

GDIF and DDIF analysis for posttest instrument of curriculum 2013 training

Komalasari

komalalpmpkalteng2015@gmail.com

Educational Quality Assurance Institution, Central Kalimantan, Indonesia

Abstract

The objective of this study was to analyze Gender Differential Item Functioning (GDIF) and District Differential Item Functioning (DDIF) in the post-test instrument of curriculum 2013 training in Central Kalimantan Indonesia. Data were secondary from Educational Quality Assurance Institution Central Kalimantan, which includes responses of 665 tenth grade teachers (414 females and 251 males; and 395 teachers from the new district and 269 teachers from the old district in Central Kalimantan Indonesia). The performed analysis of Differential Item Functioning (DIF) was applied to 30 multiple-choice items based on the Rasch model of Item Response Theory (IRT) using Conquest 4 software. The results showed that two items (item25 and item 30) exhibited gender-related DIF and five items (item 6, item15, item 22, item 24, and item 26) indicated district-related DIF. Female teachers tend to perform relatively better on analyzing a learning activity item. Otherwise, the concept of curriculum 2013 learning outcome (indicators of competence achievement) item to be relatively harder for female teachers. New district teachers tend to perform relatively better on lesson plan of curriculum 2013 and knowledge assessment (than old district teachers). The development of learning materials integrated with local content and the concept of competence for the attitude dimension of the learning process items seem to be relatively harder for new district teachers (than old district teachers). Some implications of this study are provided.

Keywords: Rasch model, DIF, gender, district, curriculum 2013

INTRODUCTION

Since a new national curriculum for primary and secondary school was implemented by Indonesian Government in 2013, the teachers must follow curriculum 2013 training. In every province of Indonesia, the training session of curriculum 2013 is conducted by LPMP (Educational Quality Assurance

Institution). The curriculum 2013 training is done in order to improve the competence of teachers and for the preparation of curriculum 2013 implementation exactly. It should be done since the implementation of the curriculum 2013 seems to be more complicated than those of the previous curriculum or school-based curriculum that is based on the school needs (Qamariah, 2015).

Educational Quality Assurance Institution of Central Kalimantan conducted the curriculum 2013 training for X grade senior high school teachers on 25-30 May 2017. It was conducted for six days or 39 hours at 20 Schools as the place of training activity. One of the assessments in curriculum 2013 training was post-test. The post-test of curriculum 2013 training was a cognitive assessment. It was held on the end of the training to assess the trainee's knowledge competency of curriculum 2013. It was also measured the extent of training effectiveness. The post-test consists of 30 items in the form of multiple choice question and 5 and prepared by Direktorat Pembinaan SMA Direktorat Jenderal Pendidikan Dasar dan Menengah Education and Cultural Ministry of Indonesia.

The participants of the training (the trainee) came from 14 districts/city and 82 schools around Central Kalimantan Province. The trainee were X grade senior high school teachers of 18 subjects and counseling teacher. Administratively, Central Kalimantan Province consists of 13 districts and 1 city. In this study, the term of old districts indicates the original district namely Palangkaraya, South Barito, North Barito, Kapuas, West Kotawaringin, and East Kotawaringin. Meanwhile, the new district is a newly formed district in 2002 consisting of 8 districts. They are Katingan, Seruyan, Sukamara, Lamandau, Gunung Mas, Pulang Pisau, Murung Raya Regency, and East Barito.

To get accurate information on training effectiveness, it is important to ensure that the cognitive assessment of teacher in curriculum 2013 training is equitable for all the teachers. Differential Item Functioning (DIF) analysis is an important part of a larger system of inquiry into issues of equity in assessments (Vista & Care, 2014). Thus, the examining of Differential Item Functioning (DIF) for curriculum 2013 training post-test instrument is very essential. This step is important to improve the quality of the items in the next version of the post test instrument of curriculum 2013 training so that it will be able to avoid gender biasness and district biasness. The utility of differential item functioning (DIF) analysis as a critical tool in test development.

Within the context of Rasch modeling an item is deemed to exhibit differential item functioning (DIF) if the response probabilities for that item

cannot be fully explained by the ability of the student and a fixed set of difficulty parameters for that item (Adam and Wu, 2010). Shepard, Camilli, & Averill (1981) define DIF as a statistical phenomenon that occurs when “two individuals with equal ability but from different groups do not have equal probability of success on the item.” In other words, DIF occurs when examinees from different groups show differing probabilities of success on (or endorsing of) the item after matching on the underlying ability that the item is intended to measure (Vista & Care, 2014).

Luc Le (2006) declares that commonly, DIF studies have examined cognitive tests for the presence of item DIF or potential test bias with respect to a number of different demographic characteristics, such as gender, education, social class, ethnicity, age and so on. In a study where the data of OECD PISA 2000 study were analyzed revealed that the gender difference in literacy scores was bigger for constructed response items than for multiple-choice items (Geske, A & Ozola, A, 2010). Gurian (2001) declares that girls do better in essay like tasks but boys in multiple-choice items since males are generally are better at deduction skills which are necessary for picking one answer from a set.

Luc Le (2006) declares that research on sources of DIF in science by gender has been reported in many studies: Some of them focus on item format effect where multiple-choice items seem to favor male examinees and open-ended items tend to favor female examinees. Some focus on the effect of item content where they found that males seem to be advantaged over females on physical science items and earth and space science items. Meanwhile, the interaction between items or domains with gender would be different from country to country (see TIMSS 2003 report, Mullis et al., 2004).

The main purpose of investigating country DIF here is to see if there is any relationship between the variations of item difficulty across countries and item characteristics (Luc Le, 2006). This study will investigate district DIF to identify if there is any relationship between the variations of item difficulty across two districts in Central Kalimantan Province (Indonesia) and item characteristics. The district is categorized as new district and old district.

Item response theory (IRT) has been widely used to detect differential item functioning (DIF) (Wang, 2000). Lord (1980) pointed out that item response functions are ideally suited to defining DIF. Since item parameters, as well as person parameters determine the functions, the detection of DIF could be made by comparing item parameters between a focal group and a reference group. Sheuneman and Subhiyah from the National Board of Medical Examiners (Bond and Fox, 2015), used merely an item estimates differences greater than

0.5 logits as the criterion for detecting DIF in 250-item medical certification test given to over 400 candidates.

Another research, conducted by Rosseni Din, et.al (2011) used three indicators to determine whether there is GDIF or not in the MeT instrument, namely: (i) t value of < -2.0 or > 2.0 (ii) DIF contrast value of < -0.5 or > 0.5 (iii) p (Probability) value < 0.05 or > -0.05 . ConQuest provides a powerful set of tools for examining DIF through the use of its multi-faceted modeling capabilities, and more particularly its ability to model interactions between facets, (Adams and Wu, 2010). Failures of variances (e.g., DIF) should alert us to potential problem with the measurement instrument or to new understanding about the underlying latent traits (Bond and Fox, 2015). Implications of DIF: DIF can increase or decrease sum scores (Wetzel & Hell, 2010).

The main objective of this study is using IRT (Rusch Model) approach to analyze DIF of the post-test items of curriculum 2013 training or variations of item difficulty parameters across the examinee groups by gender and district. This study concerns to detect the presence of differential item functioning (DIF) for a post-test instrument of curriculum 2013 training. In other words, this study will analyze the extent to which items function differently across the groups of teachers in curriculum 2013 training Central Kalimantan with gender and district as demographic variables. Items with a large DIF found during field testing have often been returned to item writers for review and revision or simply removed (Luc Le, 2006).

METHOD

The data used in this study were post-test items of curriculum 2013 training. The data were secondary from Educational Quality Assurance Institution Central Kalimantan Indonesia. The post-test of curriculum 2013 training was tested on teachers that follow the curriculum 2013 training in 2017 included 30 items of multiple choice question. All of the items were dichotomous (scored 0 or 1).

This study was population study. The data were composed of 665 tenth grade teachers (414 females and 251 males; and 395 teachers from the new district and 269 teachers from the old district in Central Kalimantan Indonesia). The summarized of data obtained can be seen in this following table.

Table 1 Data for DIF Analysis

Gender of tenth-grade teachers		District of tenth-grade teachers	
Female	Male	New	Old
414	251	396	269

The items of post-test of curriculum 2013 training that used for DIF analysis have been analyzed based on Rasch measurement model for fit statistic. All items of post-test of curriculum 2013 training are sufficient fit to the Rasch model.

The data analysis techniques in this study were performed to identify the gender differential item functioning (GDIF) and district differential item functioning (DDIF) in the instrument. The performed analysis of Differential Item Functioning (DIF) was based on the one parameter (Rasch) model of Item Response Theory (IRT). So only a difficulty of the item and teachers abilities are taken into account in this model.

Item analysis to determine if gender DIF exists in the post-test instrument of curriculum 2013 training was done using Conquest 4 software. The DIF analysis process is implemented in the following main steps:

- a. Calibrating items: Item difficulty parameter estimates by males and females from each of the studied groups are obtained separately by ACER ConQuest software (Wu et al., 1997).
- b. Computing DIF and Flagging DIF: After equating the item estimates by females into the scale of item estimates by males, the DIF value for each item is computed as the difference between the two relative difficulty estimates. The corresponding chi-square test for this difference is obtained from the DIF value and the standard errors of the estimates. To determine whether there is GDIF or not, three indicators were used namely: (i) t value of < -2.0 or > 2.0 (ii) DIF contrast or DIF magnitude value of $< -50\%$ or $> 50\%$ (iii) p (Probability) value < 0.05 or > -0.05 The three indicators were examined accordingly. An item need to meet those three conditions to be considered bias and be drop from the instrument. However, if the item meets only one of the conditions, it should not be drop.

FINDINGS AND DISCUSSION

Gender Different Item Functioning (GDIF)

ConQuest is used in this study to explore the existence of DIF with respect to gender in 30 items of multiple-choice test of curriculum 2013 training. It is a traditional DIF analysis because it is applied to dichotomously scored items and examines DIF between two groups — that is, it uses a binary grouping variable. Table 2 below shows the estimates for the gender differences in ability estimates.

Table 2. The estimates for the gender differences in ability estimates

Gender	Estimate	Error
Female	0.044	0.024
Male	-0.044	0.024
Chi-Square	3.49	
df	1	

A negative sign was used for the gender term in the item response model, so these results indicate that the male teachers have performed more poorly than the female teachers. The actual parameter estimate for the male teachers is almost two times larger than its standard error estimate, so the difference between the male and female means is obviously significant. The chi-square value of 9.63 on one degree of freedom is consistent with this finding. The conclusion that can be drawn here is that the male mean performance is lower than that of the females, this DOES NOT indicate differential item functioning. Further, the estimated difference of 0.088 is small at just over 8.8% of a teacher standard deviation.

Furthermore, the second table below gives the interaction between the item and gender facets.

Table 3. Interaction between the item and gender facets (item*gender)

Item	Female		Male	
	Estimate	Error	Estimate	Error
1	0.061	0.085	-0.061	0.085
2	0.027	0.083	-0.027	0.083
3	-0.084	0.096	0.084	0.096
4	-0.037	0.081	0.037	0.081
5	-0.053	0.081	0.053	0.081
6	-0.088	0.143	0.088	0.143
7	0.194	0.084	-0.194	0.084
8	-0.127	0.085	0.127	0.085
9	0.071	0.091	-0.071	0.091
10	-0.037	0.081	0.037	0.081
11	0.027	0.086	-0.027	0.086
12	-0.08	0.082	0.080	0.082
13	0.118	0.096	-0.118	0.096
14	-0.073	0.088	0.073	0.088
15	-0.11	0.087	0.110	0.087
16	-0.246	0.083	0.246	0.083

17	0.172	0.097	-0.172	0.097
18	-0.017	0.088	0.017	0.088
19	0.113	0.098	-0.113	0.098
20	-0.153	0.09	0.153	0.09
21	0.016	0.081	-0.016	0.081
22	0.128	0.114	-0.128	0.114
23	0.016	0.093	-0.016	0.093
24	0.226	0.098	-0.226	0.098
25	-0.302	0.087	0.302	0.087
26	-0.021	0.082	0.021	0.082
27	-0.007	0.095	0.007	0.095
28	-0.014	0.082	0.014	0.082
29	0.031	0.087	-0.031	0.087
30	0.250	0.098	-0.250	0.098

Chi-square = 50.76 Reliability = .427
Df = 29 sig Level = 0.007

The estimate of 0.061 for item 1 and females indicates that 0.061 must be added to the difficulty of this item for female teachers, similarly -0.061 must be added for the males. That is, male teachers found this item to be relatively easier than did the females. The results in this table show that sixteen items (item 3, 4, 5, 6, 10, 12, 14, 15, 16, 18, 20, and 25) are relatively easier for females than males. Fourteen items (item 1, 2, 7, 9, 11, 13, 17, 19, 21, 22, 23, 24, 29, and 30) are relatively easier for males than males. None item has the same difficulty. The significant chi-square (50.76, df=29) also shows the existence of DIF.

While this analysis has shown the existence of DIF in these items it is the magnitude of that DIF that will determine if the effect of that DIF is of substantive importance. Furthermore, to identify which items showed the existence of Gender Different Item Functioning (GDIF), this study used three indicators namely: (i) t value of < -2.0 or > 2.0 (ii) DIF contrast or DIF magnitude value of < -0.5 or > 0.5 (iii) p (Probability) value < 0.05 or > -0.05 . Table 3 indicates the magnitude and t value for each item.

Based on table 4 below, two items showed the existence of Gender Different Item Functioning (GDIF) namely items 25 and item 30. Item 25 meets all the three criteria to be given the verdict of gender bias. This item is easier for female as opposed to the male counterpart. This conclusion is drawn from the three pieces of evidence where first, the t value is -4.909 which is < -2.00 . Secondly, the DIF magnitude (DIF contrast) of 0.604 logits is way above 0.5. Lastly, it is evidenced by the p-value which is < 0.07 . Item 25 that is 60.4% of teachers standard deviation.

Item 30 meets all the three criteria to be given the verdict of gender bias. Male teachers found this item to be relatively easier than did the females. This conclusion is drawn from the three pieces of evidences where first the t value is 3.608 which is > 2.00 . Secondly, the DIF magnitude (DIF contrast) of 0.5 is way above 0.5. Lastly, it is evidenced by the p-value which is < 0.07 . Item 30 indicates that is 50% of teachers standard deviation.

Table 4. The magnitude and t value of the items

Item	Estimates		Error		Magnitude Value		t
	Female	Male	Female	Male	Log it	%	
Item1	0.061	-0.061	0.085	0.085	0.122	12.2	1.015
Item2	0.027	-0.027	0.083	0.083	0.054	5.4	0.460
Item3	-0.084	0.084	0.096	0.096	-0.168	-16.8	-1.237
Item4	-0.037	0.037	0.081	0.081	-0.074	-7.4	-0.646
Item5	-0.053	0.053	0.081	0.081	-0.106	-10.6	-0.925
Item6	-0.088	0.088	0.143	0.143	-0.176	-17.6	-0.870
Item7	0.194	-0.194	0.084	0.084	0.388	38.8	3.266
Item8	-0.127	0.127	0.085	0.085	-0.254	-25.4	-2.113
Item9	0.071	-0.071	0.091	0.091	0.142	14.2	1.103
Item10	-0.037	0.037	0.081	0.081	-0.074	-7.4	-0.646
Item11	0.027	-0.027	0.086	0.086	0.054	5.4	0.444
Item12	-0.08	0.08	0.082	0.082	-0.16	-16	-1.380
Item13	0.118	-0.118	0.096	0.096	0.236	23.6	1.738
Item14	-0.073	0.073	0.088	0.088	-0.146	-14.6	-1.173
Item15	-0.11	0.11	0.087	0.087	-0.22	-22	-1.788
Item16	-0.246	0.246	0.083	0.083	-0.492	-49.2	-4.192
Item17	0.172	-0.172	0.097	0.097	0.344	34.4	2.508
Item18	-0.017	0.017	0.088	0.088	-0.034	-3.4	-0.273
Item19	0.113	-0.113	0.098	0.098	0.226	22.6	1.631
Item20	-0.153	0.153	0.09	0.09	-0.306	-30.6	-2.404
Item21	0.016	-0.016	0.081	0.081	0.032	3.2	0.279
Item22	0.128	-0.128	0.114	0.114	0.256	25.6	1.588
Item23	0.016	-0.016	0.093	0.093	0.032	3.2	0.243
Item24	0.226	-0.226	0.098	0.098	0.452	45.2	3.261
Item25	-0.302	0.302	0.087	0.087	-0.604	-60.4	-4.909
Item26	-0.021	0.021	0.082	0.082	-0.042	-4.2	-0.362
Item27	-0.007	0.007	0.095	0.095	-0.014	-1.4	-0.104
Item28	-0.014	0.014	0.082	0.082	-0.028	-2.8	-0.241
Item29	0.031	-0.031	0.087	0.087	0.062	6.2	0.504
Item30	0.25	-0.25	0.098	0.098	0.5	50	3.608

GDIF for a post-test instrument of curriculum 2013 training can be described in figure 1 and 2.

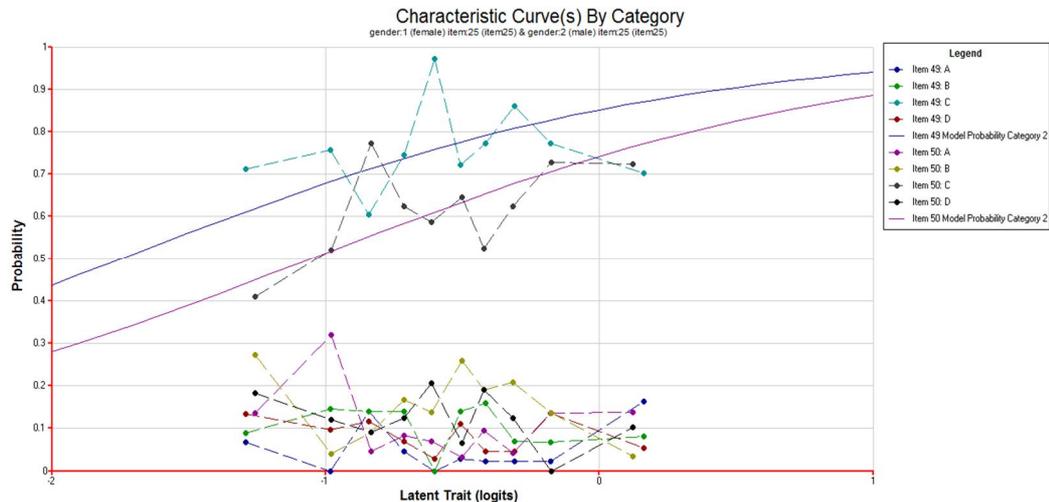


Figure 1: GDIF for Item25

For the twenty-fifth item, when referred to Fig. 1, the t value -4.909 , if rounded statistically will be -5.00 . DIF magnitude (DIF contrast value) is above 0.5 which is 0.65 . From the graph, the lines look far enough between male and female teachers. Going back to the item in question, it appears that item 25 is about analyzing a learning activity in the context of the 2013 curriculum and making a conclusion. In this matter, female teachers, in general, have been known to be able to analyze and conclude easier than male teachers counterparts. This does make sense because female tend to think complex than male. Besides, female tend to analyze things thoroughly than male. Even though item 25 is GDIF, this item should still retain from the instrument because both female and male teachers must able to analyze the learning activity, make the conclusion and follow up based on the result analysis.

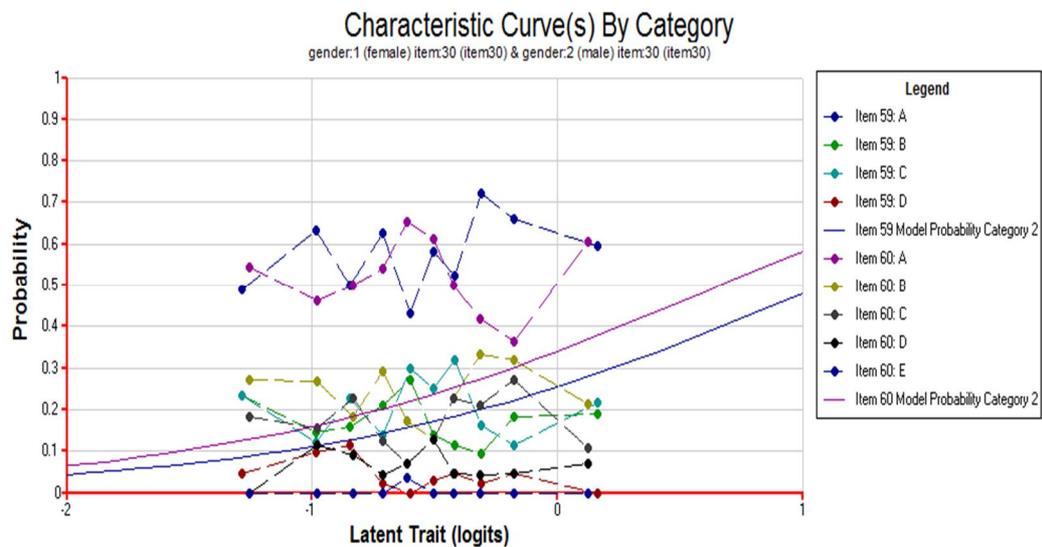


Figure 2: GDIF for Item30

Figure 2 above shows the GDIF on Item 30, when the t value is 3.608, DIF magnitude (DIF contrast value) is 0.5 which is the cut-off point for GDIF. From the graph, the lines look far enough between male and female teachers

Going back to the item in question, it appears that item 30 is about remembering and understanding the concept of curriculum 2013 learning outcome (indicators of competence achievement). In this matters, males teachers, in general, have been known to be able to remember and understand easier than female teachers counterparts. However, after analyzed the item, key answer and distractors, could be concluded that item 30 should be revised.

District Different Item Functioning (DDIF)

This study used ConQuest to explore the existence of DIF with respect to the district in 30 items of a multiple-choice test of curriculum 2013 training. The district was categorized as new district and old district. It is a traditional DIF analysis because it is applied to dichotomously scored items and examines DIF between two groups — that is, it uses a binary grouping variable. Table 5 below shows the estimates for the district differences in ability estimates.

Table 5. The estimates for the district differences in ability estimates

District	Estimate	Error
new	0.009	0.024
Old	-0.009	0.024
Chi-Square	0.19	
df	1	

A negative sign was used for the district term in the item response model, so these results indicate that the teachers from the old district have performed more poorly than the teachers of the new district. The actual parameter estimate for the new district is almost two times larger than its standard error estimate, so there is a difference between the new and old district. The chi-square value of 0.14 on one degree of freedom is consistent with this finding. The conclusion that can be drawn here is that the teacher of new district mean performance is higher than that of old district teachers, this DOES NOT indicate differential item functioning. Further, the estimated difference of 0.0 is small at just over 1.8% of a teacher standard deviation.

Furthermore, the fifth and sixth table below gives the interaction between the item and district facets and the magnitude and t value of the items.

Table 6. Interaction between the item and district facets (item*district)

Item	New district		Old district	
	Estimate	Error	Estimate	Error
1	-0.147	0.085	0.147	0.085
2	0.018	0.082	-0.018	0.082
3	-0.177	0.095	0.177	0.095
4	0	0.08	0	0.08
5	0.033	0.08	-0.033	0.08
6	-0.471	0.159	0.471	0.159
7	0.193	0.083	-0.193	0.083
8	0.079	0.083	-0.079	0.083
9	0.008	0.09	-0.008	0.09
10	0.173	0.08	-0.173	0.08
11	-0.08	0.085	0.08	0.085
12	-0.061	0.081	0.061	0.081
13	-0.149	0.098	0.149	0.098
14	-0.129	0.087	0.129	0.087
15	-0.288	0.087	0.288	0.087
16	0.023	0.08	-0.023	0.08
17	0.067	0.096	-0.067	0.096
18	0.127	0.089	-0.127	0.089
19	-0.179	0.1	0.179	0.1
20	0.211	0.087	-0.211	0.087
21	0.135	0.08	-0.135	0.08
22	-0.261	0.121	0.261	0.121
23	0.01	0.093	-0.01	0.093

24	0.286	0.097	-0.286	0.097
25	0.001	0.087	-0.001	0.087
26	0.263	0.083	-0.263	0.083
27	0.176	0.093	-0.176	0.093
28	0	0.081	0	0.081
29	0.08	0.087	-0.08	0.087
30	0.06	0.099	-0.06	0.099
Chi-square = 85.66 Reliability = .723				
Df = 29 sig Level = 0.000				

Table 7. The magnitude and t value of the items

Item	Estimates		Error		Magnitude Value		t
	Female	Male	Female	Male	Log it	%	
	Item1	-0.147	0.147	0.085	0.085	-0.294	
Item2	0.018	-0.018	0.082	0.082	0.036	3.6	0.310
Item3	-0.177	0.177	0.095	0.095	-0.354	-35.4	-2.635
Item4	0	0	0.08	0.08	0	0	0.000
Item5	0.033	-0.033	0.08	0.08	0.066	6.6	0.583
Item6	-0.471	0.471	0.159	0.159	-0.942	-94.2	-4.189
Item7	0.193	-0.193	0.083	0.083	0.386	38.6	3.288
Item8	0.079	-0.079	0.083	0.083	0.158	15.8	1.346
Item9	0.008	-0.008	0.09	0.09	0.016	1.6	0.126
Item10	0.173	-0.173	0.08	0.08	0.346	34.6	3.058
Item11	-0.08	0.08	0.085	0.085	-0.16	-16	-1.331
Item12	-0.061	0.061	0.081	0.081	-0.122	-12.2	-1.065
Item13	-0.149	0.149	0.098	0.098	-0.298	-29.8	-2.150
Item14	-0.129	0.129	0.087	0.087	-0.258	-25.8	-2.097
Item15	-0.288	0.288	0.087	0.087	-0.576	-57.6	-4.682
Item16	0.023	-0.023	0.08	0.08	0.046	4.6	0.407
Item17	0.067	-0.067	0.096	0.096	0.134	13.4	0.987
Item18	0.127	-0.127	0.089	0.089	0.254	25.4	2.018
Item19	-0.179	0.179	0.1	0.1	-0.358	-35.8	-2.531
Item20	0.211	-0.211	0.087	0.087	0.422	42.2	3.430
Item21	0.135	-0.135	0.08	0.08	0.27	27	2.386
Item22	-0.261	0.261	0.121	0.121	-0.522	-52.2	-3.050
Item23	0.01	-0.01	0.093	0.093	0.02	2	0.152
Item24	0.286	-0.286	0.097	0.097	0.572	57.2	4.170
Item25	0.001	-0.001	0.087	0.087	0.002	0.2	0.016
Item26	0.263	-0.263	0.083	0.083	0.526	52.6	4.481
Item27	0.176	-0.176	0.093	0.093	0.352	35.2	2.676

Item28	0	0	0.081	0.081	0	0	0.000
Item29	0.08	-0.08	0.087	0.087	0.16	16	1.300
Item30	0.06	-0.06	0.099	0.099	0.12	12	0.857

The estimate of 0.147 for item 1 and old district indicates that 0.147 must be added to the difficulty of this item for old district teachers, similarly -0.061 must be added for new district teachers. That is, new district teachers found this item to be relatively easier than did the old district teachers. The results in this table show that eleven items (item 1, 3, 6, 11, 12, 13, 14, 15, 19, 22, and 28,) are relatively easier for new district teachers than old district teachers. Seventeen items (item 2, 5, 7, 8, 9, 10, 16, 17, 18, 20, 21, 23, 24, 25, 26, 27, 29, and 30) are relatively easier for old district teachers than new district teachers. Two items (item 4 and item 28) has the same difficulty. The significant chi-square (85.66, df=29) also shows the existence of DIF.

While this analysis has shown the existence of DIF in these items, it is the magnitude of that DIF that will determine if the effect of that DIF is of substantive importance. Furthermore, to identify which items showed the existence of District Different Item Functioning (DDIF), this study used three indicators namely: (i) t value of < -2.0 or > 2.0 (ii) DIF contrast or DIF magnitude value of < -0.5 or > 0.5 (iii) p (Probability) value < 0.05 or > -0.05. Table 6 indicates the magnitude and t value for each item.

Based on table 6, five items show the existence of District Different Item Functioning (DDIF) namely item 6, item 15, item 22, item 24, and item 26. These items meet all the 3 criteria to be given the verdict of district bias. DDIF for post-test instrument of curriculum 2013 training can be described by figure 3 and 7.

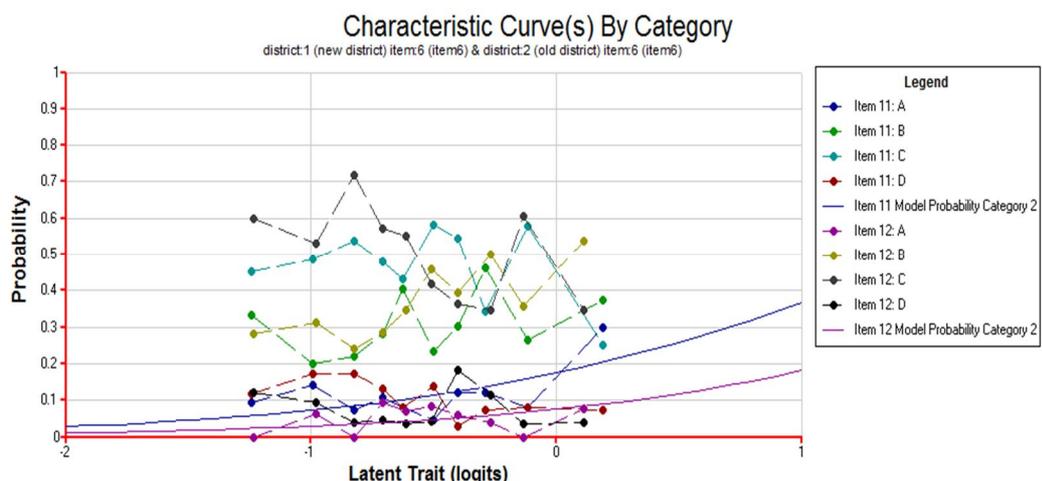


Figure 3. GDIF for Item6

For the sixth item, when referred to Fig. 3, the t value -2.446 which is < -2.00 . DIF magnitude (DIF contrast value) is below -0.5 which is -0.942 . This item is easier for new district teachers as oppose to the old district teachers counterpart. From the graph, the lines look far enough between new district and old district teachers. Going back to the item in question, it appears that item 6 is about remembering and understanding the specific concept of 2013 curriculum. In this matters, new district teachers in general have been known to be able to remember and understand easier the specific concept of 2013 curriculum than the old district teachers counterparts. However, after analyzed the item, key answer and distractors, could be conclude that the key answer was wrong and this item should be revised.

For the fifteen-item when referred to Fig. 4, the t value -4.682 which is < -2.00 . DIF magnitude (DIF contrast value) is below -0.5 which is -0.576 . This item is easier for new district teachers as oppose to the old district teachers counterpart. From the graph, the lines look far enough between new district and old district teachers. Going back to the item in question, it appears that item 15 is about analyzing the study case of knowledge assessment. In this matters, new district teachers in general have been known to be able to solve the study case easier than the old district teachers counterparts. Even though item 15 is DDIF, this item should still retain from the instrument because both new and old district teachers must able to analyze the case of knowledge assessment in order to implement curriculum 2013 in learning.

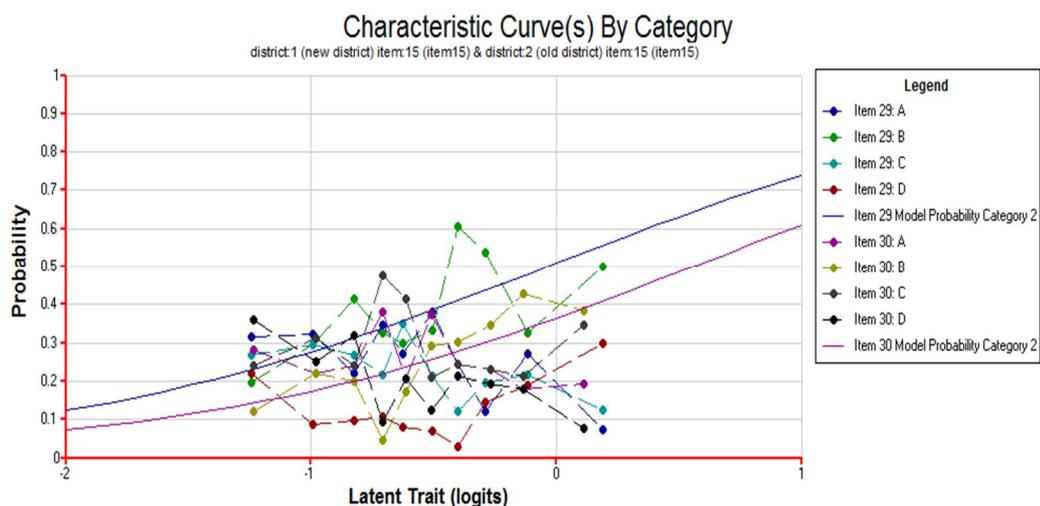


Figure 4 DDIF for Item15

Item twenty-two, when referred to Fig. 5, the t value -3.05 which is < -2.00 . DIF magnitude (DIF contrast value) is below -0.5 which is -0.522 . This item is easier for new district teachers as oppose to the old district teachers counterpart. From the graph, the lines look far enough between new district and old district teachers. Going back to the item in question, it appears that item 22 is about remembering and understanding the concept of lesson plan of curriculum 2013. In this matters, new district teachers in general have been known to be able to remember and understand lesson plan of curriculum 2013 than the old district teachers counterparts. Even though item 22 is DDIF, this item should still retain from the instrument because both new and old district teachers must able to remember and understand lesson plan of curriculum 2013 to implement curriculum 2013 in learning.

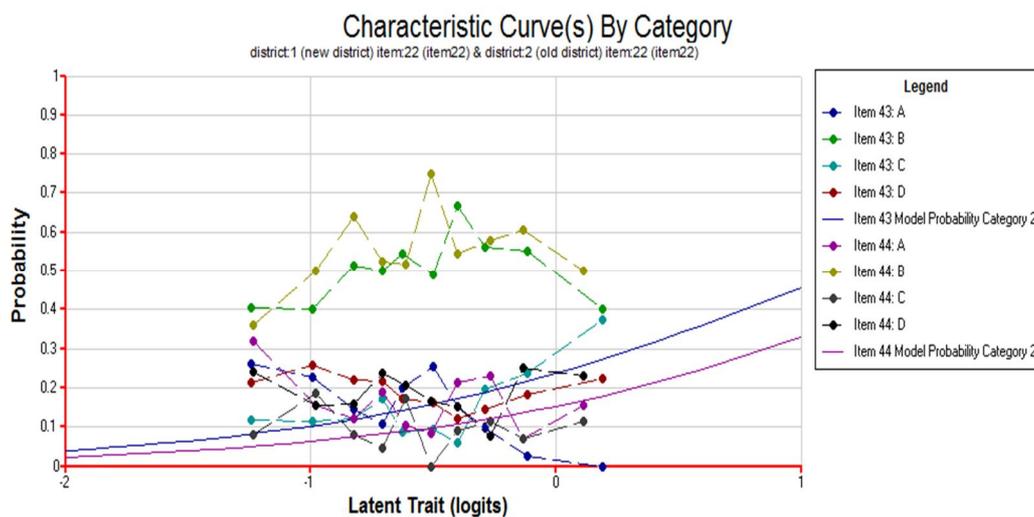


Figure 5. DDIF for Item 22

For the twenty-fourth item when referred to Fig. 6, the t value 4.17 which is > 2.00 . DIF magnitude (DIF contrast value) is above 0.5 which is -0.526 . This item is easier for old district teachers as oppose to the new district teachers counterpart. From the graph, the lines look far enough between new district and old district teachers. Going back to the item in question, it appears that item 24 is about understanding the development of learning materials integrated with local content. In this matters, the old district teachers in general have been known to be able to understand the development of learning materials integrated with local content than the new district teachers counterparts. Even though item 24 is DDIF, this item should still retain from the instrument because both new and old district teachers must able to understanding the

development of learning materials integrated with local content in order to implement curriculum 2013 in learning.

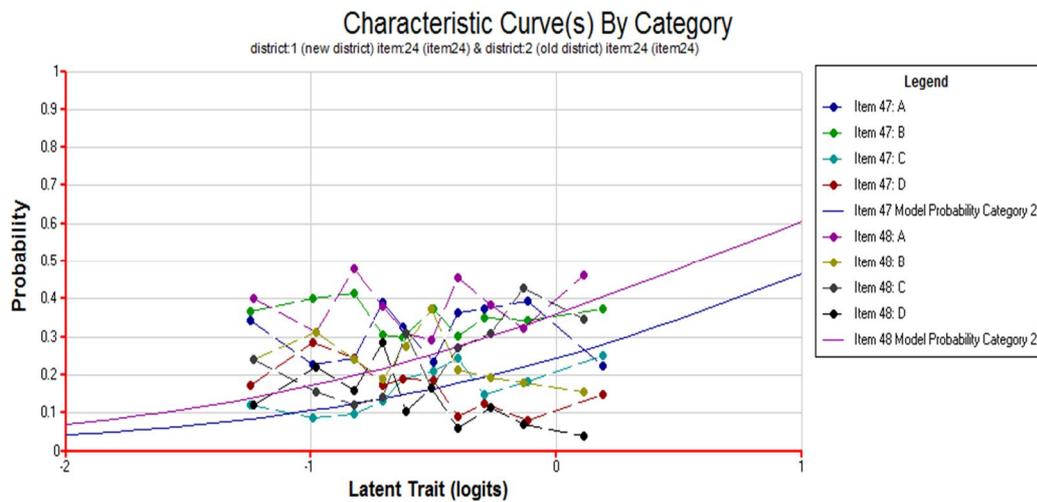


Figure 6. DDIF for Item24

For the last item as can be referred to in Fig. 7, item 26 meets all the 3 criteria to be given the verdict of district bias. This item is easier for old district teachers as oppose to the new district teachers counterpart. These conclusion is drawn from the three pieces of evidences where first, the t value is 4.48 which is > 2.00 . Secondly, the DIF contrast of 0.526 is way above 0.5. Lastly, it is evidenced by the p-value which is < 0.00 . From the graph, the lines look far enough between new district and old district teachers. Going back to the item in question, it appears that item 26 is about understanding the concept of competence for the attitude dimension of the learning process. In this matters, the old district teachers in general have been known to be able to understand the concept of competence for the attitude dimension of the learning process than the new district teachers counterparts. Even though item 26 is DDIF, this item should still retain from the instrument because both new and old district teachers must able to understanding the concept of competence for the attitude dimension of the learning process in order to implement curriculum 2013.

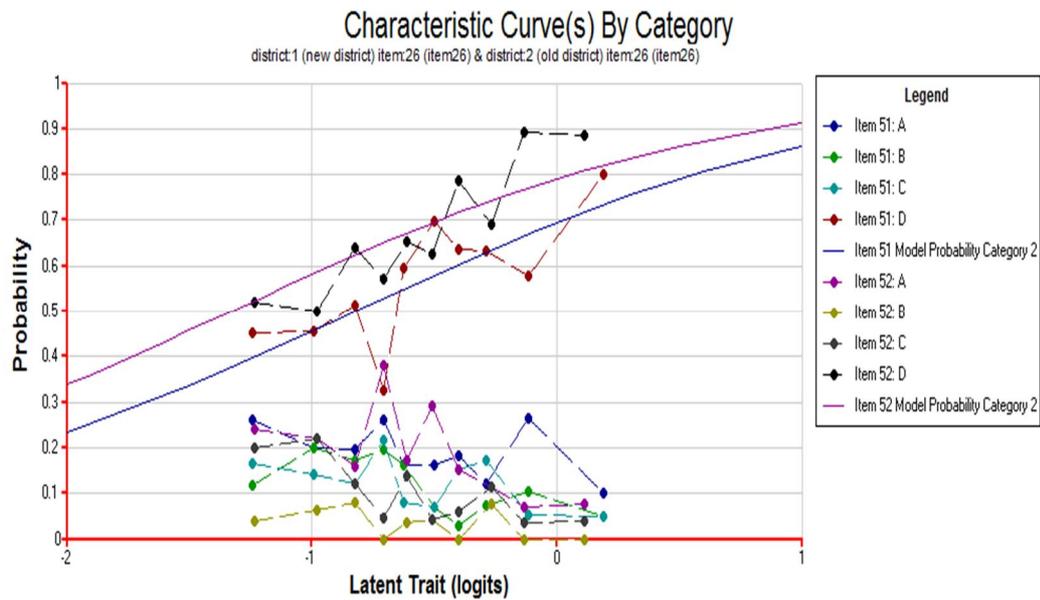


Figure 7. DDIF for Item26

There were several interesting issues in regards to the research findings. First, it was found that there was significant difference found between males and females teachers regarding analyzing a learning activity with complex/details information and making a conclusion. It shows that there is a difference in the ability of males and females teachers who participated in curriculum 2013 training to analyze a learning activity. Female teachers ability is better than males teacher. This finding is in line with an opinion from Rosseni, et al (2011) who states that females were found to be able to identify problems easier than males. This may be so due to their awareness towards details when males usually take things for granted. More often than not, males like to approach things in the simplest way possible which in turn may cause them to bypass some important details.

Secondly, new district teachers were found to be able for remembering and understanding the concept of lesson plan of curriculum 2013 and analyzing the study case of knowledge assessment easier than old district teachers. Meanwhile, regarding understanding the development of learning materials integrated with local content and understanding the concept of competence for the attitude dimension of the learning process, the old district teachers were found to be able than new district teachers. This interesting phenomenon should be a serious concern for the Educational Quality Assurance Central Kalimantan and relevant stakeholders in implementing curriculum 2013.

Lesson plan and knowledge assessment of curriculum 2013 should be a priority for curriculum 2013 assistance activities in schools for old district teachers. Otherwise, for the new district teachers, the development of learning materials integrated with local content and the concept of competence for the attitude dimension of the learning process should be a priority for curriculum 2013 assistance activities in schools.

CONCLUSION AND SUGGESTION

Based on the findings of this study, it is save to conclude that Differential Item Functioning (DIF) by gender was found to be lower than DIF by district in curriculum 2013 post-test instrument. Items 25 and item 30 show the existence of Gender Different Item Functioning (GDIF). Meanwhile, item 6, item 15, item 22, item 24, and item 26 show the existence of District Different Item Functioning (DDIF).

Item focus: Item 25 seems to be relatively easier for females teachers than male teachers. Otherwise, item 30 seems to be relatively easier for males teachers than female teachers. Item 6 and item 22 seem to be relatively easier for new district teachers than old district teachers. Meanwhile, the old district teachers seem to be relatively easier on item 24 and item 26 than the new district teachers.

Item content: female teachers tend to perform relatively better on analyzing a learning activity item. Otherwise, the concept of curriculum 2013 learning outcome (indicators of competence achievement) item to be relatively harder for female teachers. New district teachers tend to perform relatively better on lesson plan of curriculum 2013 and knowledge assessment (than old district teachers). The development of learning materials integrated with local content and the concept of competence for the attitude dimension of the learning process items seem to be relatively harder for new district teachers (than old district teachers).

Item 25 has GDIF but should be retained from the instrument because for implementing curriculum 2013, both female and male teachers must able to analyze the learning activity, make the conclusion and follow up based on the result analysis. Item 30 has GDIF and should be revised. Item 15, 22, 24 and 26 have DDIF but should be retained from the instrument because both new and old district teachers must able to understanding lesson plan, knowledge assessment, the development of learning materials integrated with local content and the concept of competence for the attitude dimension of the learning process. Meanwhile, Item 6 has DDIF and should be revised.

It is hoped that this study could inform the current curriculum 2013 training evaluation especially for post-test instrument in Central Kalimantan. The following recommendation could be drawn: Curriculum 2013 training programs for teacher need to direct the facilitators to be aware of the gender and district differences and to make them aware about their role to reduce gender and district differences. Curriculum 2013 training programs and curriculum 2013 assistance activities in schools need to provide specific strategies to encourage male teachers to analyze a learning activity with complex/details information and concluding. Specific strategies is vital in curriculum 2013 training programs and curriculum 2013 assistance activities in schools to develop the old district teachers ability in the lesson plan and knowledge assessment of curriculum 2013. Otherwise, for the new district teachers, the development of learning materials integrated with local content and the concept of competence for the attitude dimension of the learning process should be conducted in specific strategies as a priority for curriculum 2013 training and curriculum 2013 assistance activities in schools. Item that have been flagged as DIF need to be reviewed in order to retain or revise.

REFERENCES

- Adam, R & Wu, M. (2010). *Acer ConQuest Version 2.0 : Generalised Item Response Modelling Software*. Victoria: ACER Press.
- Bond, TG., & Fox, CM. (2015). *Applying the Rasch Model Fundamental Measurement in the Human Sciences*. New York : Routledge.
- Geske, A., Ozola, A. (2010) *Differential Item Functioning in the Aspect of Gender Differences in Reading Literacy* - www.iea.nl/sites/default/files/irc/IRC2010_Geske_Ozola.pdf
- Gurian, M. (2001). *Boys and Girls Learn Differently!: A Guide for Teachers and Parents*. San Francisco: Jossey-Bass
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Luc Le. (2006). *Analysis of Differential Item Functioning*. Paper prepared for the Annual Meetings of the American Educational Research Association in San Francisco, 7-11 April 2006.
- Mullis, I.V.S., Martin, M.O., Gonzalez, E.J., & Chrostowski, S.J. (2004), *TIMSS 2003 international mathematics report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.

- Qamariah, Z. (2015). Developing Islamic English instructional materials based on school-based curriculum. *Journal on English as a Foreign Language*, 5(2), 99-112.
- Rasch, G. (1960). *Probabilistic models for some intelligent and attainment tests*. Copenhagen: Institute of Educational Research. (Expanded edition, 1980. Chicago: The University of Chicago Press.).
- Rasch, G. (1960/1980). *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen: Danish Institute for Education Research, 1960. (expanded edition, Chicago: The University of Chicago Press, 1980.).
- Rossen, D., Verawati., M Faisal K., & Nur Ayu, J. (2011). Gender Differential Item Functioning (GDIF) Analysis for the Meaningful E-Learning Instrument. *Proceedings of the 10th WSEAS International Conference on E-Activities (E-ACTIVITIES '11)*. WSEAS Press.
- Shepard, L., Camilli, G., & Averill, M. (1981). Comparison of procedures for detecting test item bias with both internal and external ability criteria. *Journal of Educational Statistics*, 6(4), 317-375.
- Vista, A & Care, E. (2014). Differential Item Functioning and its Utility in an Increasingly Diverse Classroom: Perspectives from Australia. *Journal of Education and Human Development* June 2014, Vol. 3, No. 2, pp. 753-774.
- Wang, WC. (2000). The Simultaneous Factorial Analysis of Differential Item Functioning. *Methods of Psychological Research Online 2000*, Vol.5, No.1 Institute for Science Education
- Wetzel, E. & Hell, B. (2010). Differential item functioning in the AIST-R. www.iea.nl/sites/default/files/irc/IRC2010_Geske_Ozola.pdf
http://kalteng.go.id/ogi/viewarticle.asp?ARTICLE_id=1609
- Wright, B. D. and Stone, M. H. (1979). *Best Test Design*. Chicago: MESA Press.
- Wright, B. D. and Masters, G. N. (1982). *Rating Scale Analysis*. Chicago: MESA Press.
- Zumbo, B. D. (2007). Three Generations of DIF Analyses: Considering Where It Has Been, Where It Is Now, and Where It Is Going. *Language Assessment Quarterly*, 4(2), 223–233.
- Zumbo, B. D., & Gelin, M. N. (2005). A matter of test bias in educational policy research: Bringing the context into picture by investigating sociological/community moderated (or mediated) test and item bias. *Journal of Educational Research & Policy Studies*, 5(1), 1-23.