, 1998; Sanders & Rivers, 1996). This result has propelled research and policy towards the next step: how do we ensure that all teachers are effective, and how do we support them in developing the "sophisticated expertise" (Darling-Hammond & Bransford, 2005, p. 3) that characterizes outstanding teaching?

A grant supported by the Institute for Education Sciences (Award #R305A120233) supported development of assessments of preparation of teachers. As one of the early steps in this study, surveys were created to assess perceptions of preparation from the perspectives of teacher candidates, recent graduates of teacher preparation programs, and faculty members who taught in the teacher preparation programs. The purpose of this paper was to examine whether responses to items reflecting perceptions of preparation yielded different item positions across groups of survey respondents and so led to the existence of distinct latent classes. The existence of latent classes suggest that the variable takes a different definition for different groups of respondents.

This grant began with the creation of Core Competencies (CCs) or competencies considered essential for effective teaching. The survey examined here is based on the final CC's. To identify Core Competencies (CCs), documents regarding national teacher standards were examined. These included: The Interstate Teacher Assessment and Support Consortium (InTASC), The National Council for Accreditation of Teacher Education (NCATE), which is now CAEP, The National Board for Professional Teaching Standards (NBPTS), The Teacher Education Accreditation Council (TEAC), and the exam elements of the Praxis II, which is a national teacher certification test. In all, 16 sources were analyzed and represented in a matrix of the content of teacher preparation selected on the dual justification that:

1) Policy licensure and accreditation restrictions are calling for these CCs in order to teach; and

2) Programs are required to provide some evidence of how these CCs are incorporated into their program to achieve accreditation / licensure approval.

This initial mapping identified 12 potential CCs, each of which appeared in at least three of the 16 national or state sets of licensure/accreditation standards and policy recommendations. In order to focus the study, the initial 12 CCs were narrowed based on existing research and whether the CC is likely to be taught in the program (rather than being a selection criterion), is variable among programs, is observable, and is regularly employed in schools. The 12 potential CCs were grouped into 8 CCs that were considered to have less overlap, with vignettes written for each with 5-6 descriptors that would form the basis for survey items. These eight areas became: demonstrating mastery of and pedagogical expertise in content taught; managing the classroom environment; developing a safe, inclusive, respectful environment for a diverse population of students; planning and providing instruction; designing and adapting assessments, curriculum and instruction; engaging student in higher order thinking and expectation; supporting academic language development and English language acquisition; and reflection and professional growth. Details of these core competencies with descriptors can be found in Appendix A. No information was available on whether surveys would measure similar constructs for all groups. Briggs et al. (2013) analyzed data from two of the three surveys, and concluded that different approaches to examining dimensionality yielded different conclusions about program effects.

The present study examined the questionnaire validation to see if collected data yielded different sub-groups within the selected groups. The main goal of this study is to present a demonstration of usefulness of the Mixture Rasch Model for the analysis of survey data.

Method

Participants

Table 1 provides a comprehensive overview of the traits exhibited by the three groups of participants. However, it should be noted that not all variables were gathered from all participants, partly due to confidentiality concerns. The majority of respondents who were candidates or graduates were young, white females from a conventional teacher education program. Most of the faculty members who responded had full-time involvement with the program. The survey garnered responses from 296 candidates, 648 graduates, and 501 program faculty. These faculty members were drawn from teacher preparation university faculty and field supervisor teaching faculty responsible for overseeing clinical teaching preparation experiences.

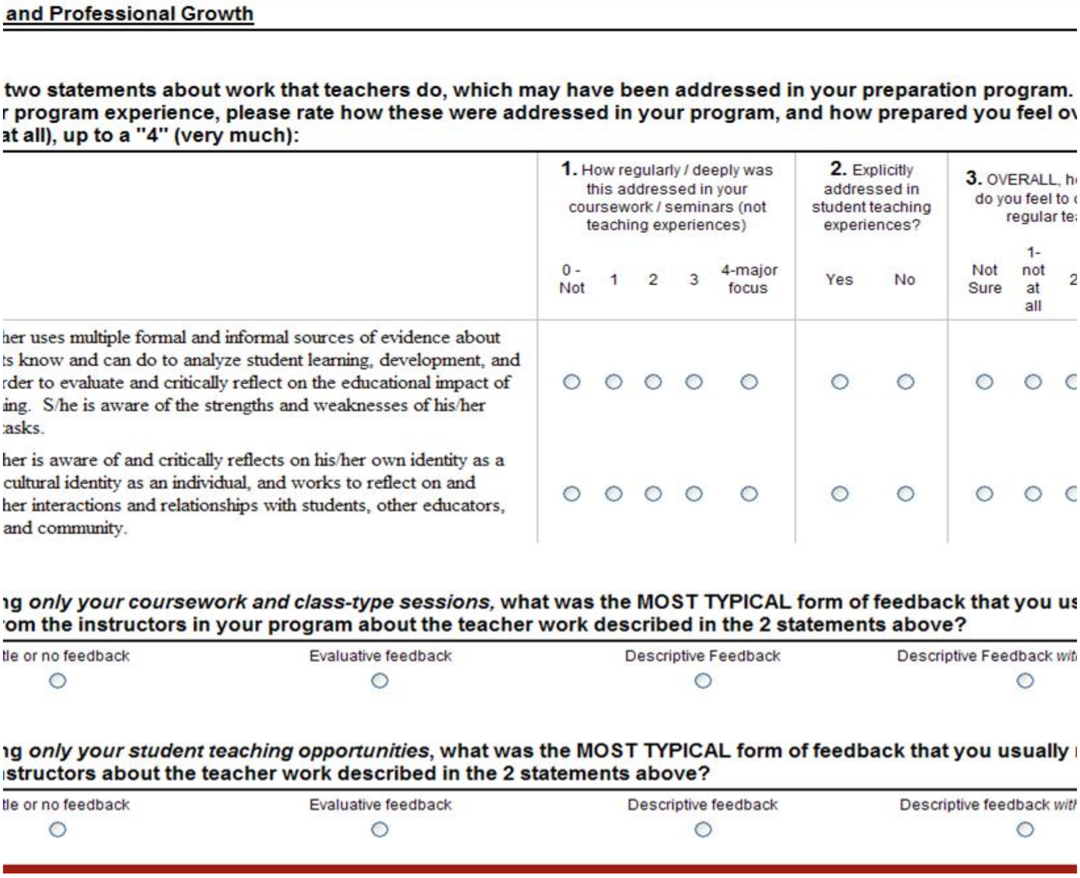
Table 1
Gender of TIMSS-2011 Subjects (based on booklet selection)

Group	Gender			
	Female		**Male**	
	Selected	**%**	**Selected**	**%**
Teacher Candidates	49.40	49	50.60	51
Program Graduates	49.10	49	50.90	51
Program Personnel	50.30	48	49.70	52

Note: Gender is shown in percentages.

Measure

The survey, as described above, was created via literature review and a comprehensive analysis of sources of standards for teacher preparation, to define eight competency areas (Hartnett-Edwards, Seidel, Whitcomb, Spurlin, Anderson, Green, & Briggs, 2013) with one additional area suggested by an advisory panel. Items were written by project personnel and vetted through teacher education program directors and a regional advisory panel. After modifications based on a series of cognitive interviews, the survey was approved by a panel of deans of colleges of education in the state. Figure 1 provides an example of a partial survey page from one of the three surveys. Further detail about the survey development and sample surveys can be found at the study's online site, www.portfolio.du.edu/IES.



and Professional Growth

two statements about work that teachers do, which may have been addressed in your preparation program.
r program experience, please rate how these were addressed in your program, and how prepared you feel ov
at all), up to a "4" (very much):

	1. How regularly / deeply was this addressed in your coursework / seminars (not teaching experiences)					2. Explicitly addressed in student teaching experiences?		3. OVERALL, h do you feel to regular te		
	0 - Not	1	2	3	4-major focus	Yes	No	Not Sure	1- not at all	2
her uses multiple formal and informal sources of evidence about ts know and can do to analyze student learning, development, and rder to evaluate and critically reflect on the educational impact of ing. S/he is aware of the strengths and weaknesses of his/her asks.	○	○	○	○	○	○	○	○	○	○
her is aware of and critically reflects on his/her own identity as a cultural identity as an individual, and works to reflect on and her interactions and relationships with students, other educators, and community.	○	○	○	○	○	○	○	○	○	○

ng only your coursework and class-type sessions, what was the MOST TYPICAL form of feedback that you us
om the instructors in your program about the teacher work described in the 2 statements above?

tle or no feedback	Evaluative feedback	Descriptive Feedback	Descriptive Feedback wit
○	○	○	○

ng only your student teaching opportunities, what was the MOST TYPICAL form of feedback that you usually
structors about the teacher work described in the 2 statements above?

tle or no feedback	Evaluative feedback	Descriptive feedback	Descriptive feedback with
○	○	○	○

Figure 1. Detail of a survey page.

The body of the survey for teacher candidates was split into nine sections, with each section eliciting views about an area of teaching competency. Each section contained three response scales oriented to appraisal of views about how extensively the content reflected in the item was covered in teacher preparation coursework-style experiences, how intensively it was addressed in teacher field-based experiences (e.g., student teaching placements), and perception of level of preparation overall. In total, the survey of teacher candidates contained 111 attitude items, 40 of which reflected overall satisfaction with the program. For additional details on these two surveys, see Briggs et al. (2013) or the study's site www.portfolio.du.edu/IES.

The survey of teacher education program faculty contained 51 items. One item asked about extent of involvement with the program and the remaining items asked "OVERALL, how well does the program prepare candidates to:" where the remainder of the statement was taken from the wording for the candidate and graduate surveys. Only the 40 items common to surveys of all three groups of respondents were used in the present study.

Procedure

The project staff generated the online surveys, consent forms, and email instructions to access the survey. This information was sent to directors of teacher preparation programs in the state. Directors of the teacher preparation programs sent a link to the survey via email to program teacher candidates with a request to complete the survey, as data were to be used for program evaluation and accountability purposes. In addition, project staff pulled publicly available district-school emails for 897 graduates which located recent programs' graduate placements in public school posts. Directors of teacher preparation programs were also sent a link to the personnel survey with a request to convey the survey to their faculty and to mentors and lead teachers associated with the program. The surveys were open from May 2012 through November 2012. Potential participants had approximately three months to respond. Qualtrics (Qualtrics.com) was used as the online survey site; when the survey was closed, data were downloaded as an Excel spreadsheet and transferred into a statistical software package. As the survey invitations were sent by individual program directors and not by the project staff, accurate response rate information is not available. However, response rates of surveys of program personnel ranged from approximately 20% to close to 100% for different programs.

Data Analysis

The responses were used to run the mixture Rasch model analysis using WINMIRA (von Davier, 2001a). Although, items had four levels, they were recoded into dichotomous responses due to WINMIRA not being able to handle the analysis. Since the data were sparse, competing models were selected by means of information criterion values which were the Pearson Chi-square value and Cressie-Read statistic (Cressie & Read, 1984). Information criteria used in this study were the Pearson Chi-square value and Cressie-Read where larger values show better fit (von Davier, 2001b). Once the latent classes were identified, item fit was examined.

Number of latent classes

To find the appropriate number of latent classes, competing models with one, two, three, and four latent classes were fit to the data for the survey. Table 2 shows p-values of the Pearson Chi-square and Cressie-Read for the four models. Table 2 suggests that there were two latent subgroups based on larger values of the Pearson Chi-square and Cressie-Read. Results showed that the mean of the raw scores of class 1 was high (M=31.88 SD=9.26), class 2 was medium (M=27.45, SD=11.10). It is crucially important to compare item parameters across different classes when deciding on number of classes.

Table 2
p-values of Model Fit Indices for the MRM

Model	Cressie Read	Pearson X^2
1-Class	0.00	0.00
2-Class	**0.08**	**0.13**
3-Class	0.00	0.17

The final dataset consisted of 40 items with 576 participants. To determine the appropriate number of classes, one, two and three class solutions were fit to the data (see Table 2). P-values for the dataset of Cressie-Read and Pearson Chi-square were .08 and .13. Since the two class model had the appropriate p-value, a two-class solution was selected. Class size values for each class shows that class 1 was expected to include about 72% of the sample. Class 2 was expected to include about 28% of the sample. According to the Q-index, there was no need to remove any items since all of the items fit each class well (.05 < p < .95) (See Table 3). To obtain additional information about the survey instrument or the IES grant project, readers can contact co-author Kent Seidel.

Table 3
Item fit assessed by the Q-index for all classes

	Class -1			**Class -2**		
Item	**Q-index**	**Zq**	***p* (X>Zq)**	**Q-index**	**Zq**	***p* (X>Zq)**
I1A	0.13	0.49	0.31	0.23	1.84	0.03
I1B	0.12	0.48	0.32	0.13	-0.43	0.67
I1C	0.12	0.19	0.42	0.11	-0.86	0.81
I1D	0.12	0.62	0.27	0.17	0.47	0.32
I1E	0.15	1.93	0.03	0.20	1.26	0.10
I2A	0.11	0.19	0.42	0.25	1.52	0.06
I2B	0.08	-0.49	0.69	0.23	1.65	0.05
I2C	0.07	-0.57	0.72	0.22	1.08	0.14
I2D	0.08	-0.40	0.66	0.13	-0.43	0.67
I2E	0.09	0.03	0.49	0.17	0.56	0.29
I3A	0.11	-0.06	0.52	0.12	-0.40	0.66
I3B	0.08	-0.50	0.69	0.11	-0.85	0.80
I3C	0.09	-0.21	0.59	0.13	-0.35	0.64
I3D	0.10	-0.04	0.52	0.17	0.37	0.35
I3E	0.11	0.63	0.26	0.17	0.65	0.26
I4A	0.09	-0.21	0.58	0.18	0.57	0.29
I4B	0.09	-0.23	0.59	0.15	-0.03	0.51
I4C	0.07	-0.67	0.75	0.06	-1.78	0.96
I4D	0.08	-0.41	0.66	0.14	-0.26	0.60
I4E	0.10	0.07	0.47	0.13	-0.32	0.63
I5A	0.09	-0.13	0.55	0.13	-0.07	0.53
I5B	0.07	-0.75	0.77	0.15	0.24	0.41
I5C	0.07	-0.77	0.78	0.09	-0.88	0.81

I5D	0.11	0.68	0.25	0.12	-0.29	0.61
I5E	0.07	-0.58	0.72	0.07	-1.59	0.94
I6A	0.13	0.20	0.42	0.06	-1.87	0.97
I6B	0.08	-0.52	0.70	0.11	-0.92	0.82
I6C	0.07	-0.56	0.71	0.11	-0.80	0.79
I6D	0.12	0.20	0.42	0.17	0.79	0.21
I6E	0.12	0.32	0.37	0.09	-1.10	0.86
I7A	0.09	-0.23	0.59	0.05	-1.85	0.97
I7B	0.12	0.86	0.19	0.19	1.10	0.14
I7C	0.09	-0.38	0.65	0.14	-0.30	0.62
I7D	0.10	-0.16	0.56	0.13	-0.21	0.58
I7E	0.10	-0.07	0.53	0.17	0.40	0.34
I8A	0.12	0.17	0.43	0.08	-1.46	0.93
I8B	0.14	0.57	0.28	0.16	0.08	0.47
I8C	0.15	0.84	0.20	0.16	0.25	0.40
I8D	0.08	-0.39	0.65	0.20	1.40	0.08
I8E	0.09	-0.33	0.63	0.29	2.06	0.02
I9A	0.13	0.49	0.31	0.23	1.84	0.03
I9B	0.12	0.48	0.32	0.13	-0.43	0.67
I9C	0.12	0.19	0.42	0.11	-0.86	0.81
I9D	0.12	0.62	0.27	0.17	0.47	0.32

Figure 2 shows that the two classes had different item difficulty parameters. Item difficulty estimates were different for the majority of items.

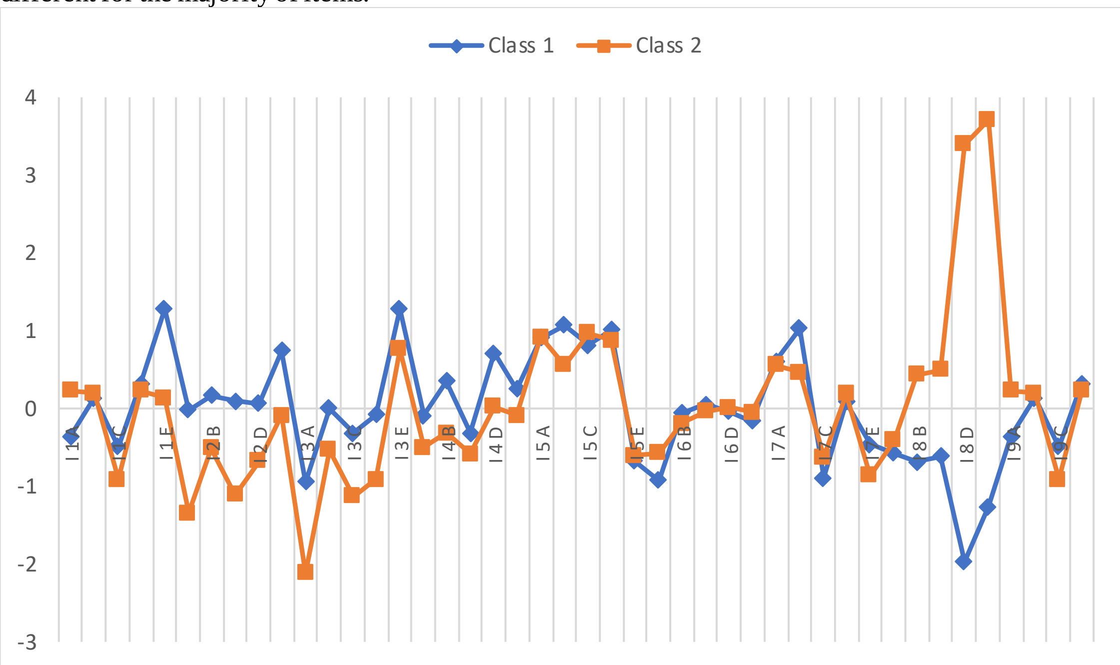


Figure 2. Class specific item parameter profiles

It can be concluded that all classes found the items to be relatively easy as logit position was generally negative (see Table 4 for specific values including standard error).

Table 4

Item parameters of Booklet One by classes

	Class-1		Class-2	
Item	Estimate	Error	Estimate	Error
I1A	-0.36	0.16	0.23	0.22
I1B	0.14	0.15	0.20	0.22
I1C	-0.49	0.17	-0.91	0.25
I1D	0.32	0.15	0.23	0.22
I1E	1.28	0.14	0.14	0.22
I2A	-0.02	0.15	-1.34	0.27
I2B	0.17	0.15	-0.51	0.23

I2C	0.09	0.15	-1.11	0.26
I2D	0.06	0.15	-0.68	0.24
I2E	0.74	0.14	-0.09	0.22
I3A	-0.94	0.19	-2.12	0.33
I3B	0.00	0.15	-0.53	0.24
I3C	-0.32	0.16	-1.13	0.26
I3D	-0.08	0.16	-0.91	0.25
I3E	1.28	0.14	0.76	0.21
I4A	-0.09	0.16	-0.50	0.23
I4B	0.35	0.15	-0.32	0.23
I4C	-0.32	0.16	-0.58	0.24
I4D	0.71	0.14	0.02	0.22
I4E	0.25	0.15	-0.09	0.22
I5A	0.91	0.14	0.92	0.21
I5B	1.07	0.14	0.56	0.21
I5C	0.82	0.14	0.97	0.21
I5D	1.01	0.14	0.88	0.21
I5E	-0.68	0.18	-0.60	0.24
I6A	-0.92	0.19	-0.57	0.24
I6B	-0.06	0.16	-0.19	0.22
I6C	0.05	0.15	-0.04	0.22
I6D	-0.04	0.16	0.01	0.22
I6E	-0.16	0.16	-0.05	0.22
I7A	0.61	0.14	0.56	0.21
I7B	1.03	0.14	0.47	0.21
I7C	-0.89	0.19	-0.63	0.24
I7D	0.08	0.15	0.19	0.22
I7E	-0.46	0.17	-0.86	0.25
I8A	-0.57	0.17	-0.41	0.23
I8B	-0.69	0.18	0.43	0.21
I8C	-0.61	0.17	0.50	0.21
I8D	-1.97	0.26	3.40	0.28
I8E	-1.26	0.21	3.71	0.30
I9A	-0.36	0.16	0.23	0.22
I9B	0.14	0.15	0.20	0.22
I9C	-0.49	0.17	-0.91	0.25
I9D	0.32	0.15	0.23	0.22

For example, item I1E had an item difficulty parameter of 1.87 and a standard error of 0.07 for class one and an item difficulty parameter of 0.13 and standard error of 0.08 for class two.

Conclusion and Discussion

The survey was created from an extensive literature review and content expert reviews of documents pertaining to teacher standards that guide teacher preparation programs. This yielded eight themes which we named “core competencies” that are essential for effective teaching. The survey was created based on these 8 CC’s with 4-6 questions for each CC. The purpose of this study was to the questionnaire validation to see if collected data yielded different sub-groups within the selected groups. Founded response patterns shows that construct validity might be called into question. This means that the same construct is not being measured same for all participants. Differences in item logit positions per class suggest further investigation into survey validity.

Based on these analyses, teacher preparation programs and even professional development personnel need to evaluate their current programming to consider what aspects are Skills related and what is Resource Use related. This in no way means that we ignore the eight core competencies, but this adds a new way to understand the focus of teacher development programs.

Recommendations

Asking what skills a teacher needs and what aspects of their program teaches students how to use resources creatively and effectively could improve the program and may lead to more effective teachers. Further research can focus on testing the validity of the survey using different techniques. Since the original development data analyzed here (collected for the Institute of Education Sciences grant), the survey instrument has been modified and used over a period of time as a formative tool in several types of preparation programs. Candidates use the survey to guide their self-assessment, and their clinical experience coaches have used the survey to guide formative evaluation. Programs for this continued research and development have included two field-based alternative route programs; two "traditional-design" university programs; and two partnership professional development school model programs, in which experienced teachers in the school where candidates are placed for clinical experiences serve as co-instructors.

Acknowledgements or Notes

This study was a part of project. A grant supported by the Institute for Education Sciences supported development of assessments of preparation of teachers. This paper is part of a research effort with many contributing participants; the authors have organized this particular work, but substantive credit should be given the entire team. Paper was presented as a poster at the annual meeting of the American Educational Research Association, San Antonio, Texas. Also, similar studies regarding questioning the validity of the survey has been presented in different conferences.

Author (s) Contribution Rate

Authors contributed equally to the study.

Conflicts of Interest

There is no conflict of interest for individuals or institutions in this research.

Ethical Approval

Ethical approval for this study was obtained from Usak University Ethics Committee on 19.01.2023.

References

Baghaei, P., & Carstensen, C. H. (2013). Fitting the mixed Rasch model to a reading comprehension test: Identifying reader types. *Practical Assessment, Research & Evaluation*, *18*(5), 1-13. https://doi.org/10.7275/n191-pt86

Bond,T. G., & Fox, C.M. (2007). *Applying the Rasch model* (2nd Ed). Erlbaum Associates.

Briggs, D. C., Circi, R., Seidel, K, & Green, K. (2013). Challenges in Measuring Core Competencies in Teacher Preparation Programs. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA, May 1, 2013

Cressie, N., & Read, T. R. C. (1984). Multinomial Goodness-Of-Fit Tests. *Journal of the Royal Statistical Society: Series B (Methodological)*, *46*(3), 440–464. https://doi.org/10.1111/j.2517-6161.1984.tb01318.x

Dallas, A. D., & Willse, J. T. (2014). Survey analysis with mixture Rasch models. *Journal of Applied Measurement*, *15*(4), 394-404.

Darling-Hammond, L., & Bransford, J. *Preparing teachers for a changing world: What teachers should learn and be able to do*. Jossey-Bass.

Ferguson, R. F. (1991). Paying for public education: New evidence on how and why money matters. *Harvard Journal on Legislation*, *28*(2), 465-498. https://heinonline.org/HOL/LandingPage?handle=hein.journals/hjl28&div=24&id=&page=

Frick, H., Strobl, C., & Zeileis, A. (2014). Rasch Mixture Models for DIF Detection. Educational and Psychological Measurement, 75(2), 208–234. https://doi.org/10.1177/0013164414536183

Frick, H., Strobl, C., & Zeileis, A. (2015). Rasch mixture models for DIF detection: A comparison of old and new score specifications. *Educational and Psychological Measurement*, *75*(2), 208-234. https://doi.org/10.1177/0013164414536183

Gauggel, S., Böcker, M., Heinemann, A. W., Lämmler, G., Borchelt, M.,& Steinhagen- Thiessen, E. (2004). Patient–staff agreement on Barthel Index scores at admission and discharge in a sample of elderly stroke patients. *Rehabilitation Psychology*, *49* (1), 21–27. https://doi.org/10.1037/0090-5550.49.1.21

Goe, L. (2007). *The link between teacher quality and student outcomes: A Research synthesis*. National Comprehensive Center for Teacher Quality.

Hanushek, E. A., & Rivkin, S. G. (2010, May). *Using Value Added Measures of Teacher Quality. Brief 9*. Urban Institute, National Center for Analysis of Longitudinal Data in Education Research (CALDER).

Hartnett-Edwards, K., Seidel, K., Spurlin, M., Anderson, S., Green, K., & Briggs, D. (2013). An exploration of novice teacher core competencies: Relationship to student achievement and depth of preparation. Presented at AACTE Annual Meeting, Orlando, Florida, February 2013.

Kaiser, F.G., Keller, C. (2001). Disclosing situational constraints to ecological behavior: A confirmatory application of the mixed Rasch model. *European Journal of Psychological Assessment, 17*, 212-221. https://doi.org/10.1027/1015-5759.17.3.212

Linacre, J. M. (2010). *A user's guide to winsteps ministep 3.70.0: Rasch-model computer programs*. Chicago: Winsteps.

Merbitz, C., Morris, J., & Grip, J. C. (1989). Ordinal scales and foundations of misinference. Archives of *Physical Medicine and Rehabilitation*, *70*(4), 308. https://doi.org/10.5555/uri:pii:0003999389901512

Mislevy, R. J., & Verhelst, N. (1990). Modeling item responses when different subjects employ different solution strategies. *Psychometrika*, *55*, 195–215. https://doi.org/10.1007/BF02295283

Pishghadam, R., Baghaei, P., & Seyednozadi, Z. (2017). Introducing emotioncy as a potential source of test bias: A mixed Rasch modeling study. *International Journal of Testing*, *17*(2), 127-140. https://doi.org/10.1080/15305058.2016.1183208

Rost, J. (1990). Rasch models in latent classes: An integration of two approaches to item analysis. *Applied Psychological Measurement*, *14*(3), 271-282. https://doi.org/10.1177/014662169001400305

Sanders, W.L., Horn, S.P. (1998). Research Findings from the Tennessee Value-Added Assessment System (TVAAS) Database: Implications for Educational Evaluation and Research. *Journal of Personnel Evaluation in Education* **12,** 247–256 https://doi.org/10.1023/A:1008067210518

Sanders, W.L., Rivers, J.C. (1996). *Cumulative and residual effects of teachers on future student academic achievement. Research Progress Report.* University of Tennessee Value-Added Research and Assessment Center.

Sanders, W.L., Wright, S.P. & Horn, S.P. (1997). Teacher and Classroom Context Effects on Student Achievement: Implications for Teacher Evaluation. *Journal of Personnel Evaluation in Education* **11,** 57–67 https://doi.org/10.1023/A:1007999204543

Şen, S. (2016). Applying the Mixed Rasch Model to the Runco Ideational Behavior Scale. *Creativity Research Journal*, *28*(4), 426-434. https://doi.org/10.1080/10400419.2016.1229985

Şen, S. & Toker, T. (2021). An Application of Multilevel Mixture Item Response Theory Model . Journal of Measurement and Evaluation in Education and Psychology , 12 (3) , 226-238 . https://doi.org/10.21031/epod.893149

Toker, T. & Green, K. (2021). A Comparison of Latent Class Analysis and the Mixture Rasch Model Using 8th Grade Mathematics Data in the Fourth International Mathematics and Science Study (TIMSS-2011) . *International Journal of Assessment Tools in Education* , *8* (4) , 959-974 . DOI: 10.21449/ijate.1024251

Smith, E. V., Jr. (2001). Evidence for the reliability of measures and validity of measure interpretation: A Rasch measurement perspective. *Journal of Applied Measurement, 2* (3), 281–311.

von Davier, M. (2001). *WINMIRA* [Computer software]. Institut für die Pädagogik der Naturwissenschaften

von Davier, M. (2001b). *WINMIRA user manual* [Computer software manual]. Institut für die Pädagogik der Naturwissenschaften

Wenglinsky, H. (2002). The link between teacher classroom practices and student academic performance. *Education Policy Analysis Archives, 10,* 1-30. https://doi.org/10.14507/epaa.v10n12.2002

Wright, B. D., & Stone, M. H. (2004). *Best test design*. MESA Press.