ATLANTIS PRESS

Advances in Social Science, Education and Humanities Research, volume 295

1st International Conference on Educational Sciences and Teacher Profession (ICETeP 2018)

Developing and Validating an Instrument of In-service Teachers Responses to Knowledge-Based Teacher, Engagement, and Expectation in Teacher Profession Education Program in Indonesia:

Integrating factor analysis with Rasch modeling

Patuan Raja, Bambang Setiadi
English Education Study Program, Faculty of Teacher Training and Education
University of Lampung
Lampung, Indonesia

Abdurrahman Abdurrahman
Science Education Graduate Program, Faculty of Teacher Training and Education
University of Lampung
Lampung, Indonesia
Abdurrahman.1968@fkip.unila.ac.id

***Abstract*—This paper reports on the use of several quantitative analytical approaches, including Rasch analysis, to examine teacher responses to questionnaire items that explore opinions related to knowledge-based teacher, involvement, and expectations in the Teacher Professional Education Program in Indonesia. While many research reports have presented the impact of training-based teacher professional development programs in Indonesia, this paper seeks to improve these results by utilizing Rasch model analysis to identify items with scaled score sequences in all response categories, and make them new explorations and factor confirmation analyzes. The 3-factors scale produced (knowledge-based teacher, teacher involvement, teacher expectations) is proven to be acceptable in terms of confirmatory factor analysis and in terms of analysis of Rasch items. Furthermore, this paper briefly discusses the implications of these results with respect to the capacity of improved instruments to gather information about how teachers see and hope about the Indonesian Teacher Professional Program, including analysis of gender disparity.**

Keywords—knowledge-based teacher; teacher profession; rasch model

I. INTRODUCTION

Over the past decade, there has been a dramatic increase in research on teacher professional development (TPD) program. Even today it has become a global agreement that teacher effectiveness has become the top of the education policy agenda criteria, because many countries have become convinced that teaching is one of the most important factors related to school especially in student achievement [1,2]. In addition, teacher preparation and development are the key to developing effective teachers across the world [3]. For example, the State of Finland has strengthened a high quality education system that is equitable by establishing a quality teaching profession where all teachers hold at least a 2-year master's degree which includes mastery of strong subject matter content and mature pedagogical preparation, and is integrated in research and successful practice. The impact is that teaching has become the most sought after profession, and many teachers continue their studies to pursue PhD degrees and then remain teachers. In one generation, Finland jumped from relatively low-educated countries to the current literacy rate of 96%, with high graduation and college levels and top scores in all fields of PISA assessment [3-5]. Likewise with other great countries such as the United States, Britain, Australia, Canada, Singapore, and others, they already have a very massive teacher development professional development system and model [3,6-8].

Furthermore, almost all countries in the world include teacher certification programs as one of the best efforts in teacher professional development programs. Many studies have tried to explain exploratively whether certification programs affect student achievement [9-12]. However, the general findings from the literature have not been very good. Not all professional development experiences are effective to enhance student learning achievement [13]. TDP had not been realized as an effort to improve the quality of learning in schools, so that students' academic achievements did not significantly increased. In addition, research that attempts to see the impact of teacher certification in developing countries on student performance is very rarely found in literature. One of the research result about quality teacher in Indonesia revealed that the teacher certification program in Indonesia has not seriously impacted the improvement of student learning outcomes, so TPD models are needed that are more effective in equipping teachers to be competent in all their professional duties [12]. In the context of preparing professional teachers in Indonesia, currently the teacher professional education program (PPG) has been carried out in the form of college-based continuing teacher professional education to increase teacher knowledge and skills in supporting the students' achievement of learning outcomes in the form of complex thinking skills students need

to learn and work in the 21st century. Sophisticated forms of teaching are needed to develop student competencies such as deep mastery of challenging content, critical thinking, complex problem solving, effective communication and collaboration, and self-direction. Thus, an effective teacher professional development model is needed to help teachers learn and refine the pedagogical knowledge and skills needed to teach these skills [14]. However, many studies show that the TPD initiative model that emerges is not effective in supporting changes in teacher learning practices and student learning outcomes. For the example, one recent study found that even with large financial investments in TDP, both teacher performance and student learning showed little enhancement [15]. The study found that evaluating teacher performance remained the same even decreasing in the span of 2-3 years, while more than $18,000 of TDP per teacher had been spent. The causes are very diverse why the TPD implementation process is not successful including the implementation of TPD has not been based on an analysis of teacher needs, initial knowledge, the involvement of teachers in the previous TPD, and the teacher's expectations in developing himself as a professional teacher. While the entire domain of professional knowledge, professional practice and professional involvement, as well as the teacher's expectations about professional development greatly determine the success of the implementation of the teacher professional education program [3,16,17].

In addition, it is very important to understand the teacher's perspective on their teacher preparation program so that the TPD works well [18,19]. While [20] found a relationship between teacher perceptions of teacher preparation programs and their effectiveness as professional teachers. Furthermore, [20] notes that although a teacher's feelings for their preparation may not reflect their actual classroom practice, their preparation is highly correlated with teacher self-efficacy, which is also highly correlated with student achievement. The definition of results in teacher education programs and the ability to measure these correlations is very important to assist in the process of reform and policy making in teacher education [21]. This stage is even more important for the current education administration system that seeks to support their current teachers, beginning with understanding their preparation to play an active role in the overall professional development of teachers [22].

Given the urgency of the teacher's perspective on TDP so that TDP can be realized optimally, this study was conducted. Teacher's perspective regarding the development of TDP may not be identified without using instruments that are right on target. Therefore, we developed an instrument and validated it to examine whether the general items for the teacher preparation perception survey in this position function in an equivalent manner. This study examines the general set structure of the 30 items that we have developed. This survey involved 234 respondents as participants in the teacher professional education program at one of the state universities in the province of Lampung, Indonesia. This analysis provides an exploration of the constructs we made and ways to verify whether this construct is the same across groups of teacher categories such as age, length of teaching, gender, and teacher school graduates. But an exploratory in-depth analysis using Rasch models focuses on differences in gender views about the adequacy of teacher knowledge (content and pedagogy), teacher involvement in professional development, and teacher expectations for teacher development programs.

II. METHOD

A. Participants

Characteristics of the participants are detailed in table 1. Not all variables were collected for all participants, in part due to confidentiality concerns. Most the repsondents responding were young and graduate from a traditional teacher education program. Most of them responding had full-time involvement with the survey. Responses were received from 234 in-servive teachers who are participating in the TDP through teacher professional education programs in Indonesia.

TABLE I. DESCRIPTION OF THE SAMPLES

Variable	**Participants**	
Sex	***n***	***%***
Male	49	20.94
Female	184	79.16
Age		
Mean	42.98	
SD	11.12	
Graduation		
Public Teacher Education	154	
Private Teacher Education	80	
Length of Teaching		
Mean	20.07	
SD	11.10	

B. Instrument and Procedure

The survey, as described above, was created via literature review and a comprehensive analysis of sources of standards for teacher preparation, to define three competency areas [23], with one additional area suggested by an advisory panel. Items were written by project personnel and vetted through teacher education program coordinator and a peer advisory professor. After modifications based on a series of cognitive interviews, the survey was approved by a panel of deans of colleges of education in the college. The body of the survey for teacher candidates was split into three sections, with each section eliciting views about an area of teaching competency, engagement in TPD, and expectation about TPD. In total, the survey of in-service teacher profession education contained 30 perceive items which reflected overall satisfaction with the program. The survey sent to in-service teacher students of TPD in Lampung University was divided into the same three competency areas. Then, participants response the survey through a google form online link.

C. Analysis

The Rasch model presents uni-dimensional constructs arranged in a regular pattern of increasing throughout the same interval [24]. When data matches the Rasch model, estimates of materials and people are made by natural log transformation opportunities for raw data [25]. Rasch modeling is a broad collection of literature in the fields of education and social sciences [25,26]. The instruments examined through Rasch

analysis allow us to determine the extent to which items function to consistently measure one variable from easy to difficult in a monotonically and regularly increasing mode. The Rasch model consists of a model family that applies to dichotomous, polytomous, and continuous data. The Rasch rating scale model was used in this study because responses were given on a 0-4 point rating scale, with the same step scale used for all items [27]. Rasch analysis allows researchers to evaluate the extent to which uni-dimensional scales are made by items in size. A rasch fit index is used to determine whether each item or person contributes to a single construct measurement by assessing the extent to which an item or person performs as expected. That is, with the right items that are quite difficult to be supported by fewer people than easy items. Likewise, respondents with less measured construction (for example, class management competencies) support fewer "difficult" items than respondents with more constructs measured. Fit mean square is modeled to 1.0 when the data matches the model. In addition, the analysis of the residual main components is used to determine whether the second factor seems to be present in the data. Furthermore, [28] describes that instruments tend to be uni-dimensional if the variance shown by the first dimension is substantial, meanwhile eigenvalues for the first contrast (in the context of identical exploratory factors with eigenvalues for the second factor) are less than or equal to 2.0, and the variance marked by the first contrast is worth less than 5%. Goods reliability index and people estimate the replicability of the placement of goods and orders of people. Separation of persona identifies the number of subgroups of people who can be discriminated against by the instrument. Separation and reliability of separation illustrates reliability in different ways [29]. The Rasch reliability index, along with Rasch estimates of item difficulties and people's abilities, is based on linear size rather than raw or ordinal data and is more suitable for subsequent parametric calculations of mean and standard deviation [30]. Finally, Differential Item Functional Analysis (DIFF) was used to conduct the analysis of the differences of each item response toward the differences of gender.

III. Result and Discussion

Based on the results of research conducted, the data obtained were analyzed using the Rasch Model using the Winstep 3.73 application. Winstepor checks the average score per item and response category [31]. The research instrument was tested on 234 PPG students. The research instrument is a scaled data that measures three aspects, namely the teacher's perception of his ability, activities to improve professionalism, and expectations for professional development. The instrument tested was 30 items, each aspect had 10 question items.

A research instrument that can be used correctly has good validity and reliability, the instrument can be accessed with indicators that will be achieved. The results of the analysis obtained that 30 reliable questions with Cronbach alpha value of 0.95 were included in the category of questions with very good reliability. Cronbach's alpha value to measure reliability is the interaction between the same person and item in table 2 and table 3.

TABLE II. People's Realibility

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT		OUTFIT	
					MNSQ	*ZSTD*	*MNSQ*	*ZSTD*
MEAN	59.5	30.0	-1.60	.38				
S.D.	13.6	.0	1.78	.21				
MAX.	101.0	30.0	2.41	2.41				
MIN.	30.0	30.0	-7.81	-7.81	.11	-6.1	.10	-5.8
REAL RMSE		.47	TRUE SD	1.72	SEPARATION	3.66	Person RELIABILITY	.93
MODEL RMSE		.44	TRUE SD	1.73	SEPARATION	3.96	Person RELIABILITY	.93
S.E. OF PERSON MEAN = .12								

Person RAW SCORE-TO-MEASURE CORRELATION = .96

CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .95

Table 2 shows that the research instrument has a value of person reliability of 0.93 [32], meaning that the instrument is in a very good category. This shows that the respondent answered the whole item earnestly and not carelessly.

TABLE III. Reliability Items

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT		OUTFIT	
					MNSQ	*ZSTD*	*MNSQ*	*ZSTD*
MEAN	463.9	234.0	.00	.12	.99	-.2	1.00	-.1
S.D.	42.3	.0	.63	.01	.27	2.6	.28	2.4
MAX.	559.0	234.0	1.05	.13	1.58	5.2	1.68	6.0
MIN.	398.0	234.0	-1.34	.11	.62	-4.2	.62	-3.8
REAL RMSE	.13	TRUE SD	.62	SEPARATION	4.75	Item RELIABILITY	.96	
MODEL RMSE	.12	TRUE SD	.62	SEPARATION	4.99	Item RELIABILITY	.96	
S.E. OF PERSON MEAN = .12								

Table 3 shows the value of item reliability of 0.96 included in the special category [30], meaning that all items identified have accuracy with the model and are quality items so that items this can be used to measure teacher perceptions of their abilities, activities to improve professionalism, and expectations for professional development.

In addition, good instruments are instruments that do not contain bias. The instrument is said to be biased when there is one individual with certain characteristics more advantageous than an individual with other characteristics. For example, instruments are more easily answered by male students than females, this indicates the existence of gender-biased items. Detection of the overall bias items can be seen in table 4 and figure 1.

TABLE IV. DIFFERENTIAL ITEM FUNCTION (DF)

Person CLASSES	SUMMARY DIF CHI SQUARE	D.F.	PROB.	BETWEEN-CLASS MEAN-SQUARE	t=ZSTD	Item Number	Name
2	1.8941	1	.1687	.5933	.1326	1	K1
2	2.0556	1	.1517	.6394	.1776	2	K2
2	2.4292	1	.1191	.7648	.2900	3	K3
2	.0525	1	.8188	.0181	-1.0928	4	K4
2	.3597	1	.5487	.1180	-.6093	5	K5
2	.0249	1	.8745	.0120	-1.1640	6	K6
2	.0000	1	1.0000	.0007	-1.4651	7	K7
2	1.3681	1	0.2421	.4365	-.0407	8	K8
2	7.0999	1	.0077	2.2936	1.1476	9	K9
2	.6234	1	.4265	.2052	-.3986	10	K10
2	.0574	1	.8106	.0206	-1.0687	11	K11
2	.3254	1	.5684	.0975	-.6734	12	K12
2	.9675	1	.3253	.3149	-.2066	13	K13
2	.1927	1	.6607	.0625	-.8079	14	K14
2	.0000	1	1.0000	.0011	-1.4302	15	K15
2	1.1637	1	.2807	.3276	-.1234	16	K16
2	.0383	1	.8448	.0172	-1.1028	17	K17
2	.3178	1	.5729	.1034	-.6543	18	K18
2	.5996	1	.4387	.1902	-.4300	19	K19
2	.1478	1	.7007	.0475	-.8816	20	K20
2	.5754	1	.4481	.1901	-.4302	21	K21
2	.1443	1	.7041	.0422	-.9114	22	K22
2	.2432	1	.6219	.0724	-.7659	23	K23
2	.0401	1	.8413	.0125	-1.1575	24	K24
2	.2254	1	.6350	.0667	-.7895	25	K25
2	.3725	1	.5417	.1121	-.6270	26	K26
2	2.7494	1	.0973	.8821	.3846	27	K27
2	2.9040	1	.0884	.9308	.4213	28	K28
2	.1096	1	.7406	.0365	-.9461	29	K29
2	1.5756	1	.2094	.5259	.0623	30	K30

Table 4 shows that of the 30 instruments used in the study there was one instrument that contained gender bias, namely the K9 instrument with a probability value of 0.0077 (<0.05). This is in accordance with the statement Suminton and Widhiarso which reveals that a probability of more than 5% is considered to have no bias (difference) [32].

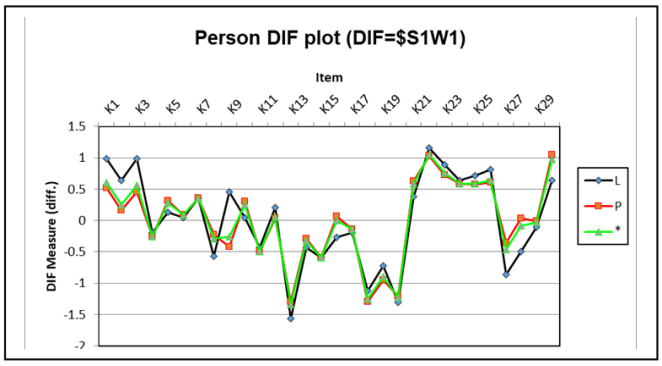

Fig. 1. Person DIF plot.

Figure 1 shows that there are three curves based on the sex of the student, namely L (male), P (female), and * (star) indicates the average value. From the graphic above, it can be seen in the K9 instrument that the ability of men to make learning media (including ICT-based) so that students can learn well, the knowledge L (Male) gained while studying was higher than that of P (Female). To find out specifically about the existence and absence of item bias in each aspect, namely the teacher's perception of their abilities, activities to improve professionalism, and expectations for professional development can be seen in table 5, table 6, and table 7.

TABLE V. DIF TEACHER'S PERCEPTIONS OF THEIR CAPABILITIES

Person CLASSES	SUMMARY DIF CHI SQUARE	D.F.	PROB.	BETWEEN-CLASS MEAN-SQUARE	t=ZSTD	Item Number	Name
2	1.4788	1	.2240	.4643	-.0073	1	K1
2	1.7186	1	.1899	.5336	.0707	2	K2
2	2.1485	1	.1427	.6781	.2138	3	K3
2	.0799	1	.7774	.0261	-1.0206	4	K4
2	2.0099	1	.1563	.6469	.1847	5	K5
2	.7432	1	.3886	.2375	-.3362	6	K6
2	.3029	1	.5821	.0995	-.6671	7	K7
2	4.7052	1	.0301	1.5060	.7816	8	K8
2	8.7709	1	.0031	2.8505	1.3579	9	K9
2	2.8058	1	.0939	.9039	.4011	10	K10

Table 5 shows that there are two instruments that contain biases, namely the K8 and K9 instruments with probability values sequentially of 0.0301 and 0.0031, meaning that there are differences in abilities between men and women, in detail can be seen in figure 2.

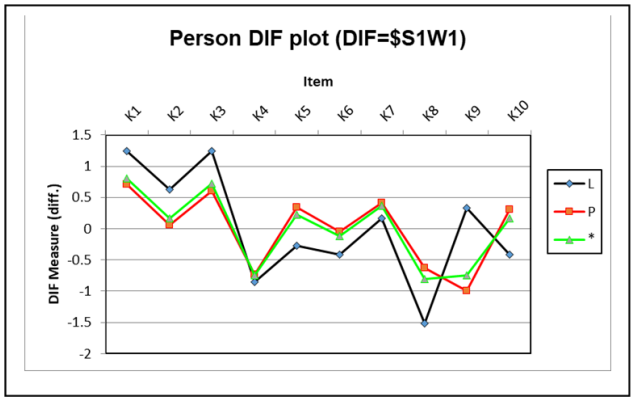

Fig. 2. Person DIF plot of teachers's perception of their capability.

Figure 2 shows that in the K8 instrument the ability of men to make good classroom action research, the research methodology L (male) obtained from college was lower than that of women. While, instrument K9 shows that women's ability to make learning media (including ICT-based) so students can learn well the knowledge that P (Female) get is lower than that of L (male). K8 and K9 survey item were.

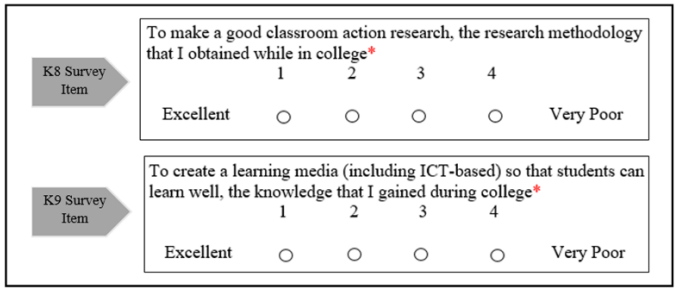

Fig. 3. K8 and K9 survey items.

TABLE VI. DIF ACTIVITIES TO IMPROVE PROFESSIONALISM

Person CLASSES	SUMMARY DIF CHI SQUARE	D.F.	PROB.	BETWEEN-CLASS MEAN-SQUARE	t=ZSTD	Item Number	Name
2	.1636	1	.6859	.0505	-.8657	1	KP1
2	.5125	1	.4740	.4740	-.5018	2	KP2
2	1.0607	1	.3031	.3031	-.1659	3	KP3
2	.1487	1	.6998	.6998	-.8786	4	KP4
2	.0274	1	.8685	.8685	-1.1898	5	KP5
2	1.2212	1	.2691	.2691	-.0947	6	KP6
2	.0173	1	.8954	.8954	-1.2272	7	KP7
2	.5481	1	.4591	.4591	-.4543	8	KP8
2	.9706	1	.3245	.3245	-.2138	9	KP9
2	.1161	1	.7333	.7333	-.9459	10	KP10

Table 6 shows that is no gender difference in items that measure activities to improve professionalism because the probability value of all items is more than 5%.

TABLE VII. DIF EXPECTATIONS FOF PROFESSIONALISM DEVELOPMENT

Person CLASSES	SUMMARY DIF CHI SQUARE	D.F.	PROB.	BETWEEN-CLASS MEAN-SQUARE	t=ZSTD	Item Number	Name
2	.2314	1	.6305	.0734	-.7619	1	PP1
2	1.1631	1	.2808	.3436	-.1642	2	PP2
2	1.2209	1	.2692	.3553	-.1474	3	PP3
2	.4820	1	.4875	.1377	-.5546	4	PP4

Table 7. Cont.

2	1.0303	1	.3101	.2990	-.2314	5 PP5
2	1.4357	1	.2308	.4192	-.0624	6 PP6
2	2.3137	1	.1282	.7875	.3091	7 PP7
2	2.8435	1	.0917	.9197	.4130	8 PP8
2	.0049	1	.9440	.0028	-1.3501	9 PP9
2	.8871	1	.3463	.2784	-.2647	10 PP10

Rasch Model analysis to measure aspects of teacher perceptions of their abilities, activities to improve professionalism, and expectations of professional development are used Mean Square values of Infit and Oufit as presented in figure 4, figure 5, and figure 6.

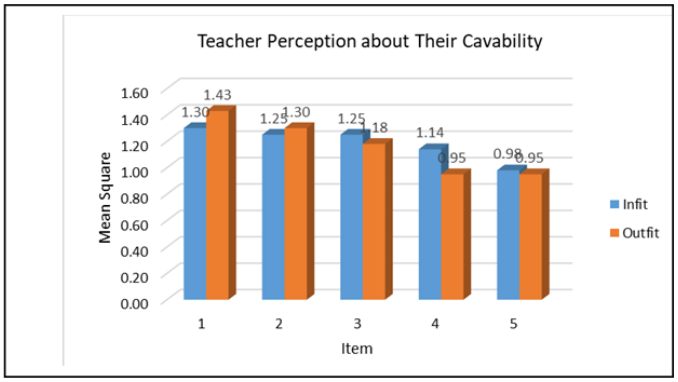

Fig. 4. Chart of infit and outfit measure per item teacher's perception of their capability

Figure 4 shows that the mean square fit values for infit and outfit per item all occupy a range of values from 0.5-1.7 where the item is not too difficult and not too easy [33,34]. The response category is based on the teacher's perception of his ability, there is no 100% who respond strongly or disagree. Based on the results of the study using Rasch Model analysis, in this study using confirmatory factor analysis that will optimize the results of the solution factors. Of the 10 items used to measure teacher perceptions of their abilities, it was found that all items had a high level of suitability based on the Rasch Model analysis.

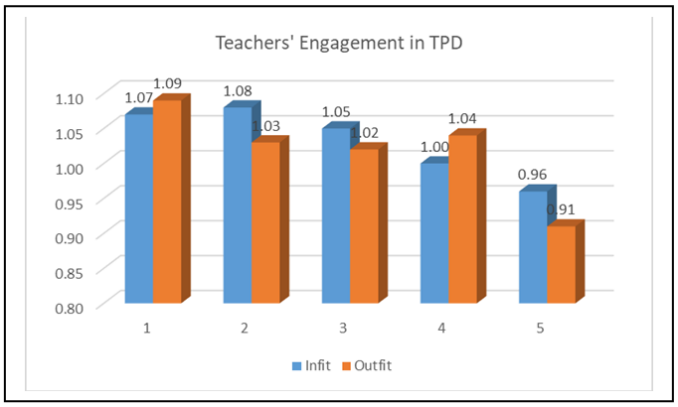

Fig. 5. Chart of infit and outfit measure per activity to improve professionalism.

Figure 5, shows that the mean square fit values for infit and outfit per item all occupy a range of values from 0.5 to 1.7 where the item is not too difficult and not too easy [33,34] Response categories are based on activities to improve professionalism, there are no 100% who respond strongly or disagree. The results of the study using Rasch Model analysis, where in this study using confirmatory factor analysis that will optimize the results of the solution factor. Of the 10 items used to measure activities to improve professionalism, it was found that all items had a high level of suitability based on the Rasch Model analysis define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, sc, dc, and rms do not have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable.

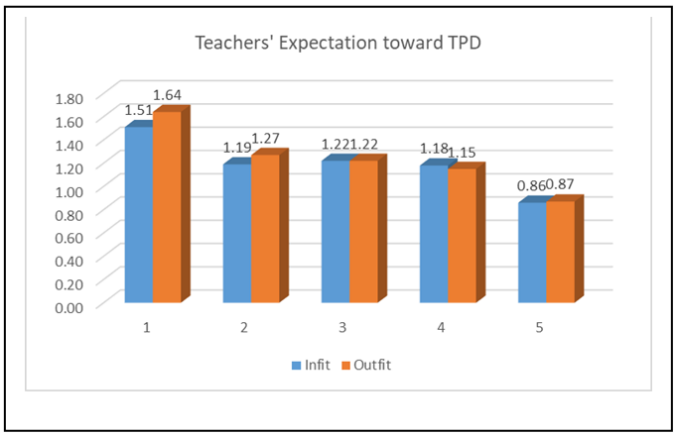

Fig. 6. Chart of infit and outfit measure per item of expectation for professional development.

Figure 6, shows that the mean square fit values for infit and outfit per item all occupy a value range from 0.5-1.7 where the item is not too difficult and not too easy [33,34]. Response categories are based on expectations for developing professionalism, there is no 100% who respond strongly or disagree. Based on the results of the study using Rasch Model analysis, in this study using confirmatory factor analysis that will optimize the results of the solution factors. Of the 10 items used to measure expectations for developing professionalism, it was found that all items had a high level of suitability based on Rasch Model analysis.

Based on the research that has been done it is found that the data that has been analyzed using Rasch and confirmatory factor analysis has produced an instrument with a statistically and conceptually elegant factor structure, and can be used for teacher surveys. This is consistent with research conducted by Grimbeek and Nisbet which states that the resulting factors proved acceptable in terms of exploratory factor analysis and

confirmation and in terms of analysis of Rasch items so that teachers see the reporting system numerical Queensland [35].

IV. CONCLUSION

The overall results of the Rasch model analysis on survey items on teacher perceptions in teacher professional education programs include three TDP attributes, namely the teacher's ability or teacher capability, activity or involvement in the TDP, and hopes for the implementation of TPD in teacher professional education, indicating that the survey instrument developed according to the quality of the instrument survey criteria with the validity and reliability of the items indicated by a consistent and consistent suitability with aspects of teacher's needs and perceptions in the teacher's professional education program. The higher significance of all factor items observed in the Rasch model analysis shows that instrument items can be used on a larger research scale to obtain a hypothetical model that is more complex than the teacher professional education development model. Important findings from this survey are still related to aspects of pedagogical content knowledge teachers (PCK) which require special emphasis in the TDP process through teacher professional education programs. In addition, there are still many teachers who feel lacking in mastering ICT, so most of them are not confident in developing ICT-based learning. PCK aspects of items. Furthermore, the analysis of gender differences using Differential Item Functioning (DIFF) shows that male teachers are better prepared for ICT than female teachers, but female teachers are superior in motivation to be actively involved in each TDP program. However, there was almost no significant difference between male teachers and female teachers about their perspectives on expectations in the TDP, all respondents wanted a more intensive and effective professional coaching process so that they achieved adequate competency.

ACKNOWLEDGMENT

We gratefully acknowledge valuable assistance from the members of the PPG Management Faculty of Teacher Training and Education University of Lampung that were involved in our study. Besides, this work is supported by the grant research from Faculty of Teacher Training and Education, University of Lampung 2018.

REFERENCES

[1] OECD (Organization for Economic Cooperation and Development), Education at a Glance 2016. OECD Indicators, Paris: OECD Publishing, 2016.

[2] D. Holtsch, J. Hartig, and R. Shavelson, Do Practical and Academic Preparation Paths Lead to Differential Commercial Teacher "Quality"?. Vocations and Learning, 2018, 1-24.

[3] L. Darling-Hammond, M.E. Hyler, and M. Gardner, Effective teacher professional development, Palo Alto, CA: Learning Policy Institute, 2017.

[4] I.S. Jenset, K. Klette, and K. Hammerness, "Grounding teacher education in practice around the world: an examination of teacher education coursework in teacher education programs in Finland, Norway, and the United States," Journal of Teacher Education, vol. 69, 2018.

[5] D. Reimer, B. Sortkear, M. Oskarsson, T. Nilsen, M. Rasmusson, and K. Nissinen, Northern Lights on TIMSS and PISA 2018. Nordic Council of Ministers, 2018.

[6] D.H. Gitomer, "Teacher Quality in a Changing Policy Landscape: Improvements in the Teacher Pool. Policy Information Report." Educational Testing Service, 2007.

[7] C. Campbell, K. Zeichner, A. Lieberman, and P. Osmond-Johnson, Empowered Educators in Canada: How High-Performing Systems Shape Teaching Quality, John Wiley & Sons, 2017.

[8] L. Darling-Hammond, S.P. Newton, and R.C. Wei, "Developing and assessing beginning teacher effectiveness: The potential of performance assessments", Educational Assessment, Evaluation and Accountability, vol. 25, pp. 179-204, 2013.

[9] D.N. Harris and T.R. Sass, "Teacher training, teacher quality and student achievement", Journal of public economics, vol. 95, pp. 798-812, 2011.

[10] L. Darling-Hammond, B. Berry, and A. Thoreson, "Does teacher certification matter? Evaluating the evidence," Educational evaluation and policy analysis, vol. 23, pp. 57-77, 2001.

[11] C. Jepsen and S. Rivkin, "Class size reduction and student achievement the potential tradeoff between teacher quality and class size," Journal of human resources, vol. 44, pp. 223-250, 2009.

[12] B. Rahman, A. Abdurrahman, B. Kadaryanto, and N.E. Rusminto, "Teacher-Based Scaffolding as a Teacher Professional Development Program in Indonesia," Australian Journal of Teacher Education, vol. 40, pp. 11, 2015.

[13] J.H. Holloway, "Connecting Professional Development to Student Learning Gains," Spring, vol. 15, pp. 37-43, 2006.

[14] L. Darling-Hammond, M.E. Hyler, and M. Gardner, Effective teacher professional development. Palo Alto, CA: Learning Policy Institute.

[15] TNTP, The mirage: Confronting the hard truth about our quest for teacher development. Brooklyn, NY: TNTP, 2015.

[16] OECD (Organization for Economic Cooperation and Development), Building a Teaching Profession: Lessons from around the World, Paris: OECD Publishing, 2011.

[17] K. Kiemer, A. Gröschner, M. Kunter, and T. Seidel, "Instructional and motivational classroom discourse and their relationship with teacher autonomy and competence support—findings from teacher professional development," European Journal of Psychology of Education, vol. 33, pp. 377-402, 2018.

[18] C. Donovan, K.E. Green, and K. Seidel, "Differential Item Functioning on a Measure of Perceptions of Preparation for Teachers, Teacher Candidates, and Program Personnel," Leadership and Research in Education, vol. 4, pp. 27-54, 2017.

[19] R.M. van der Lans, W.J. van de Grift, and K. Van Veen, "Developing an instrument for teacher feedback: using the rasch model to explore teachers' development of effective teaching strategies and behaviors," The journal of experimental education, vol. 86, pp. 247-264, 2018.

[20] L. Darling-Hammond, Securing the right to learn: Policy and practice for powerful teaching and learning DeWitt Wallace-Reader's digest distinguished lecture, Journal of Education, vol. 189, 2009, pp. 9-21.

[21] J. Perryman, S.J. Ball, A. Braun, and M. Maguire, "Translating policy: governmentality and the reflective teacher," Journal of Education Policy, vol. 32, pp. 745-756, 2017.

[22] C. Donovan, K.E. Green, and K. Seidel, "Differential Item Functioning on a Measure of Perceptions of Preparation for Teachers, Teacher Candidates, and Program Personnel," Leadership and Research in Education, vol. 4, pp. 27-54, 2017.

[23] K. Chval, S. Abell, E. Pareja, K. Musikul, and G. Ritzka, "Science and Mathematics Teachers' Experiences, Needs, and Expectations Regarding Professional Development," Eurasia Journal of Mathematics, Science & Technology Education, vol. 4, 2008.

[24] G. Rasch, Probabilistic models for intelligence and attainment tests (expanded edition), Chicago: University of Chicago Press, 1980.

[25] T.G. Bond and C.M. Fox, Applying the Rasch model (2nd Ed) , Mahwah: Lawrence Erlbaum Associates, 2007.

[26] B.D. Wright and M.H. Stone, Best test design, Chicago: MESA Press, 2004.

[27] B.D. Wright and G. Masters, Rating scale analysis. Chicago: MESA Press, 1982.

[28] J.M. Linacre, A user's guide to winsteps ministep 3.70.0: Rasch-model computer programs, Chicago: Winsteps, 2010.

[29] E.V. Smith Jr, "Evidence for the reliability of measures and validity of measure interpretation: a Rasch measurement perspective", Journal of applied measurement, vol. 2, pp. 282-311, 2001.

[30] C. Merbitz, J. Morris, and J.C. Grip, "Ordinal scales and foundations of misinference", Archives of physical medicine and rehabilitation, vol. 70, pp. 308-312, 1989.

[31] J.M. Linacre, WINSTEPS Rasch measurement computer program (Version 3.53) [Computer software], Chicago: Winsteps.com, 2004.

[32] B. Suminton and W. Widhiarso, Rasch Modeling Application on Educational Assessment, Cimahi: Trimkomunikata, 2015.

[33] R.M. Smith, R.E. Schumacker, and J. J. Bush, "Examining replication effects in Rasch fit statistics", Objective measurement: Theory into practice, vol. 5, pp. 303-317, 2000.

[34] B. Wright, "Reasonable mean-square fit values", Rasch Meas Trans, vol.8, 1994.

[35] P. Grimbeek and S. Nisbet, "Surveying primary teachers about compulsory numeracy testing: Combining factor analysis with Rasch analysis", Mathematics Education Research Journal, vol. 18, pp. 27-39, 2006.