TELL, Vol. 6, No. 2, Fall & Winter 2012, pp. 25-47

Using Rasch to validate a measure of English language teacher prejudice

Reza Pishghadam[1]
Associate Professor, Ferdowsi University of Mashhad
Purya Baghaei
Assistant Professor, Islamic Azad University, Mashhad Branch
Elahe Bazri
MA, Ferdowsi University of Mashhad
Shoorangiz Ghaviandam
MA, Ferdowsi University of Mashhad

Received on August 1, 2012
Accepted on December 25, 2012

Abstract

The major objective of this study was to construct and validate an English Language Teacher Prejudice Scale (ELTPS). To this end, 180 English language teachers in private language institutes in Mashhad, Iran were asked to participate in this study. Rasch model was employed to substantiate the construct validity of the scale. The results of analyses exhibited that the whole scale was uni-dimensional and only one item misfitted. Rating scale statistics showed that the middle category of 'No idea' was redundant and should be removed from the scale. Some suggestions are made to fully validate the scale and use it in language education.

[1] Corresponding author at Ferdowsi University of Mashhad
Email: pishghadam@um.ac.ir

Keywords: teacher prejudice, English language, validation, Rasch analysis

1. Introduction

Being a multi-faceted profession, teaching deals with different personal, social, educational, and cognitive aspects. During the last decades, the cognitive aspect has been the center of attention in teaching education (Moini, 2009). Mentality of teachers serves a central role in developing students` learning process. In fact, "the beliefs, intentions, and personalities of all teachers play a more significant role in the success or failure of individual students than the curriculum, materials, class size, and so forth" (Obidah & Teel, 2001, p. 107; cited in Pang, 2002). A spate of research has been done in order to offer a better understanding of these constructs in English language teaching (e.g. Moini, 2009; Pishghadam, Askarzadeh Torghabeh, & Navari, 2009).

One of the major constructs of teacher cognition which can affect the teaching process is teacher prejudice. Prejudice is considered to be "a negative judgment or opinion formed about a group without knowledge of the facts" (Kreidler, 1997, p. 265). Like every person, language teachers form prejudices and biases in their mind, bringing them into classroom consciously or unconsciously. Consequently, their biases affect the success of their students. As Pang (2002) suggests teacher prejudice is one of the most challenging issues to tackle in teacher education.

To our knowledge, this construct has been widely ignored in language education. Therefore, due to the paucity of research in this area, this study was built on understanding the nature of prejudice among English language teachers. To this end, we attempted to construct a valid measure to scrutinize the kinds of prejudices and biases surrounding English language teachers.

2. Literature Review

As already mentioned, teacher prejudice is part of teacher cognition. Teacher cognition deals with "what teachers think, know, and believe" (p. 1), and consequently its main focus is the mental lives

of teachers (Borg, 2001). The practice of language teachers can be considered as a complex interplay between their cognition, methodology, and the context in which they teach (Borg, 2006). Teacher cognition has considerable influence on the way they teach which has important interaction with their practice.

One important part of teachers' cognition is beliefs. Beliefs play a considerable role in many aspects of teaching. Teachers' belief is referred to as "teachers' pedagogic beliefs or those beliefs of relevance to an individual's teaching" (Borg, 2001, p. 187). Many studies suggest that teachers' beliefs and values are essential in their practice (e.g. Evrim, Göçek, & Enisa, 2009; Fang, 1996; Hall, 2005; Stipek, Givvin, Salmon, & MacGyvers, 2001). The belief system of teachers form the basis of activities they do during teaching (Moini, 2009). Apart from how knowledgeable teachers are, the governing force of their action in the class in large part is their belief rather than the amount of knowledge they have (Hall, 2005). Teachers' belief also may influence their attitude towards the teaching and learning process.

Kagan (1992) discussed that not only a teacher's personal knowledge, but also their professional knowledge might be considered as beliefs; In fact, the majority of English teachers have particular predetermined beliefs in order to approach English teaching in the best way. As demonstrated by several researchers (Johnson, 1992; Richards & Lockhart, 1996; Smith, 1996), teachers are often affected by their previous educational experiences, cultural backgrounds, and social interaction, which might substantially influence their English language teaching (Liao, 2007). He maintained that, since every English language teacher has their unique sets of beliefs towards language teaching, these beliefs are mostly considered to play an important role in their teaching practices. In the same vein, Kagan (1992) referred to beliefs as personal knowledge, stating that beliefs are the major components of teachers' professional knowledge. Flores (2001) also demonstrated that prior experiences have a considerable impact on teachers' beliefs.

Another important factor that is highly influential in the description of teachers' cognition is their biases and prejudices. Prejudices can be "part of our thinking process" (Woolfolk, Winne, & Perry, 2003, p. 170). Valencia (1997) and Solorzano (1997)

explain that "teachers' deficit thinking causes them to make biased judgments of students' intelligence, ability, and behavior" (cited in Cooper, 2003, p. 103). Teachers' biased ideas about learning and teaching can have serious impact on students. Most studies done about teacher's prejudice relate this notion to a sociological basis mostly about the effect of teachers' biased ideas against ethnic and racial minorities (e.g., Adler, 2011; Cooper, 2003; Thijs, Westhof, & Koomen, in press; van den Bergh, Denessen, Hornstra, Voeten, & Holland, 2010). All of these studies have focused on teacher prejudice in general education.

3. Purpose of the Study

Granted the fact that teacher prejudice is of vital importance in learning and teaching processes, delving more closely into this construct is required. To our knowledge, in the realm of English language teaching, there seems to be no research done to take teacher prejudice into account. In this regard, one of the major elements which can pave the way for other researchers to channel their studies into this area is designing a valid scale of teacher prejudice. Therefore, in this study we intend to construct a valid questionnaire for examining this construct.

4. Methodology

4.1 Participants

To collect the required data, the questionnaire was distributed among English language teachers, teaching English in different private language institutes in Mashhad, a city in Iran. The total number of participants was 180 EFL (English as a Foreign Language) teachers. They were both male (30 %) and female (70%), whose age ranged between 20-58 years. The number of respondents holding bachelor's degree and master's degree were equal (B.A.: 88; M.A.: 88), and only two held a PhD degree. With regard to their major, 85 had a degree in English language teaching, 33 majored in English literature, 37 were educated in the field of translation, and 18 majored in other fields of study. In Iran,

everybody who is highly proficient in English is allowed to teach it in private language institutes. As to their teaching experiences, all participants had at least a minimum of 1 and maximum of 40 years of teaching experience. The average of teaching experience was 8 years.

4.2 Scale Development

The present study involved the designing and administration of a questionnaire for English language teacher prejudice. A checklist of factors of teachers' prejudices was provided based on the guidelines laid down by the experts in the field (e.g. Fang, 1996; Pishghadam & Saboori, 2011). The checklist includes factors such as teachers' biases towards learners' accent, western culture, learners' fluency, participation, and mistakes. Consultations were held by the experts in the field of foreign language learning and teaching to ensure about the plausibility of the factors. Then, based on the checklist, for each factor about 4 or 5 items were developed as a Likert-type scale. The scale comprised a set of 32 items. Following this stage, to remove the ambiguities in the items, two experienced teachers were asked to read and think aloud their opinions concerning each individual item. The items were revised based on their comments, and then 6 English teachers similar to the target group were asked to answer the questionnaire to make sure the items were not ambiguous or unclear. By applying the comments gained in the piloting stage, items were revised and 3 items were dropped. Ultimately, 29 items remained for further analyses (see Appendix). The revised version of the scale was administered to EFL teachers in different private language institutes.

4.3 The Measurement Model

The Rasch rating scale mode (RSM; Andrich, 1978) was employed as the appropriate measurement model to analyze the ploytomous data obtained from Likert-type response categories in this study. In contrast to classical test theory, the Rasch model relies on some probabilistic assumptions to scale persons and items and evaluate the psychometric properties of measurement instruments. The Rasch model transforms ordinal person and item raw scores to

interval measures which can be located on the same metric. Examining the distributions of person and item locations on the logit scale indicates the extent to which the test is relevant for the sample. If the data fit the Rasch measurement model, the item and person parameter estimates do not depend on the specific sample of persons or items used for scaling. The other appealing property of the Rasch model is that, unlike 2-PL and 3-PL, no assumptions as regards the normality of the distribution of person traits need to be made. Primarily for these reasons the Rasch model was used in the present study to evaluate the psychometric properties of the English teacher prejudice scale. The Rasch rating scale model is formally expressed as follows:

$$P(X_{ni}=x)=\frac{exp\sum_{k=0}^{x}[\theta_n-(\delta_{i+\tau_k})]}{\sum_{x=0}^{m}exp\sum_{k=0}^{x}[\beta_n-(\delta_{i+\tau_k})]}$$

$$x=0,1,\ldots,m$$

where $P(X_{ni}=x)$ is the probability that person n will be observed in rating scale category x on item i, which has m+1 rating scale categories, $\theta_n$is the persons location on the trait continuum, $\delta_i$is the item's difficulty (endorsability) and $\tau_k$is the threshold parameter.

The parameters of the model were estimated by WINSTEPS Rasch program (Linacre, 2009) in which joint maximum likelihood estimation is implemented. The respondents' estimates indicate their levels of prejudice. The item calibrations refer to the difficulty of endorsing the items as descriptions of the respondents' prejudiced behavior in the classroom.

5. Results

Participants endorsed their levels of agreeability/disagreeability on a 5-point Likert scale. Some of the items had to be reverse-scored so that higher scores show higher levels of prejudice. Initial analysis of the data with the rating scale model showed that only Item 7

misfits the Rasch model. This item reads "I ignore students who speak Persian in class discussions". The analysis yielded a reliability of .74, a person separation of 1.70 and an item separation of 6.47. Item measures ranged from -.90 (If more proficient students keep silent in class discussions I do not insist them to speak.) to 1.25 logits (I prefer both fluent and non-fluent students to take part in class discussions equally). The *root mean square error* (RMSE) for items was 0.08 and for persons is 0.20 which indicate quite precise measurement.

5.1 Rating Scale Diagnostics

Table 1 shows category statistics for each response option. 'Count' indicates how many respondents chose a particular category, summed across all items. "Irregularity in observation frequency across categories may signal aberrant category usage. A uniform distribution of observations across categories is optimal for step calibration. Other substantively meaningful distributions include uni-modal distributions peaking in central or extreme categories, and bimodal distributions peaking in extreme categories" (Linacre, 1999, p. 110). The table shows that the distribution of observations for categories is bimodal with peaks at Categories 2 and 4, which is an instance of irregular observation distribution.

'Average Measure' is the mean of the trait estimates (in logits) for all persons who chose the corresponding category. For example, the average of the trait estimates of those who chose category 1 on any item in the scale is -.86 (Bond & Fox, 2007). These values should monotonically increase to indicate that those with higher trait estimates choose the higher categories and vice versa (Bond & Fox, 2007). As Table 1 shows average measures are monotonically increasing for the categories in our data.

Table 1: Rating scale statistics

Category	Count	Average Measure	Infit Mean square	Threshold
1	729	-.86	.99	None

2	2096	-.43	.97	-1.69
3	612	-.12	1.00	.84
4	1370	.10	1.01	-.82
5	358	.31	1.04	1.57

For Likert scales, infit mean squares greater than 1.40 indicate that the category was used in an unexpected way and there is unexplained randomness in the observations. Mean squares less than .60 indicate over-predictability in the data (Bond & Fox, 2007). Table 1 shows that infit mean squares are close to their perfect value, i.e., 1 in these data.

Rating scales imply that as the level of the latent trait increases in respondents a progression should be observed in the categories of the rating scale. Each category of the rating scale is expected to be most probable (to be chosen) for a certain group of respondents; persons higher on the trait continuum are expected to choose higher categories and vice versa.

Thresholds are estimated difficulties of observing one response category over the category below. They show how difficult it is to observe each category. They are the points on the rating scale where the probability of being observed in a category and the category below is equal. For example, it is equally probable for a person with a trait estimate of -1.69, on average, to endorse categories1 and 2 on all items. If the trait estimate is lower than -1.69 the probability of observing category 1 increases and if it is greater than -1.69 the probability of observing category 2 increases. The first category has no lower category so there is no measure for that. We expect threshold estimates to increase with category values.

Table 1 displays that the thresholds for categories 3 and 4 are disordered. This has happened because of the irregularity in the distribution of observations (Linacre, 1999). Disordering in threshold estimates, i.e., thresholds which do not advance with category values indicate that the category is rarely endorsed

and has a narrow interval on the variable or the "concept is poorly defined in the minds of respondents" (Linacre, 1999, p. 114).

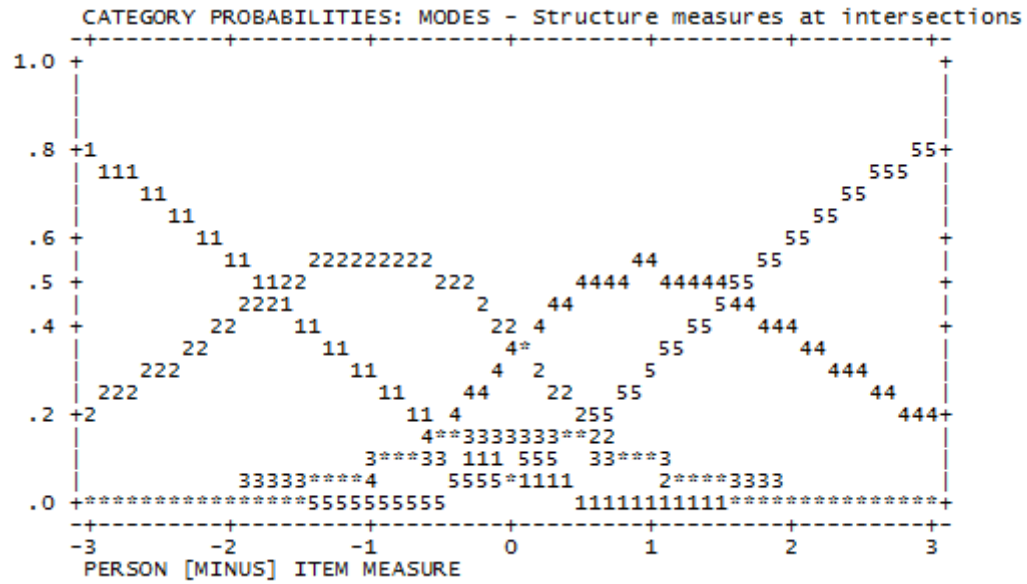

Figure 1: Probability curves for the 5-point category scale

Figure 1 shows that the probability curve for category 3 is flat. Each category should have a peak to show that each is the most probable response for some regions of the scale. Category 3 is redundant and does not define a specific section of the trait continuum. Thresholds are in fact intersections of rating scale categories.

Table 2: Rating scale statistics after collapsing categories

Category	Count	Average Measure	Infit Mean square	Threshold
1	729	-.78	.96	None
2	2096	-.39	1.00	-1.89

4	1982	.07	1.02	-.15
5	358	.26	1.04	2.04

In order to remedy the disordered threshold problem the variable was re-categorized. Category 3 was collapsed down, i.e., categories 3 and 4 were both scored three (12334) and the data were reanalyzed. Table 2 shows that category statistics improved after collapsing categories 3 and 4. Other scale statistics including reliability (.74), person separation (1.68), item RMSE (.11) and person RSMSE (.30) deteriorated after collapsing Categories 3 and 4. The lower precision indices are due to smaller number of categories. This sounds like loss of information. However, disordered categories are counterproductive to measurement and the information they produce has not much value. Reliability drops because the number of score points is reduced. Higher reliability due to score points that do not conform to the measurement model is artificially inflated reliability.

The distance between the first thresholds and second thresholds is 1.74 logits and between second and third is 1.89logits. Linacre (1999) states that threshold distances which define distinct proportions of the variable should be greater than 1.4 logits and less than 5 logits. The distances among the thresholds after re-categorization meets the criteria set by Linacre. Alternatively, it is possible to collapse another category and devise a 3-point scale with categories which cover longer portions on the variable. Nevertheless, this will result in the loss of information and lower reliability.

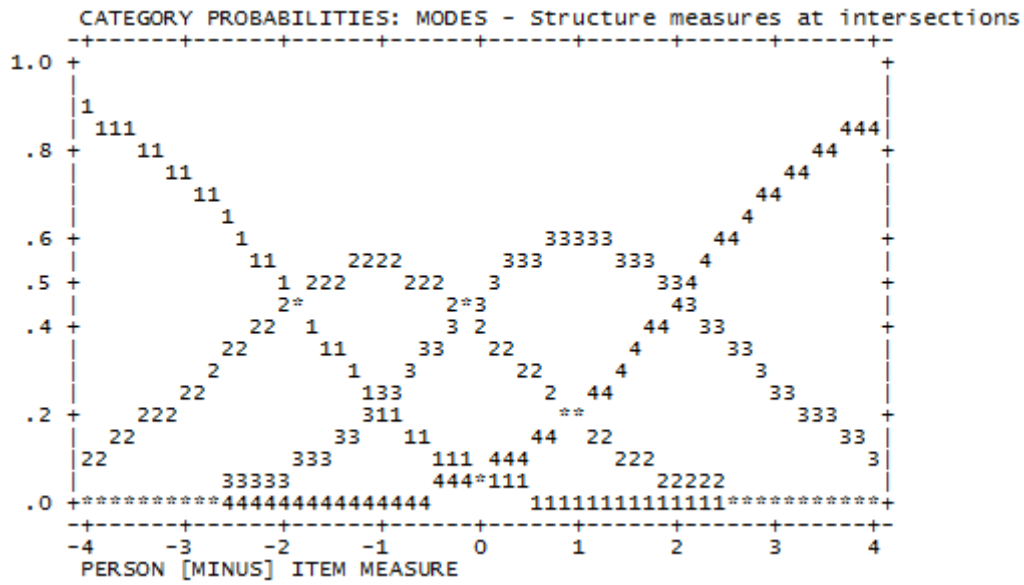

Figure 2: Probability curves for the 4-point category scale (12334 collapsing)

Re-categorization of the variable with three response categories was also examined. Configurations of '33221' and'32221' were examined. Although the distances among the thresholds increased, the reliability in '32221' dropped to .60. However, the reliability of '32221' was .73 which is slightly lower than the reliability of the 4-point scale.

Figure 2 shows the category curves for the scale after collapsing Category 4 downward into Category 3. The curves show that each category represents a unique section of the measured construct, i.e., each category is most probable for respondents falling on certain sections of the trait continuum. Moreover, none of the infit mean square indices was greater than 2 (Linacre, 1999) indicating that no noise crept into the measurement process.

5.2 Item Statistics

Item estimates and their fit values after collapsing categories are shown in Table 3. The table shows that Item 7 still misfits with an

infit mean square value of 1.36. This item should be removed from the scale. Other items have acceptable mean square infit values (.70-1.30) (Bond & Fox, 2007).

Table 3: Item estimates and fit statistics

Item	Estimate	Error	Infit MNSQ
1	-.59	.07	.99
2	-.52	.07	1.14
3	1.14	.10	1.13
4	-.31	.07	.91
5	.28	.07	1.04
6	-.05	.07	.75
7	.00	.07	1.36
8	.40	.07	1.26
9	.12	.07	.98
10	.58	.08	.99
11	.02	.07	1.04
12	.16	.07	.95
13	.43	.07	1.07
14	-.31	.07	1.05
15	.60	.08	1.12
16	-.83	.08	1.28
17	1.00	.09	1.13
18	.05	.07	.94
19	-.03	.07	.99

20	-.07	.07	1.03
21	-.17	.07	.95
22	-.45	.07	1.14
23	-.13	.07	.80
24	-.15	.07	1.03
25	.28	.07	.86
26	-.73	.08	.84
27	-.15	.07	.87
28	-.36	.07	1.01
29	.25	.07	.95

Figure 3 is a Wright map of the item and person distributions on the logit scale. The map shows that the instrument is well targeted, with the item calibrations spanning much the same range as the person measures. The map shows that the item locations, on the right, cover a difficulty span of 1.97 logits (-.83 to 1.14); the person measures, on the left, cover a trait span of 2.81 logits (-1.83 to .98). The test is composed of Likert-type items which have a threshold span of 3.51 logits (-1.62 to 1.89). This indicates that the operational range of the items is much wider than what the map shows.

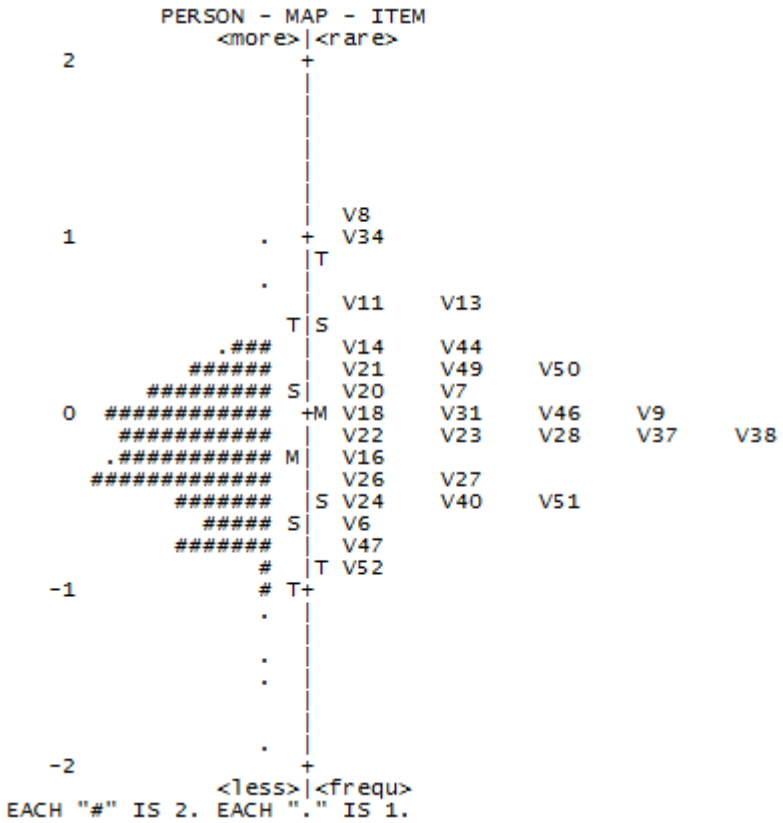

Figure 3: Map of trait distributions and item parameter estimates

Figure 4 shows category thresholds for the items in the instrument. The items are ordered vertically in descending order of difficulty on the right. The horizontal lines on the bottom and top indicate the Rasch logit scale. The numbers in the rectangle show the location of each threshold indicated by its corresponding category score. The first category has no threshold, so there are no 1's in the rectangle. Category 3 was combined with category 4, so there are no 3's either. The location of threshold categories for each item can be read off from the logit scales on the top and bottom of the box.

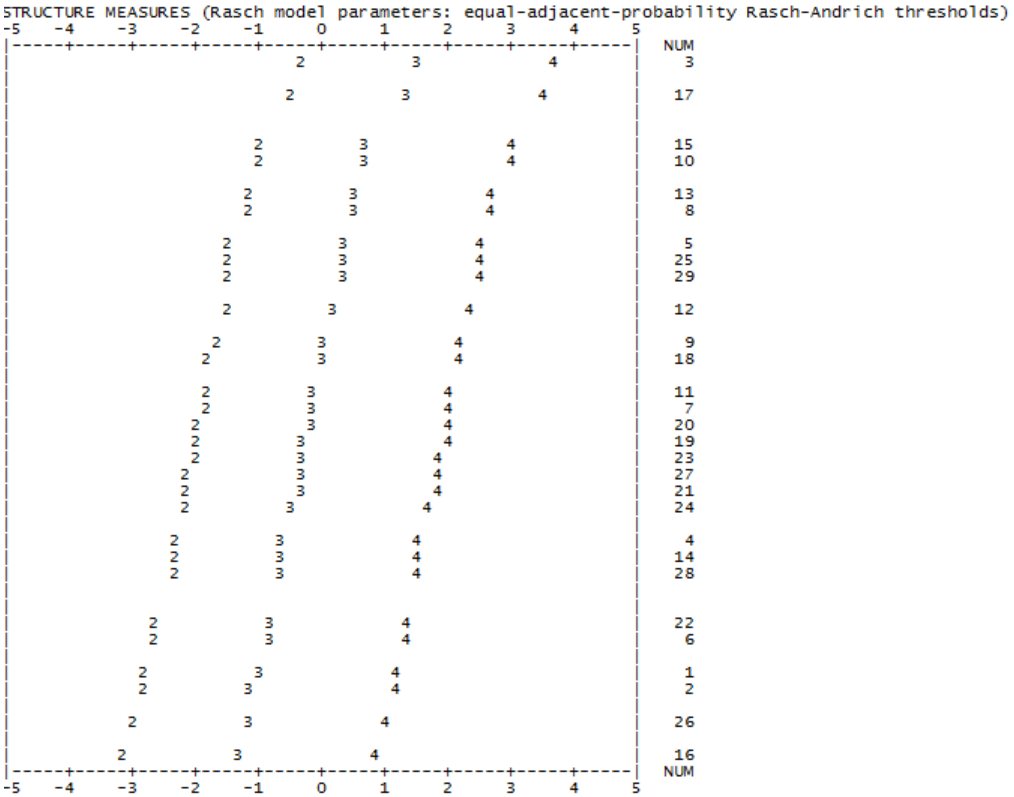

Figure 4: Map of category thresholds

Unlike partial credit model (PCM) (Masters, 1982) where a set of thresholds is estimated for every single item, in rating scale model (RSM) one set of thresholds is estimated for all the items in the instrument. In RSM the distances among the points on the scale are assumed to be equal across all the items. That is why one set of thresholds suffices for all the items, although the distances within the items can vary. PCM is less restrictive, i.e., the distances among the points on the scale need not to be equal, neither across nor within the items. That is, every item is a unique scale and has a different set of threshold estimates (Baghaei, 2010).

Although in RSM there is only one set of thresholds shared among all the items it is possible to have unique thresholds for each item relative to item difficulty. The item threshold estimates under

RSM are computed by adding the estimated thresholds to the item parameters. If there are three estimated RSM thresholds indicated by τ_1, τ_2, and $\tau_3$the thresholds for Item *i* with difficulty δ_i will be $\delta_i+\tau_1$, $\delta_i+\tau_2$, and $\delta_i+\tau_3$, respectively.

As Figure 4 shows category 2 of Item 16, which is the easiest item, defines the lowest threshold in the instrument. Category 4 of Item 3, which is the hardest item, defines the highest threshold in the instrument. Other item thresholds fall in between. The map shows that item categories define an operational range wider than the overall item estimates shown in Figure 1. The lowest threshold is -3.11logits (Item 16) and the highest is 3.64logits (Item 3).

6. Discussion

As already mentioned, the major aim of this study was to design and validate a scale on English language teacher prejudices. Based on the literature on teacher prejudice, a questionnaire with 29 items was constructed and Rasch model was utilized to analyze the data.

Rasch model analyses showed that only one item misfitted and therefore should be removed from the questionnaire. Rating scale category statistics indicated that the thresholds for Categories 3 and 4 were disordered. After collapsing up these two categories the thresholds were ordered. This finding further corroborates the previous research that the middle category of “No Idea”, “Undecided”, “Neutral” in rating scales leads to category malfunctioning and should be avoided in rating scale construction (Garland, 1991; Nunnally, 1967).

Figure 3 showed that item locations, on the right, cover a difficulty span of 1.97 logits (-.83 to 1.14); persons, on the left, cover a trait span of 2.81 logits (-1.83 to .98). The spread of the trait estimates is .84 logits wider than the spread of item locations. But considering that the scale is composed of Likert-type items which have a wide threshold span indicates that the operational range of the items is much wider than what Figure 3 shows. Figure 4, map of thresholds, showed that item categories define a wide range of more than 5 logits which cover a broad span of trait estimate.

The overall analysis of the results demonstrated that the questionnaire is uni-dimensional and valid for measuring English language teacher prejudice. As already indicted, the new version of the questionnaire should be used without Item 7 and with a 4-point Likert scale, i.e., without the middle category of 'No Idea'. Seemingly, this category adds nothing to the required information, which makes it a redundant category.

As already stated, this newly-designed instrument measures English language teacher prejudice against western culture, accent, fluency, mistakes, and so on. These biases can hinder the process of learning, having a detrimental effect on the minds of the learners. For instance, one of the language teachers' prejudices is their biases in favor of the western culture (Pishghadam & Sadeghi, 2011). Teachers with biases against their home culture try to behave as native speakers, devaluing their own culture. This may affect identity formation and cultural attachment of their students and ruin the potentials of English language class to "be the sites for developing the cultural and national identity of the learners" (Pishghadam, 2011, p. 14). As an example, Item 6 measures teachers' inclination to talk more about western culture rather than their own culture.

Another important aspect of teachers' prejudice is their biases towards learners' accent. Some English language teachers push learners to achieve a native-like accent (Pishghadam & Saboori, 2011). In the era of World Englishes in which notions such as "the death of native speaker" (Widdowson, 2003) is well accepted, all efforts for having a native-like accent is in vain. If teachers show positive tendency towards putting on American or British accent, this may create misleading conceptions in learners about the learning of English language. Moreover, teachers' biased ideas about fluent and non-fluent learners may have consequences for learners. Teachers may tend to give more opportunity to fluent learners for participation and in this way exclude non-fluent learners indirectly. Another factor which is important in defining teachers' prejudice is their biases towards learners' mistakes. Showing an inappropriate reaction to learners' mistakes can demotivate them or spoil their risk-taking behaviors.

7. Conclusion and Implications

The outcomes of this study demonstrated that the scale is uni-dimensional and valid for measuring teacher prejudice in the realm of English education in Iran. This instrument measures the potential biases English language teachers may display in language learning classes. Detecting diverse types of biases helps English language teachers get conscious of their own biases, trying to modify or remove them. Moreover, teacher educators are expected to be more familiar with these biases, sensitizing prospective teachers with the wrong beliefs which might hinder the learning process. All in all, digging and delving into the minds of teachers can illuminate the beliefs and values, which are hidden and unrevealed to teachers themselves. These beliefs and values act as a map for teaching, and of course if it is drawn wrongly, it may lead to mis-teaching and misguiding.

In the end, to more substantiate the validity of the scale, it is also recommended that the association of this newly-made questionnaire with belief scales be measured. Moreover, determining the underlying factors of the scale is of high importance. Exploratory and confirmatory factor analyses can be used to unravel the underlying sub-constructs of the instrument. Corroborating the association of the scale with teacher success can also be a fruitful endeavor. Finally, it is our hope that this instrument can shed more light on the construct of teacher prejudice and that can be widely employed in English language teaching settings to make teachers conscious of their biases, trying to remove them.

References

Adler, S., M. (2011). Teacher epistemology and collective narratives: Interrogating teaching and diversity. *Teaching and Teacher Education, 27*, 609-618.

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43*(4), 561-573.

Baghaei, P. (2010). A comparison of thee polychotomous Rasch models for super-item analysis. *Psychological Test and Assessment Modeling, 52*(3), 313-322.

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.

Borg, M. (2001). Teachers' beliefs. *ELT Journal, 55*(2), 186-188.

Borg, S. (2006). *Teacher cognition and language education: Research and practice*. London: Continuum.

Cooper, C. W. (2003). The detrimental impact of teacher bias: Lessons learned from the standpoint of African American mothers. *Teacher Education Quarterly, 13*, 101-116.

Evrim, E., Gökçe, K., & Enisa, M. (2009). Exploring the relationship between teacher beliefs and styles on classroom management in relation to actual teaching practices: A case study. *Procedia Social and Behavioral Sciences, 1*, 612–617.

Fang, Z. (1996). A review of research on teacher beliefs and practices. *Educational Research, 38*, 47-65.

Flores, B. B. (2001). Bilingual education teachers' beliefs and their relation to self-reported practices. *Bilingual Research Journal, 25*(3), 275-299.

Garland, R. (1991). The mid-point on a rating scale: Is it desirable? *Marketing Bulletin, 2*, 66-70. Available: http://marketing-bulletin.massey.ac.nz

Hall, L. (2005). Teachers and content area reading: Attitudes, beliefs and change. *Teaching & Teacher Education, 21*(4), 403-414.

Johnson, K. E. (1992). Learning to teach: Instructional actions and decisions of pre-service ESL teachers. *TESOL Quarterly, 26*(3), 507-535.

Kagan, D. M. (1992). Implication of research on teacher belief. *Educational Psychologist, 27*(1), 65-90.

Kreidler, W. J. (1997). *Conflict resolution in the middle school*. Cambridge: Educators for Social Responsibility.

Liao, P. (2007). Teachers' beliefs about teaching English to elementary school children. *English Teaching & Learning, 31*(1), 43-76.

Linacre, J. M. (1999). Investigating rating scale category utility. *Journal of Outcome Measurement, 3*(2), 103-122.

Linacre, J. M. (2009). WINSTEPS [Computer Software]. Version 3.72.3, Chicago, IL: winsteps.com.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, *47*(2), 149-174.

Moini, M. R., (2009). The impact of EFL teachers' cognition on teaching foreign language grammar. *Pazhuhesh-e Zabanha-ye Khareji*, *49*, 141-164.

Nunnally, J. C. (1967). *Psychometric theory*. New York: McGraw-Hill.

Obidah, J. E., & Teel, K. M. (2001). *Because of the kids: Facing racial and cultural differences in schools.* New York: Teachers College Press.

Pang, V. O. (2002). The courage to change: Teacher prejudice [Review of the book *because of the kids: Facing racial and cultural differences in schools*]. *Journal of Teacher Education, 53*(5), 444-446.

Pishghadam, R. (2011). Introducing applied ELT as a new approach in second/foreign language studies. *Iranian EFL Journal*, *7*(2), 9-20.

Pishghadam, R., Askarzadeh Torghabeh, R., & Navari, S. (2009). Metaphor analysis of teachers' beliefs and conceptions of language teaching and learning in Iranian high schools and language institutes: A qualitative study. *Iranian EFL Journal*, *4*, 6-40.

Pishghadam, R., & Saboori, F. (2011). A Qualitative analysis of ELT in the language institutes of Iran in the light of the theory of world Englishes. *Journal of Language Teaching and Research, 2*, 569-579.

Pishghadam, R., & Sadeghi, M. (2011). Home culture attachment and access to social/cultural capital: A case of Iranian English foreign language teachers. *African Journal of Social Sciences, 1*(3), 56-65.

Richards, J. C., & Lockhart, C. (1996). *Reflective teaching in second language classrooms*. Cambridge: Cambridge University Press.

Smith, D. B. (1996). Teacher decision making in the adult ESL classroom. In D. Freeman & J. C. Richards (Eds.), *Teacher*

learning in language teaching (pp. 197-216). Cambridge: Cambridge University Press.

Solorzano, D. (1997). Images and words that wound: Critical race theory, racial stereotyping, and teacher education. *Teacher Education Quarterly, 24*(3), 5-19.

Stipek, D. J., Givvin, K. B., Salmon, J. M., & MacGyvers, V. L. (2001). Teachers' beliefs and practices related to mathematics Instruction. *Teaching and Teacher Education, 17*, 213-226.

Thijs, J., Westhof, S., & Koomen, H. (in press). Ethnic incongruence and the student-teacher relationship: The perspective of ethnic majority teachers. *Journal of School Psychology*.

Valencia, R. (1997). Conceptualizing the notion of deficit thinking. In R. Valencia (Ed.), *The evolution of deficit thinking*. London: Falmer Press.

van den Bergh, L., Denessen, E., Hornstra, L., Voeten, M., & Holland, R. W. (2010). The implicit prejudiced attitudes of teachers: Relations to teacher expectations and the ethnic achievement gap. *American Educational Research Journal*, *47*(2), 497-527.

Widdowson, H. G. (2003). *Defining issues in English language teaching.* Oxford: Oxford University Press.

Woolfolk, A. E., Winne, P. H., & Perry, N. A. (2003). *Educational psychology* (2nd ed.). Boston: Allyn & Bacon, Inc.

Appendix

English Language Teacher Prejudice Scale (ELTPS)

No.	Statement	SD	D	U	A	SA
1	My students should achieve excellence in everything I teach them.	1	2	3	4	5
2	I think English teachers and students must always speak English in class.	1	2	3	4	5
3	I prefer both fluent and non-fluent students to take part in class discussions equally.	5	4	3	2	1
4	I do not pay attention to their grammar if students speak comprehensibly.	5	4	3	2	1
5	I am intolerant of what non-fluent students talk about if they take too much time speaking.	1	2	3	4	5
6	I prefer to give more examples from western cultures while teaching English.	1	2	3	4	5
7**	I ignore students who speak Persian in class discussions.	1	2	3	4	5
8	I do not like to waste too much time on silent (weak) students.	1	2	3	4	5
9	If students speak comprehensibly, I ignore their mistakes.	5	4	3	2	1
10	I will be disappointed if my students make mistakes.	1	2	3	4	5
11	I prefer to explain some difficult points in Persian.	5	4	3	2	1
12	I prefer students with native-like accent participate more in discussions.	1	2	3	4	5
13	I think that young students have lower language proficiency.	1	2	3	4	5
14	I am not very patient with students' excuses for poor work.	1	2	3	4	5
15	I like my students to understand the differences between Iranian and western cultures.	5	4	3	2	1

16	If more proficient students keep silent in class discussions I do not insist them to speak.	5	4	3	2	1
17	I like to involve all learners in class discussions irrespective of their ages.	5	4	3	2	1
18	I believe that students should learn to speak with no mistake.	1	2	3	4	5
19	I sometimes give or ask the meaning of new words in Persian.	5	4	3	2	1
20	I prefer to talk more about western culture rather than Persian culture in class.	1	2	3	4	5
21	I cannot tolerate students' careless repetitive mistakes.	1	2	3	4	5
22	I think that pronunciation is not important if students can communicate.	5	4	3	2	1
23	I believe language teachers and learners must be representative of the target language culture.	1	2	3	4	5
24	Sometimes I give instructions in Persian.	5	4	3	2	1
25	I cannot tolerate students' mistakes on pronunciation even if it's comprehensible.	1	2	3	4	5
26	I like my students to be familiar with western culture as much as possible.	1	2	3	4	5
27	I'll be demotivated when students don't maintain the standards I assign.	1	2	3	4	5
28	I believe Persian culture must be highlighted in English classes.	5	4	3	2	1
29	I like my students to be able to communicate, their accent is not important for me.	5	4	3	2	1

* SD = Strongly Disagree; D = Disagree; U = Undecided; A = Agree; SA = Strongly Agree

** Misfitting item