

# Differential item functioning in PLAB 1 items by sex and ethnicity: PLAB Annual Report 2024

Dr Matt Homer, Leeds Institute of Medical Education, University of Leeds

[m.s.homer@leeds.ac.uk](mailto:m.s.homer@leeds.ac.uk)

## Contents

List of tables .....	1
List of figures .....	2
Executive summary .....	3
Study overview .....	3
Key findings .....	3
Conclusions .....	3
Introduction .....	4
Methodology .....	4
The data samples .....	4
Rasch analysis and differential item functioning .....	6
Main findings .....	8
Aggregate percentage score distributions .....	8
Aggregate percentage scores by candidate sex and ethnicity .....	10
DIF analysis February 2024 by sex .....	14
DIF analysis May 2024 by sex .....	18
DIF analysis February 2024 by ethnicity .....	19
DIF analysis May 2024 by ethnicity .....	22
Discussion and conclusion .....	24
References .....	25
Appendix .....	26
DIF analysis February 2024 by sex full details .....	26
DIF analysis May 2024 by sex full details .....	35
The collapsing of ethnicity into three categories .....	39
DIF analysis February 2024 by ethnicity full details .....	40
DIF analysis May 2024 by ethnicity full details .....	48

## List of tables

Table 1: Items and candidates in sample data .....	5
Table 2: Candidates demographics for DIF analysis .....	6
Table 3: Items with DIF by sex for February 2024 .....	15
Table 4: Items with DIF by sex for May 2024 .....	19

Table 5: Items with DIF by ethnicity for February 2024.....	20
Table 6: Items with DIF by ethnicity for May 2024.....	23
Table 7: Summary DIF data by sex for February 2024.....	27
Table 8: Summary DIF data by sex for May 2024.....	35
Table 9: Summary DIF data by ethnicity for February 2024.....	41
Table 10: Summary DIF data by ethnicity for May 2024.....	49

## List of figures

Figure 1: Person-item distributions for February and May 2024.....	8
Figure 2: Aggregate percentage scores for February and May 2024.....	9
Figure 3: Aggregate percentage scores by sex for February and May 2024.....	11
Figure 4: Aggregate percentage scores by ethnicity in six categories for February and May 2024.....	12
Figure 5: Aggregate percentage scores by ethnicity in three categories for February and May 2024.....	13
Figure 6: ICC by sex for Item 098 from February 2024.....	16
Figure 7: ICC by sex for Item 125 from February 2024.....	17
Figure 8: ICC by sex for Item 062 from February 2024.....	18
Figure 9: ICC by ethnicity for Item 085 from February 2024.....	21
Figure 10: ICC by ethnicity for Item 050 from February 2024.....	21
Figure 11: ICC by ethnicity for Item 125 from February 2024.....	22
Figure 12: ICCs by sex for February 2024.....	34
Figure 13: ICCs by sex for May 2024.....	38
Figure 14: ICCs by ethnicity for Item 007 from February 2024.....	39
Figure 15: ICCs by ethnicity for February 2024.....	48
Figure 16: ICCs by ethnicity for May 2024.....	56

## Executive summary

### Study overview

This report quantifies the extent of differential item functioning (DIF) in PLAB1 items (questions) by *sex* and by *ethnicity*. DIF analysis is a psychometric method that investigates whether or not sub-groups of individual candidates at the same overall level of ability perform differently on specific items. Where DIF is present this can be seen as undermining the validity of the item and, perhaps, the overall exam and its associated outcomes, although evidence of DIF is not automatically evidence of actual bias in an item.

Two 2024 sittings of PLAB1 were investigated (February and May). A descriptive analysis was carried out at the exam level, and this was followed by a Rasch/DIF analysis by *sex* and *ethnicity* at the item level for each of the two sittings.

### Key findings

A preliminary descriptive analysis indicates more variation in PLAB1 total exam scores by *ethnicity* than by *sex* (Figure 3 and Figure 4).

The main DIF analysis identifies 11% and 6% of items with statistically significant levels of DIF by *sex* for February 2024 and May 2024 (pages 14 and 18 respectively). The corresponding percentages for DIF by *ethnicity* are 25% and 23% (pages 19, 22) - so that the prevalence of DIF is higher for *ethnicity* compared to *sex*.

Graphs (item characteristic curves) for items exhibiting DIF were inspected and a characterisation of each in terms of how particular sub-groups of candidates perform relative to each other is given (Table 3 to Table 6).

On aggregate, any apparent bias for or against particular sub-groups of candidates tends to balance out overall within each exam – as might be expected based on the overall methodological approach employed (page 4).

One additional finding of the Rasch analysis is that PLAB1 might benefit from greater proportions of more difficult items (page 7). This could help to ensure that the targeting of the exam more closely aligns with the typical ability of candidate groups, thereby improving the measurement from a psychometric point of view. However, there are reasons why this might also prove challenging in practice (e.g. changes to expectations regarding cut-scores – see page 8).

### Conclusions

More work is needed to better understand any causes of DIF by *sex* or *ethnicity*. Item writers and content specialists should review the numeric and graphical data for those items identified as having DIF. This will help to clarify if there are specific problems with the phrasing of items that might cause genuine bias, or perhaps indicate that there are PLAB1 content areas that are more difficult for certain groups of candidates of otherwise equal abilities for whatever reasons.

This work can help inform future developments of the PLAB1 exam.

## Introduction

There is an ongoing interest in issues of fairness and equity in assessment (Kakara Anderson et al., 2024; General Medical Council, 2024). This interest has influenced the subject of some recent PLAB reports and related research which has investigated various influences on scoring in PLAB exams to better understand sources of unwanted ('error') variance in these high-stakes assessments. Analysis of influences on overall candidate pass/fail decisions (Homer, 2022a) and, most recently, examiner effects on station-level scoring (Homer, 2023a; Homer, 2023b) have been carried out. To complement this earlier work, and to provide additional validity evidence regarding the PLAB1 applied knowledge exam, the current report investigates the extent of item (question) level potential bias<sup>1</sup> (Osterlind and Everson, 2009) by candidate *sex* and *ethnicity* in this exam.

In formal psychometric language, differential item functioning (DIF) is a statistical method that quantifies the extent to which individual items behave differently (i) across key sub-groups of candidates (i.e. might be biased by, for example, *sex* or *ethnicity*) – *at the same level of overall ability*. In other words, DIF analysis can identify items that have different average levels of difficulty across sub-groups having controlled for overall ability as measured in the exam. Usually, evidence of DIF is seen as a threat to the validity of the scores associated with the item, and reviewing such items sometimes shows that items are biased for or against particular sub-groups of candidates.

The greater the prevalence of DIF items, the more the validity of the overall exam and use of its pass/fail outcomes for high-stakes decision-making is undermined (Cook et al., 2015). However, it is important to note that the issue at the exam level is actually quite subtle because, by design, DIF analysis as employed in this report has to balance out overall. For example, if a single item shows DIF in favour of a particular sub-group of exam takers, then on aggregate the remainder of the items in the exam must show the opposite bias (Tennant and Pallant, 2007). This means that DIF cannot be used to detect bias in the exam overall but remains a very useful technique for highlighting specific items that have shown bias based on their psychometric performance. In turn, such analysis can be used to revise specific problematic items, and to inform future item and exam development.

In the remainder of this report, I describe the data samples and methodologies employed to quantify DIF, then present the findings and make some concluding remarks.

## Methodology

### The data samples

Anonymised item-level candidate scores (i.e. 0/1 for incorrect/correct) were analysed separately from the two most recently available PLAB1 exams - February and May sittings from 2024.

Each sitting was a 180-item exam, but initial screening indicated that both exams included two items each that had 0% facility (i.e. where all candidates had answered incorrectly)<sup>2</sup>. Such 'extreme' items essentially contribute nothing to the measurement process and cannot

---

<sup>1</sup> The word 'bias' is used here and throughout in a technical sense, and evidence of bias does not necessarily imply, for example, that item writers are genuinely biased against particular sub-groups of candidates.

<sup>2</sup> Items 53 and 109 from February, and 65 and 152 from May 2024.

be included in a DIF analysis. Hence, they were removed so that the final dataset for this report consisted of 178 items for each of the two PLAB1 exams analysed.

Candidate numbers for the two exams are shown in Table 1.

<b>Exam</b>	<b>Number of items analysed</b>	<b>Number of candidates</b>
<b>February 2024</b>	178	5,211 <sup>3</sup>
<b>May 2024</b>	178	5,204

**Table 1: Items and candidates in sample data**

Candidate *sex* and *ethnicity* were captured for each exam – and the profiles by these characteristics for the two examinations are shown in Table 2. These are the two facets that are to be used in the DIF analysis but, for *ethnicity*, it was decided to collapse the four smaller categories (Mixed, Other Ethnic Groups, Prefer not to say, and White) into one (Other) – as shown in the lower section of the table.

This collapsing has been done to avoid numbers of candidates within certain ethnic groups at specific ability levels becoming too small to make sense of in the DIF analysis for many items. I justify this approach in more detail in part of the Appendix (page 39).

---

<sup>3</sup> One candidate was removed as sex was missing.

Candidate characteristic	February 2024		May 2024	
	N	%	N	%
<b>Sex</b>				
<b>Man</b>	2,385	45.8	2,286	43.9
<b>Woman</b>	2,826	54.2	2,918	56.1
<b>Ethnicity (six groups)</b>				
<b>Asian or Asian British</b>	2,984	57.3	2,905	55.8
<b>Black or Black British</b>	1,193	22.9	1,178	22.6
<b>Mixed</b>	64	1.2	69	1.3
<b>Other Ethnic Groups</b>	601	11.5	614	11.8
<b>Prefer not to say</b>	181	3.5	242	4.7
<b>White</b>	188	3.6	196	3.8
<b>Ethnicity (three groups)</b>				
<b>Asian or Asian British (Asian)</b>	2,984	57.3	2,905	55.8
<b>Black or Black British (Black)</b>	1,193	22.9	1,178	22.6
<b>Other</b>	1,034	19.8	1121	21.6

**Table 2: Candidates demographics for DIF analysis**

The balance of sexes is relatively consistent across the two sittings – with women being in a small majority over men. In terms of *ethnicity*, there are two dominant categories with Asian or Asian British, and Black or Black British making up over three-quarters of examinees in total in both exam sittings.

### **Rasch analysis and differential item functioning**

A DIF study usually involves first carrying out an item response theory analysis of the items in the exam (Downing, 2003). For the current study, I have used the simplest, but arguably the most elegant, approach to such an initial analysis – a Rasch analysis within each PLAB1 exam (Bond and Fox, 2007; Panayides et al., 2010).

In brief, a Rasch analysis assumes that each item and each candidate have a specific level of difficulty and ability respectively. These difficulty/ability parameters are estimated during the analysis based on the sole assumption that the probability of a specific candidate getting a specific item correct is determined precisely by the difference between these two parameters. This makes intuitive sense – for example, a candidate with ‘high’ overall ability, will have a higher probability of getting a specific item correct compared to a candidate of lesser ability.

Once the initial Rasch analysis for each exam has been carried out, a DIF study for each candidate characteristic (e.g. *sex* or *ethnicity*) can be carried out to highlight items that show statistically significant differences in levels of difficulty of across sub-groups at the same

overall ability level. I include both graphical and tabular evidence when assessing the degree and nature of DIF for particular items.

I focus on what is called uniform DIF (Osterlind and Everson, 2009, p.10) – where the pattern of differential performance is similar (hence, uniform) across the ability range. Similar methods can be used to identify non-uniform DIF but there was little evidence of that in this study.

Further details of the methodology and interpretation of findings are given at the relevant points in the Main findings section. In terms of statistical software, I use SPSS (IBM Corp, 2021) for descriptive analysis and data cleaning, and RUMM2030 (Andrich et al., 2012) for the Rasch and DIF analyses.

Before presenting the main findings, I first give some details of the overall fit to the Rasch model, and an analysis of how well the candidate ability matches that of the items in the two PLAB1 exams.

#### Overall Rasch model fit

The overall fit to the Rasch model was good for both exams with high reliability and high person separation (all  $\geq 0.93$ ). The power of analysis of fit was classed as 'excellent' for both exams which gives confidence that item level analyses and measures of item fit are reliable. There was, however, some evidence of overall misfit to the Rasch model with the aggregate item-trait interaction giving a statistically significant result (indicating evidence of overall misfit). However, this is known to be a problematic test (Müller, 2020) and does not automatically undermine the efficacy of the DIF analysis. There was some evidence of individual item misfit to the Rasch model but the extent of this has not been specifically investigated in this report.

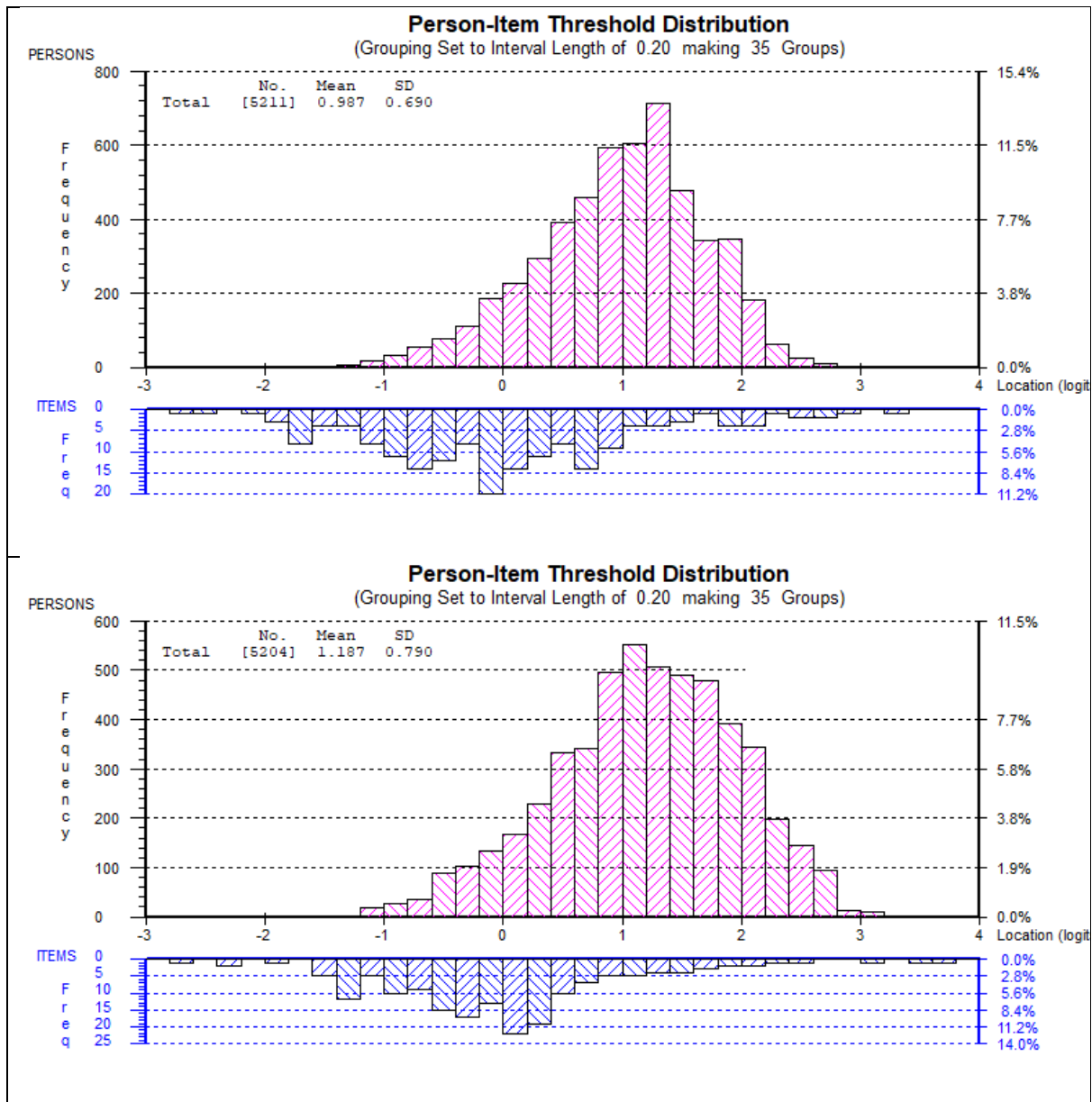
#### The targeting of the exams

A Rasch analysis can provide insight into the targeting of the exam under study – that is, how well the combined difficulty of the items matches (or otherwise) the overall ability of the candidate group. Ideally, the two distributions should match each other approximately – this matching is optimal in terms of providing the best accuracy of estimates for candidate ability and item difficulty.

Figure 1 show what are called the item-person distributions for the two PLAB1 exams. The top half of each graph shows the distribution of Rasch candidate ability estimates, whilst the bottom half shows the distribution of the estimates of the item difficulties.

One of the elegant properties of the Rasch model is that it places persons and items on the same (ability/difficulty) scale – from left (less able/easier) to right (more able/harder) – as demonstrated in Figure 1.

For both exams, Figure 1 indicates that there is good overlap between the two distributions of estimates. However, it is also clear that the peak of each distribution is higher for persons compared to items. From a strict measurement perspective, this implies that the exams might be improved (i.e. be more discriminating and provide better measurement) if they were to include greater proportion of more difficult items.



**Figure 1: Person-item distributions for February and May 2024**

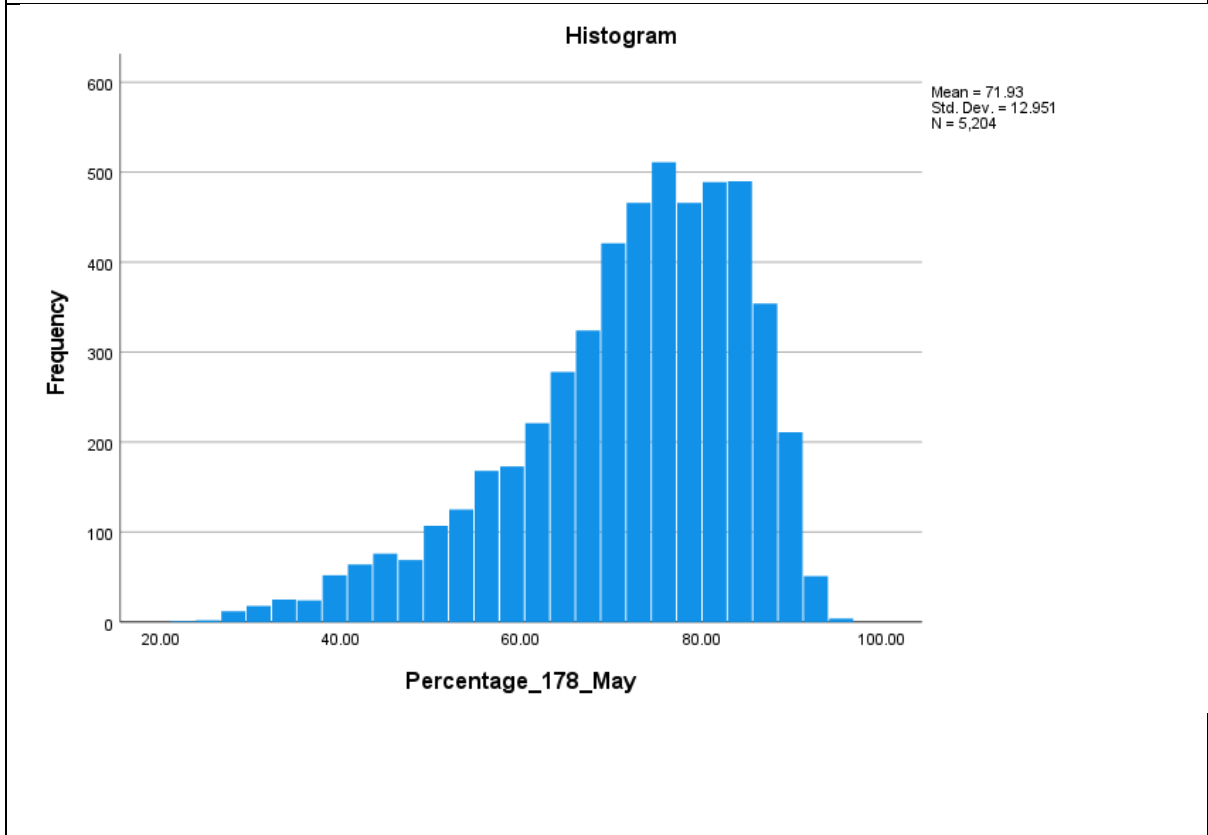
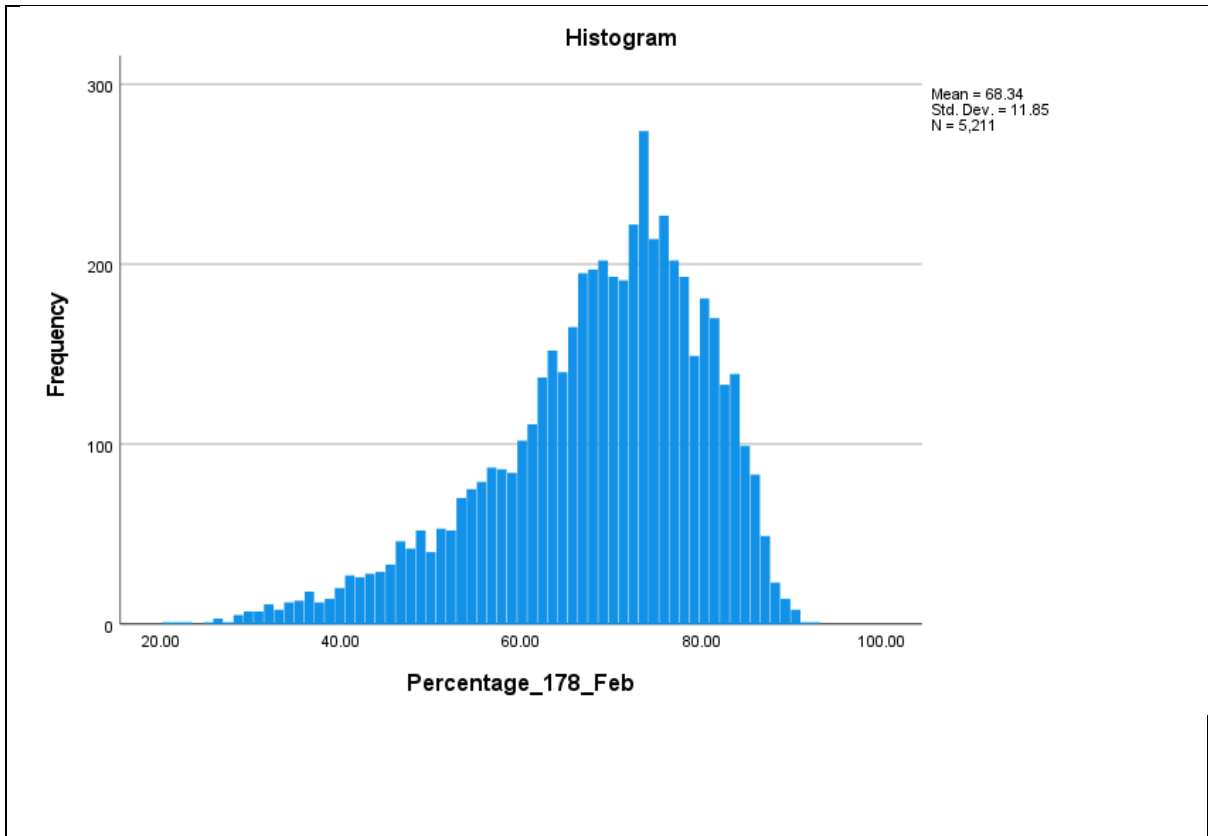
Having outlined the main methodological approaches to the analysis, I now present the main findings.

## Main findings

To provide important contextual data about the performance in the two PLAB1 exam outcomes, I begin the findings with analysis of aggregate score distributions across these, and of differences in aggregate performance by *sex* and *ethnicity*. Then I move on to the DIF analysis proper.

### Aggregate percentage score distributions

Histograms for the aggregate percentage scores for each exam are shown in Figure 2. Both distributions have some negative skew and a mean value of around 70%.

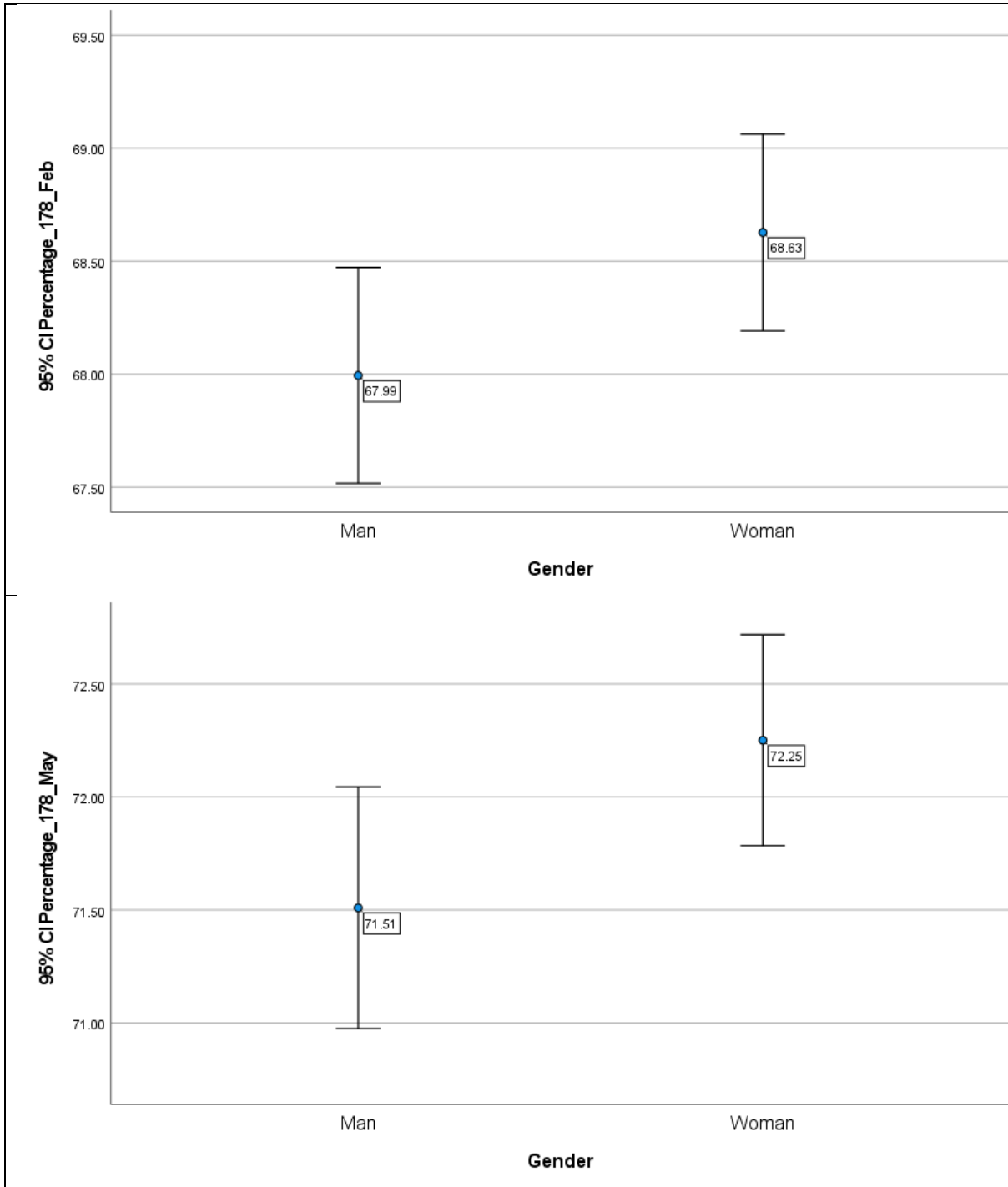


**Figure 2: Aggregate percentage scores for February and May 2024**

That the mean score on each exam is around 70% is consistent with the Rasch-based evidence around the exam targeting (page 7). From a strictly psychometric point of view, an exam with an average performance around 50% might be preferred, although it is recognised that there are other perspectives and factors that might make this a challenge – for example, candidate experience of the exam, the challenge of writing additional difficult items, and the normative expectations around the standard setting process.

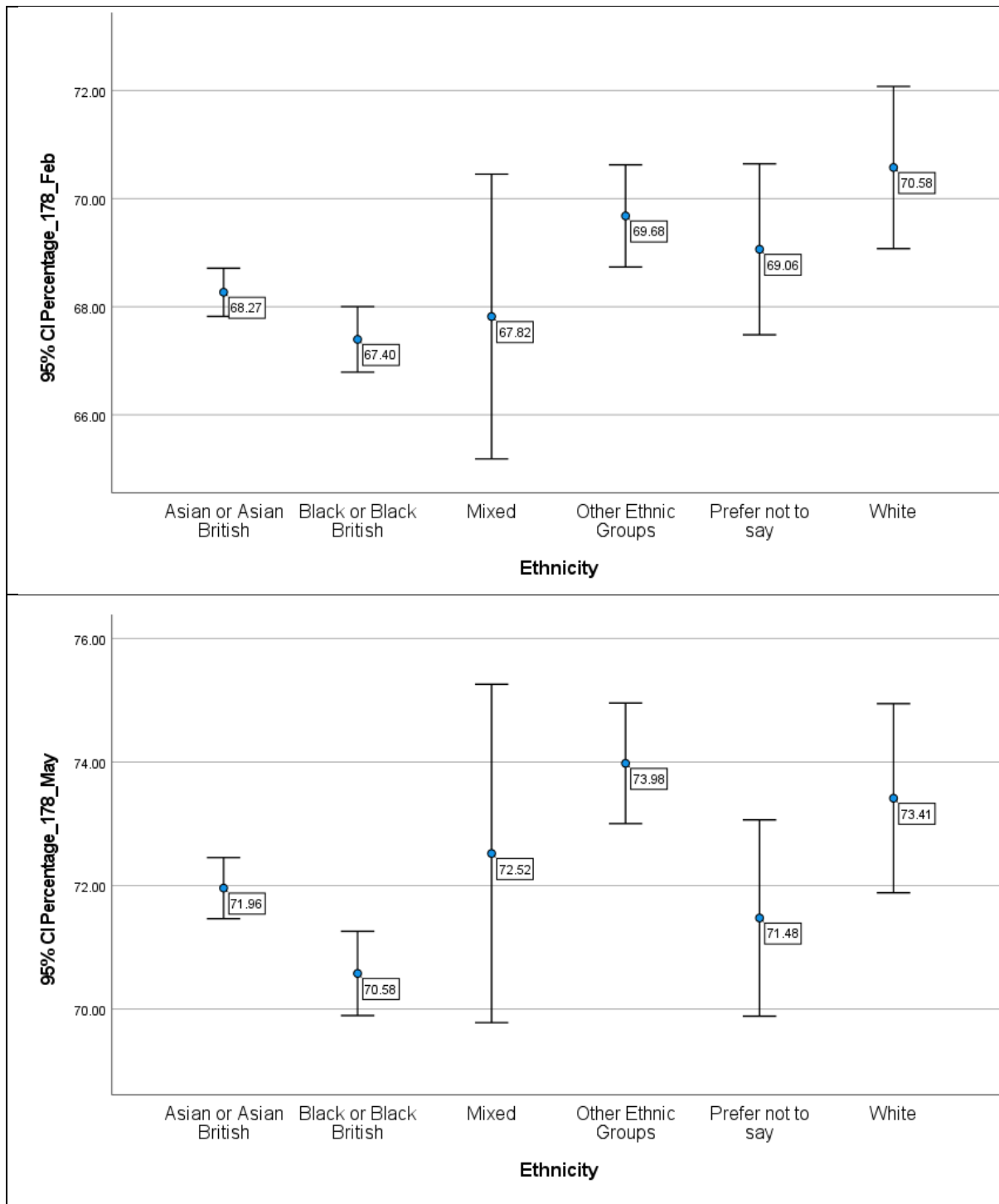
### **Aggregate percentage scores by candidate sex and *ethnicity***

Figure 3 shows the mean score and corresponding 95% confidence intervals for the aggregate exams scores by sex. Women are outscoring men by around 1 percentage point in both exams. The R-squared values for sex are approximately 0.001 for each exam indicating quite a small effect size for difference by sex.



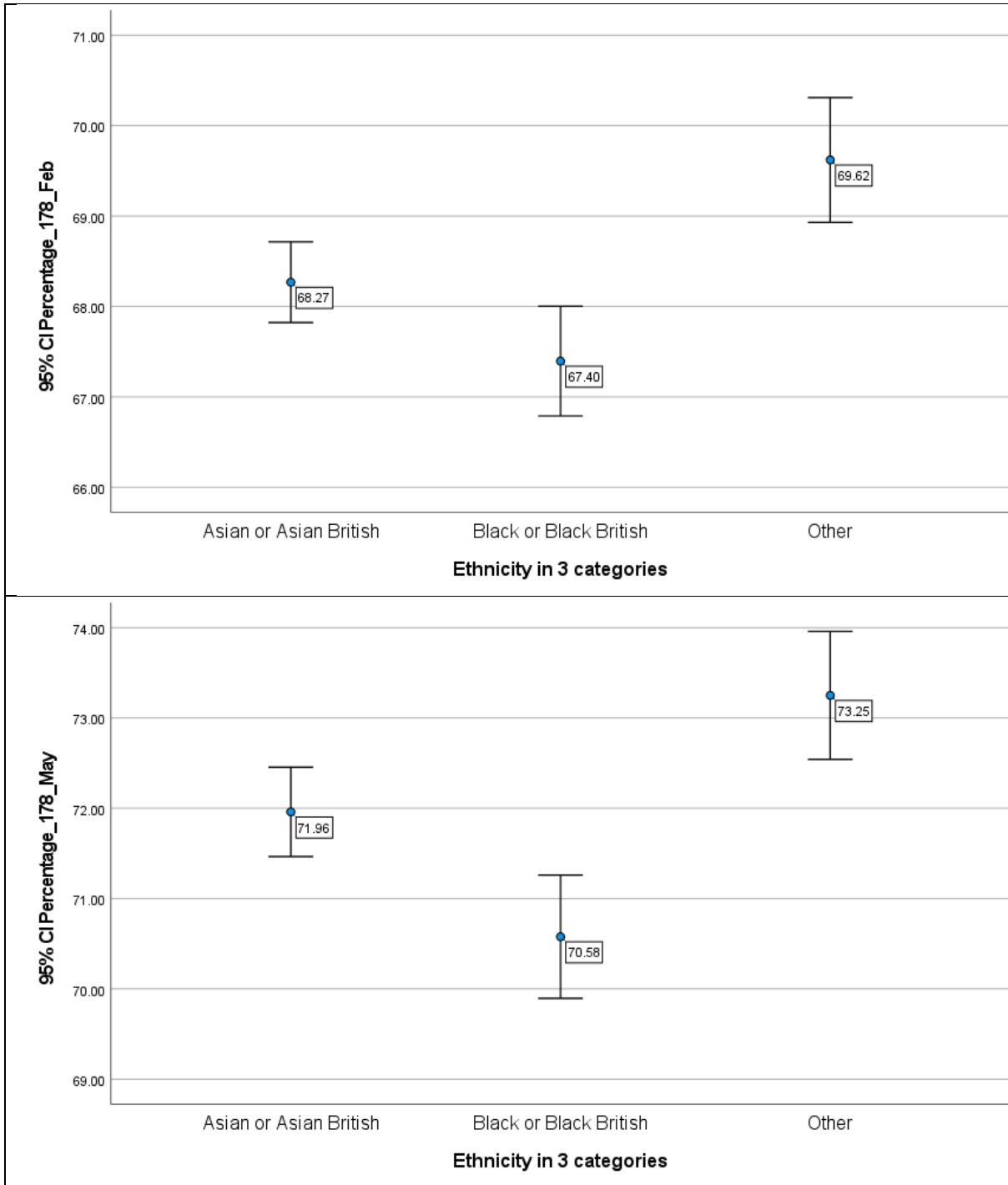
**Figure 3: Aggregate percentage scores by sex for February and May 2024**

Figure 4 shows the mean score and corresponding 95% confidence intervals for the aggregate exam scores by *ethnicity* in the original six categories. There are some differences in the pattern of mean values, and the R-squared values for *ethnicity* are 0.006 for both exams. Comparing with the analysis by sex, this suggests that differences in overall PLAB1 performance are greater by *ethnicity* than they are by *sex*.



**Figure 4: Aggregate percentage scores by ethnicity in six categories for February and May 2024**

For completeness, Figure 5 repeats the graphs in Figure 4 but with *ethnicity* collapsed into three categories as per the lower section of Table 2. For both exams, the combined Other group outperforms the Asian or Asian British, and Black or Black British groups (R-squared now of the order of 0.004 for both exams).



**Figure 5: Aggregate percentage scores by ethnicity in three categories for February and May 2024**

I now move on to the DIF analysis itself - first by sex for February and then May 2024, and then by *ethnicity*.

### **DIF analysis February 2024 by sex**

There are 20 items (11.2%) from the February 2024 exam identified with statistically significant DIF by sex. These items are listed in Table 3 – ordered approximately from most to least DIF.

Based on inspection of appropriate graphical evidence, Table 3 also includes a qualitative description of the precise nature of the DIF in four 'best fit' categories (men OR women outperforming the other group across OR across most of the ability range). I do this in a similar way for each DIF analysis in this report.

For February 2024, Table 3 shows that there are 12 items where women outperform men, and eight vice versa. Given that overall DIF must balance out across men and women (as discussed earlier in the Introduction), one would expect that there would be an approximate balance in the precise nature of DIF by sex identified.

<b>Item</b>	<b>Nature of DIF</b>
<b>I0098</b>	Men outperforming women across the ability range
<b>I0125</b>	Women outperforming men across the ability range
<b>I0085</b>	Women outperforming men across the ability range
<b>I0128</b>	Men outperforming women across the ability range
<b>I0062</b>	Men outperforming women across the ability range
<b>I0013</b>	Men outperforming women across the ability range
<b>I0160</b>	Men outperforming women across most of the ability range
<b>I0126</b>	Women outperforming men across the ability range
<b>I0076</b>	Men outperforming women across most of the ability range
<b>I0164</b>	Men outperforming women across most of the ability range
<b>I0172</b>	Women outperforming men across most the ability range
<b>I0034</b>	Women outperforming men across most of the ability range
<b>I0072</b>	Men outperforming women across most of the ability range
<b>I0152</b>	Women outperforming men across most of the ability range
<b>I0103</b>	Women outperforming men across most of the ability range
<b>I0047</b>	Women outperforming men across the ability range
<b>I0001</b>	Women outperforming men across most of the ability range
<b>I0180</b>	Women outperforming men across most of the ability range
<b>I0063</b>	Women outperforming men across the ability range
<b>I0130</b>	Women outperforming men across most of the ability range

**Table 3: Items with DIF by sex for February 2024**

The Appendix contains full details of statistical tests, and all graphs for these items (page 26).

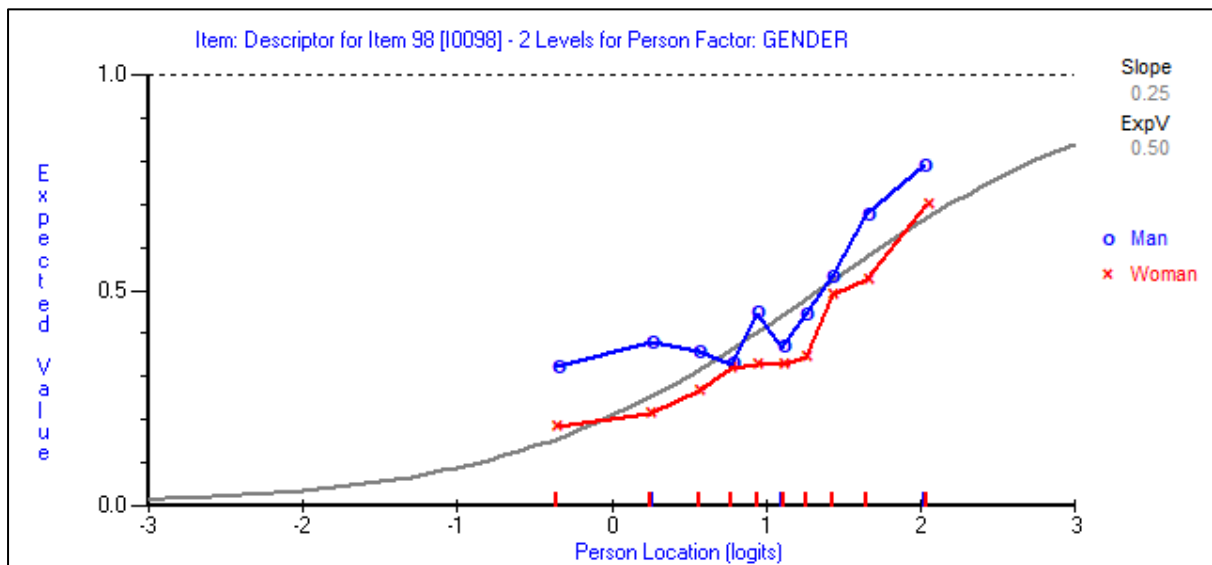
In this main text, I now focus on three items from this exam to graphically exemplify the nature of the differential item performance by sex, and to illustrate some specific psychometric issues.

Figure 6 shows what is known as an item characteristic curve (ICC) – for item 098. This indicates the expected likelihood of getting the item correct across the full ability spectrum based on the results of the Rasch modelling.

The horizontal axis shows the ability ('location') scale (from low ability, left, to high, right), whilst the vertical axis shows the predicted likelihood of getting the item correct.

The grey line shows the overall predicted Rasch model likelihood of getting the item correct for all candidates. In other words, this line represents the idealised Rasch model performance for an item of this difficulty.

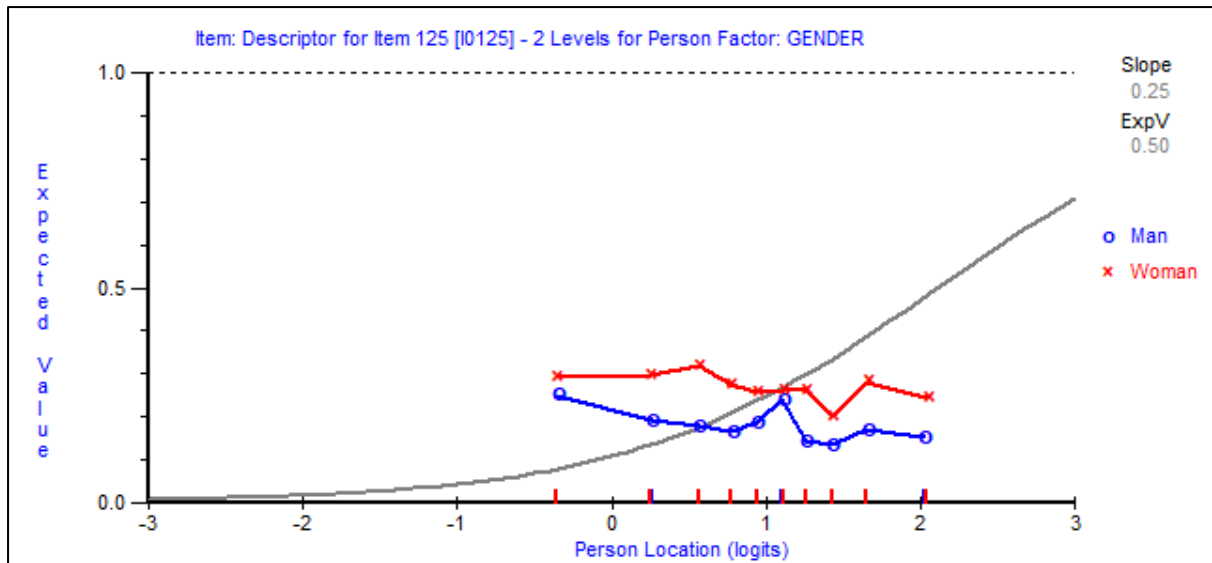
The blue (red) lines are the overall model-based likelihood of men (women) getting the item correct at specific levels of ability as determined automatically by the RUMM2030 software.



**Figure 6: ICC by sex for Item 098 from February 2024**

There is quite a large vertical gap between the blue and red lines across most of the ability range indicating that men are outperforming women on this item, having taken account of overall performance on the exam.

The two coloured lines are approximately either side of the grey line suggesting that the overall Rasch fit is reasonable. The next item is different in this regard – see Figure 7.



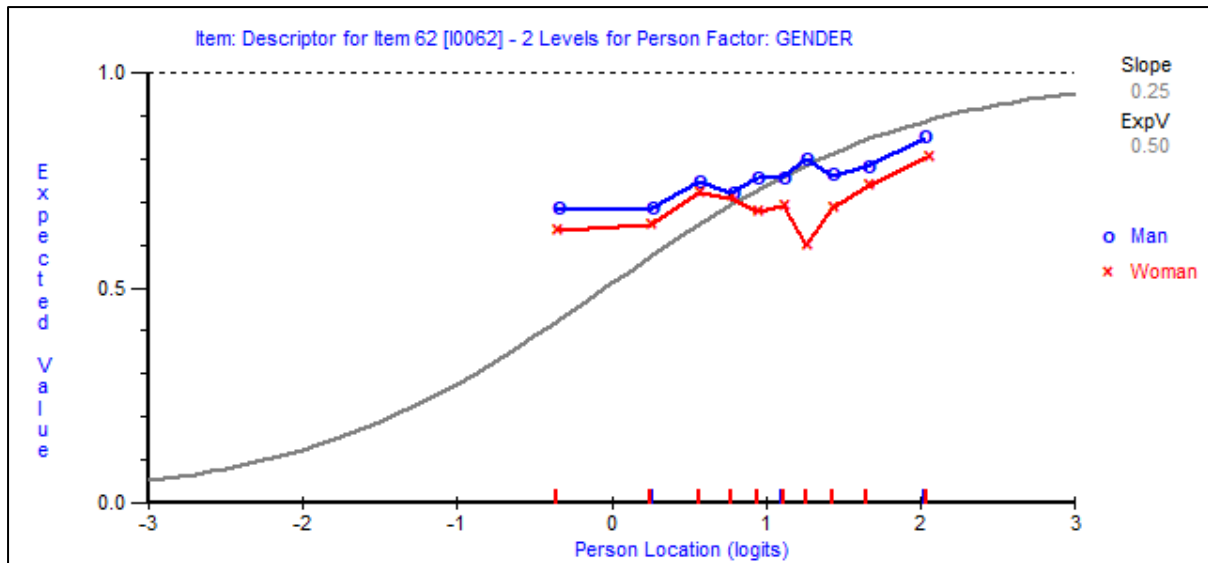
**Figure 7: ICC by sex for Item 125 from February 2024**

In Figure 7 the picture is quite different to that in Figure 6, with both coloured lines slopping downwards from left to right and crossing the grey line. This indicates that more able candidates overall of both sexes are less likely to get this item correct compared to those less able. In other words, it is a negatively discriminating item. It also quite a difficult item compared to that shown in Figure 6 – with relatively low likelihood of getting the item correct regardless of ability.

Differential item functioning is also playing a part in the performance of this item, with women more likely to get the item correct compared to men at the same overall ability level.

In summary, the Rasch/DIF analysis suggests that this item has two potential flaws – first that it is negatively discriminating, and second DIF by sex is present.

I move on to the final illustrative example from February 2024 (by sex) in Figure 8.



**Figure 8: ICC by sex for Item 062 from February 2024**

This is a relatively easy item as can be seen by the higher likelihood of getting it correct across the full ability range. Partially as a consequence, it is a poor discriminator - the two coloured lines are quite flat and do not track the grey line particularly well. This implies that the fit to the Rasch model is relatively poor for this item. However, there is also clearly some DIF in favour of men across the ability range.

The next sub-section summarises the DIF analysis by sex for the second PLAB1 exam, that administered in May 2024.

### **DIF analysis May 2024 by sex**

The number of DIF items for May 2024 is 10 (5.6%), half of that from the February sitting. These items are listed with associated DIF categorisation in Table 4, which shows that there is balance between those items where men have outperformed women (5) and vice versa (5). Beyond this, there are no additional issues worthy of exemplification beyond those already discussed in the previous analysis for February.

The Appendix contains full details of statistical tests, and all graphs for these items (page 35).

Item	Nature of DIF
I0105	Men outperforming women across the ability range
I0170	Women outperforming men across the ability range
I0005	Women outperforming men across most of the ability range
I0127	Men outperforming women across the ability range
I0010	Women outperforming men across the ability range
I0133	Women outperforming men across most of the ability range
I0112	Women outperforming men across the ability range
I0078	Men outperforming women across most of the ability range
I0033	Men outperforming women across the ability range
I0143	Men outperforming women across the ability range

**Table 4: Items with DIF by sex for May 2024**

The next sub-sections move on to a summary of the DIF by *ethnicity* analysis – for February and then May 2024.

#### **DIF analysis February 2024 by *ethnicity***

There are 45 items (25.3%) from the February 2024 exam identified with statistically significant DIF by *ethnicity*. The 20 items with the most DIF are listed in Table 5 – ordered approximately from most to least DIF. The Appendix contains full details of statistical tests (Table 9), and all item characteristic curves (ICC) for these 20 items (Figure 15). Based on inspection of these graphs, Table 5 also includes a qualitative description of the precise nature of the DIF in a range of ‘best fit’ categories. Given that there are three *ethnic* categories this is more complex than it was for sex.

Looking across Table 5 there are a range of DIF categorisations that, to an extent, balance out across the table (as might be expected).

Item	Nature of DIF
I0085	Asian group performing best across the ability range
I0098	Other group performing best across the ability range
I0159	Black group performing best across the ability range
I0083	Asian group performing best across the ability range
I0014	Asian and Other group performing best across most of the ability range
I0089	Black and Other group performing best across the ability range
I0115	Black and Other group performing best across the ability range
I0140	Black group performing best across most of the ability range
I0087	Asian and Other groups performing best across most of the ability range
I0096	Asian group performing best across most of the ability range
I0050	Asian and Other groups performing best across most of the ability range
I0157	Other group performing best across most of the ability range
I0101	Black and Other group performing best across most of the ability range
I0007	Black and Other group performing best across most of the ability range
I0158	Asian group performing best across some of the ability range
I0009	Other group performing best across some of the ability range
I0064	Black group performing best across most of the ability range
I0099	Black group performing best across most of the ability range
I0125	Asian group performing best across most of the ability range
I0162	Other group performing best across most of the ability range

**Table 5: Items with DIF by ethnicity for February 2024**

As with the analysis of DIF by sex, I choose three items from this exam to graphically exemplify the nature of the differential item performance by *ethnicity*.

Figure 9 shows the ICC for item 085 – this is the item with the largest DIF by ethnicity in the February 2024 exam. Across the full ability range, the blue (Asian) line is above the green (Other) and red (Black) lines. This means that Asian candidates are more likely to answer this item correctly at the same overall ability level compared to Other and Black candidates.

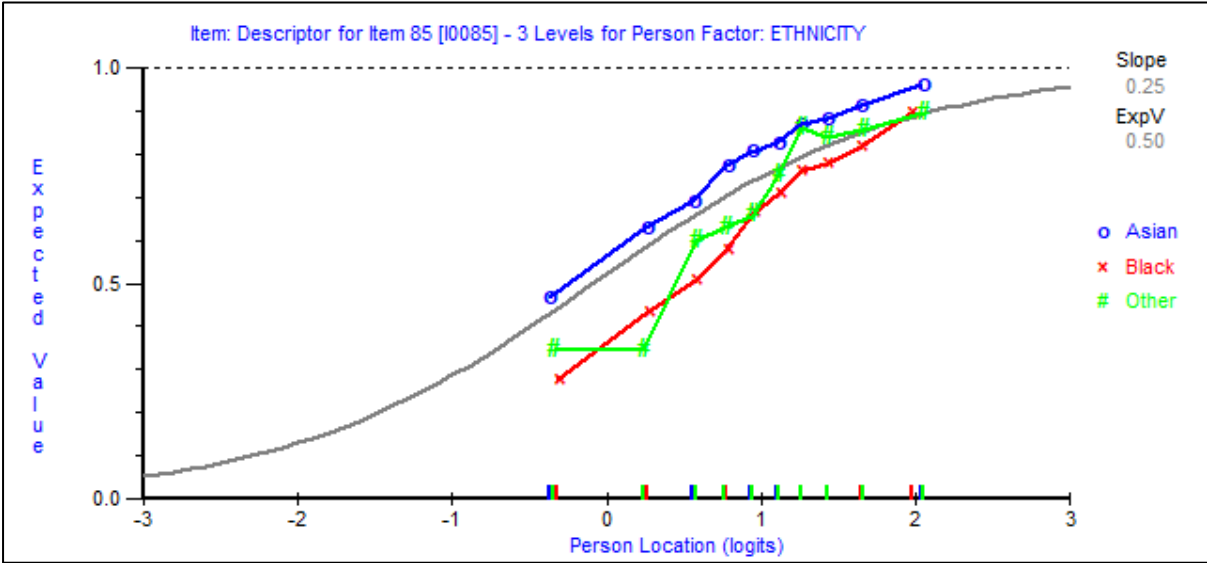


Figure 9: ICC by ethnicity for Item 085 from February 2024

By contrast, Figure 10 is the ICC for item 050. Comparing to Figure 9, this item is more difficult and less discriminating than item 085. The nature of the DIF is also distinct, with the relative performance of candidates in the Other ethnic category being both lower and higher than other candidates depending on overall ability (i.e. location on the horizontal scale).

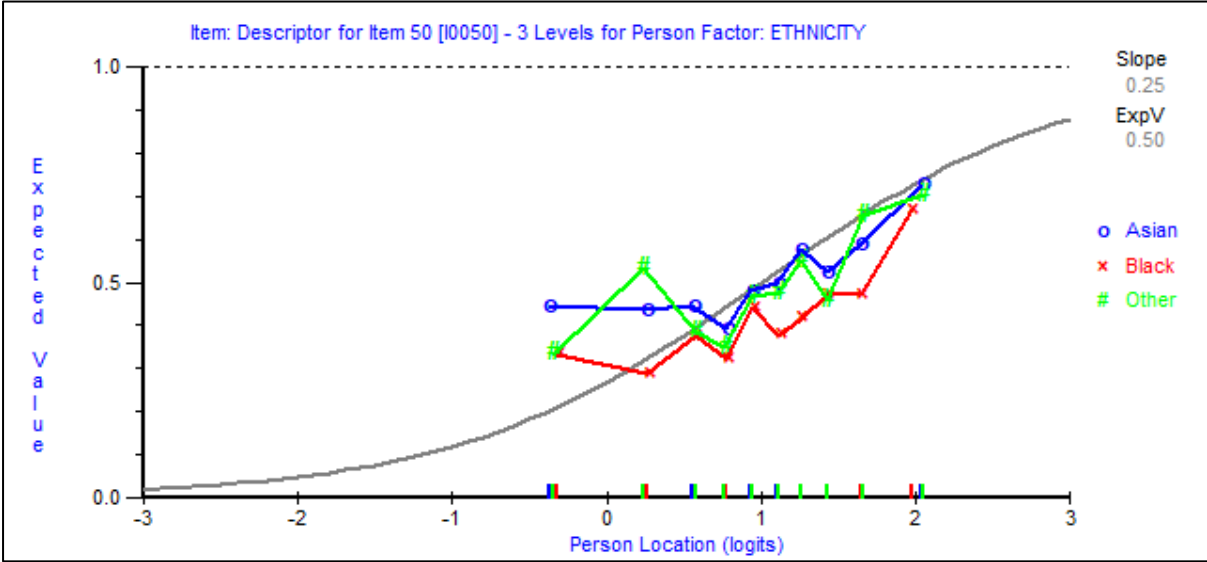
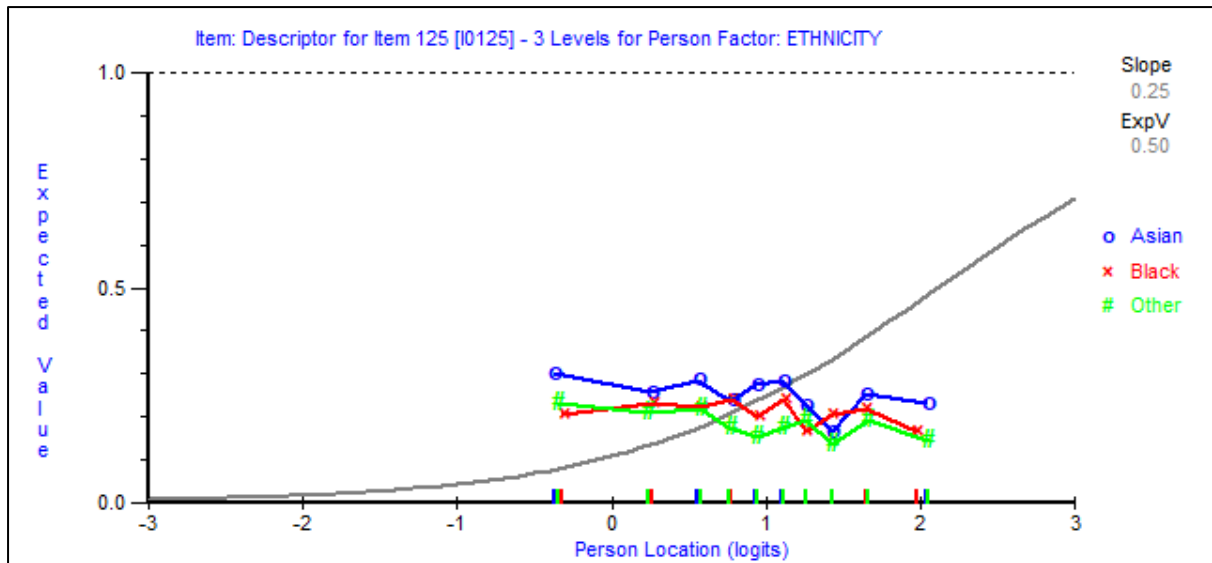


Figure 10: ICC by ethnicity for Item 050 from February 2024

As a final illustrative example, Figure 11 shows the ICC for item 125. This is a more difficult item with Asian candidates performing best across most of the ability range. However, the item itself is negatively discriminating – with candidates at lower levels of overall ability more likely to answer this item correctly compared to those at higher levels of ability. This pattern is largely the same across all three ethnic groups.



**Figure 11: ICC by ethnicity for Item 125 from February 2024**

The final sub-section of the Main findings summarises the DIF by *ethnicity* analysis, for the May 2024 PLAB1 exam.

#### **DIF analysis May 2024 by *ethnicity***

There are 40 items (22.5%) from the May 2024 exam identified with statistically significant DIF by *ethnicity*. In a similar way as for February's exam, the 20 items with the most DIF are listed in Table 6 – ordered approximately from most to least DIF, and with a 'best fit' qualitative description of the nature of the DIF.

The Appendix contains full details of statistical tests (Table 10), and all item characteristic curves (ICC) for these 20 items (Figure 16).

Again, and as expected, there are a range of DIF categorisations, that to an extent balance out across the table.

<b>Item</b>	<b>Nature of DIF</b>
<b>I0016</b>	Black group performing best across the ability range
<b>I0091</b>	Asian and Other groups performing best across the ability range
<b>I0063</b>	Other group performing best across most of the ability range
<b>I0008</b>	Black group performing best across most the ability range
<b>I0163</b>	Other group performing best across most of the ability range
<b>I0119</b>	Black group performing best across the ability range
<b>I0030</b>	Asian group performing best across most of the ability range
<b>I0109</b>	Black group performing best across most of the ability range
<b>I0094</b>	Black and Other group performing best across the ability range
<b>I0151</b>	Asian group performing best across most of the ability range
<b>I0087</b>	Asian and Other group performing best across the ability range
<b>I0031</b>	Other group performing best across the ability range
<b>I0055</b>	Black and Other group performing best across most of the ability range
<b>I0123</b>	Other group performing best across the ability range
<b>I0092</b>	Black group performing best across the ability range
<b>I0153</b>	Black group performing best across most of the ability range
<b>I0049</b>	Asian group performing best across some of the ability range
<b>I0079</b>	Black and Other group performing best across most of the ability range
<b>I0103</b>	Asian group performing best across some of the ability range
<b>I0078</b>	Other group performing best across the ability range

**Table 6: Items with DIF by ethnicity for May 2024**

As with *sex*, there are no particular issues worthy of exemplification beyond those already discussed in the previous *ethnicity* analysis for February.

Having completed the findings, the final section of the report is the Discussion and conclusion.

## Discussion and conclusion

This work investigates the prevalence of items that have differential item functioning in two 2024 administrations of PLAB1, first by *sex* and then by *ethnicity*. Overall, there is quite a high percentage of items that show such DIF – 5 to 10% of items for *sex*, and of the order of 25% for *ethnicity*. Given that descriptive analysis shows that overall performance in PLAB1 is more varied by *ethnicity* than it is by *sex*, it is perhaps not a surprise that there is a greater prevalence of DIF by the former than by the latter, although psychometrically this does not automatically follow.

It is important to emphasise that DIF at the item-level does matter. Ideally, a ‘fair’ exam should not show DIF at the item level, and the existence of items with DIF is a threat to the validity of PLAB1 and its outcomes (American Educational Research Association, 2014, chap. 3). However, further extensive evidence is needed to better understand the drivers of the differential functioning of the items – and evidence of DIF in particular items does not automatically imply bias for or against particular groups of candidates either at the item or exam level.

By definition, PLAB1 candidates are a varied global group of international medical graduates, and it could be that specific areas of applied medical knowledge are more ‘standard’ or more ‘specialised’ in particular contexts (either by *ethnicity* and/or by *sex*) – which in turn could lead to DIF. This kind of argument implies that evidence of DIF is not necessarily evidence of genuine bias on the part of PLAB1 item writers. Further, it does not automatically imply that the overall PLAB1 outcomes are biased against specific groups. Indeed, the overall methodology employed in this study cannot alone be used to say anything specifically about the wider validity of the overall PLAB1 outcomes. As already stated, the methods used to investigate DIF across the full set of items necessarily assume that aggregate DIF is zero overall across the whole exam.

An obvious next step in developing understanding of the potential issues revealed here will be to consider the individual items identified in this work as exhibiting DIF, and to review them in terms of what they are intending to assess. This can help to investigate if there are important content areas that are seemingly more difficult for some candidate groups compared to others, and to assess the extent to which this is as a result of the design/wording of the item, or perhaps more a reflection of differentiated levels of specific applied medical knowledge for particular groups of candidates relative to other groups. Subject matter experts and item writers can then use all of the available evidence to make judgements as to whether there is any evidence of genuine bias in some of these items. In turn, this work will help improvements in the ongoing quality of the PLAB1 exams and associated outcomes.

## References

- American Educational Research Association 2014. *Standards for Educational and Psychological Testing*. Washington, D.C: American Educational Research Association.
- Andrich, D., Sheridan, B. and Luo, G. 2012. RUMM2030: Rasch unidimensional models for measurement.
- Bond, T. and Fox, C.M. 2007. *Applying the Rasch Model: Fundamental Measurement in the Human Sciences, Second Edition* 2 edition. Mahwah, N.J: Routledge.
- Cook, D.A., Brydges, R., Ginsburg, S. and Hatala, R. 2015. A contemporary approach to validity arguments: a practical guide to Kane's framework. *Medical Education*. **49**(6), pp.560–575.
- Downing, S.M. 2003. Item response theory: applications of modern test theory in medical education. *Medical education*. **37**(8), pp.739–745.
- General Medical Council 2024. Tackling differential attainment. [Accessed 26 September 2024]. Available from: <https://www.gmc-uk.org/education/standards-guidance-and-curricula/guidance/tackling-differential-attainment>.
- Homer, M. 2023a. *Influences on scoring and grading in PLAB2: PLAB2 Annual Report 2022*. London: General Medical Council.
- Homer, M. 2023b. Towards a more nuanced conceptualisation of differential examiner stringency in OSCEs. *Advances in Health Sciences Education*.
- IBM Corp 2021. IBM SPSS Statistics for Windows, Version 28.0.
- Kakara Anderson, H.L., Govaerts, M., Abdulla, L., Balmer, D.F., Busari, J.O. and West, D.C. 2024. Clarifying and expanding equity in assessment by considering three orientations: Fairness, inclusion and justice. *Medical Education*. **n/a**(n/a).
- Müller, M. 2020. Item fit statistics for Rasch analysis: can we trust them? *Journal of Statistical Distributions and Applications*. **7**(1), p.5.
- Osterlind, S.J. and Everson, H.T. 2009. *Differential Item Functioning*. SAGE Publications.
- Panayides, P., Robinson, C. and Tymms, P. 2010. The assessment revolution that has passed England by: Rasch measurement. *British Educational Research Journal*. **36**(4), pp.611–626.
- Tennant, A. and Pallant, J. 2007. DIF matters: A practical approach to test if Differential Item Functioning makes a difference. [Accessed 26 September 2024]. Available from: <https://www.rasch.org/rmt/rmt204d.htm>.

## Appendix

The statistical significance level for all hypothesis tests in this report is Bonferroni adjusted at  $0.05/534=0.000094$  with those items reaching this threshold included and highlighted in the relevant tables<sup>4</sup>. This corresponds to the methodology used in the RUMM2030 software to correct for the issue of multiple hypothesis testing.

The central columns in the following tables (labelled 'DIF main') show the main DIF analysis results, and each table is ordered by the 'prob' column within this (smallest to largest – i.e. starting with the item with the most statistically significant DIF).

The columns to the left of each table ('Rasch item fit') indicate tests of overall fit to the Rasch model for each item, and there are many items that show poor fit – although the prevalence of such items has not been investigated systematically in this report. However, often misfit is a result of over or under-discrimination, and does not automatically undermine the DIF analysis – the graphical evidence in the item characteristic curves confirms the existence of the DIF regardless of overall Rasch fit.

The column to the right ('DIF interaction') indicates the extent to which DIF between groups might be different at different levels of ability for each item. However, this only meets the level of statistical significance for a single item/analysis in this report<sup>5</sup> and is not considered any further.

Below each table the corresponding item characteristic curves are shown in the same order as in the table (i.e. starting with the most statistically significant DIF).

### DIF analysis February 2024 by sex full details

Table 7 and Figure 12 give the full tabular and graphical results of the DIF analysis by sex for February 2024.

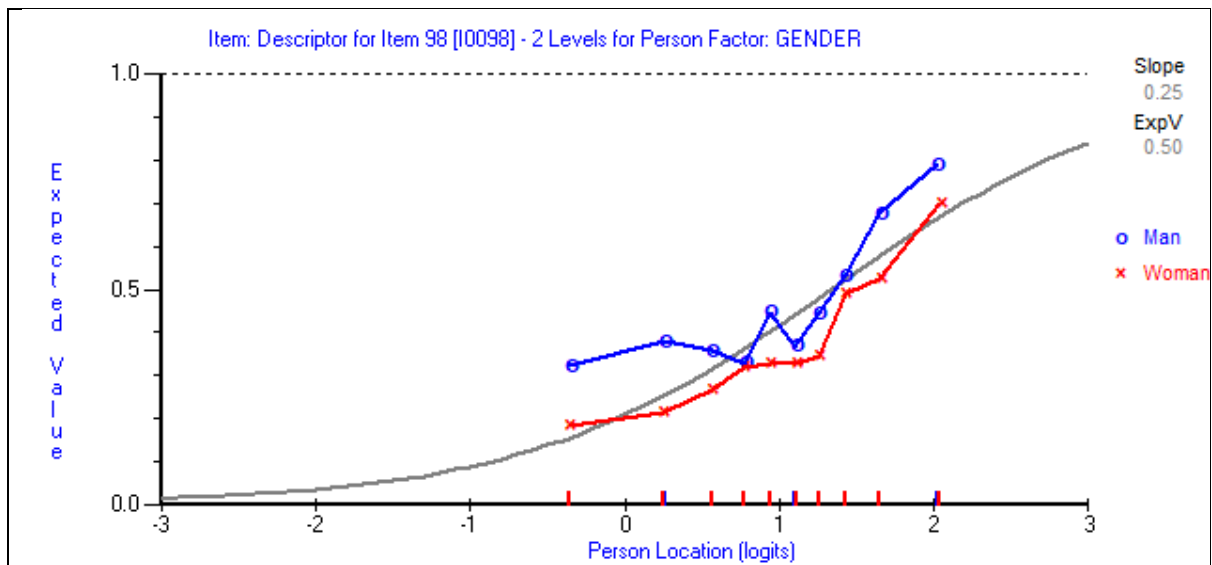
---

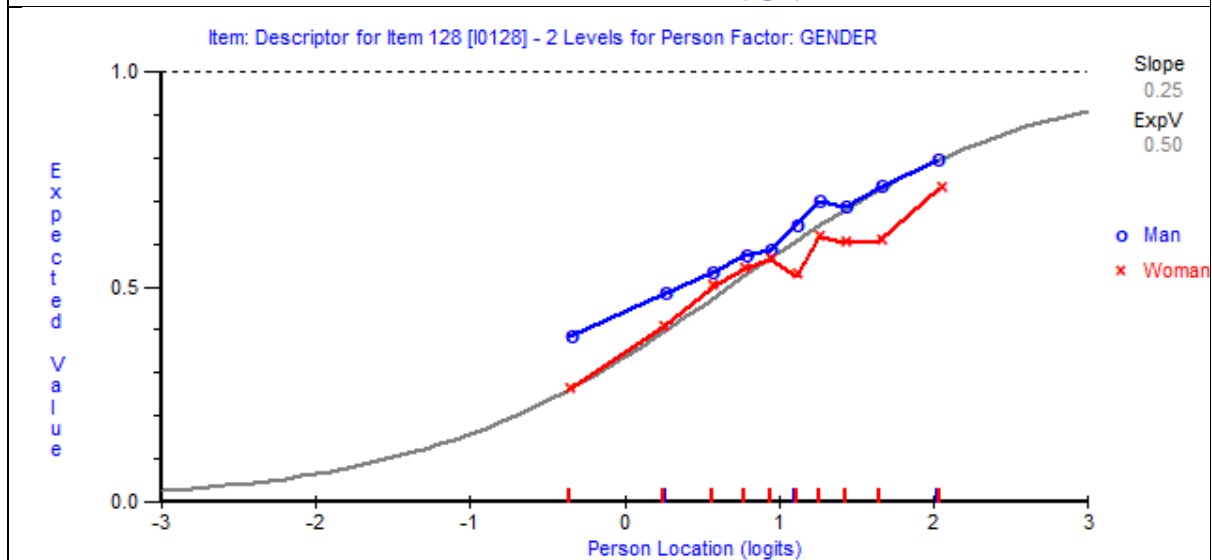
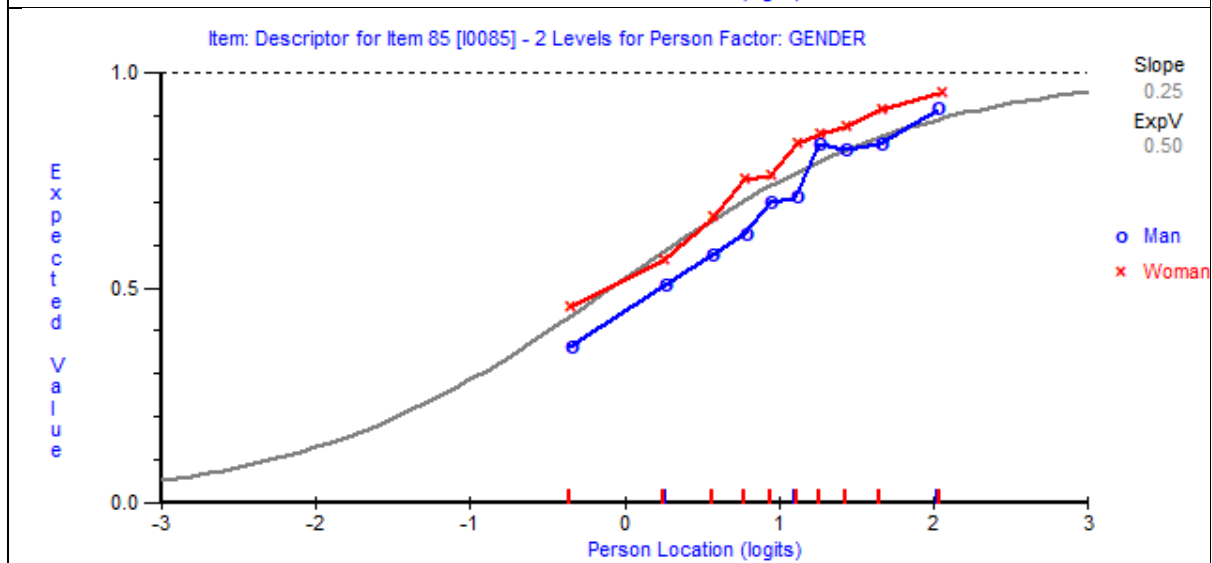
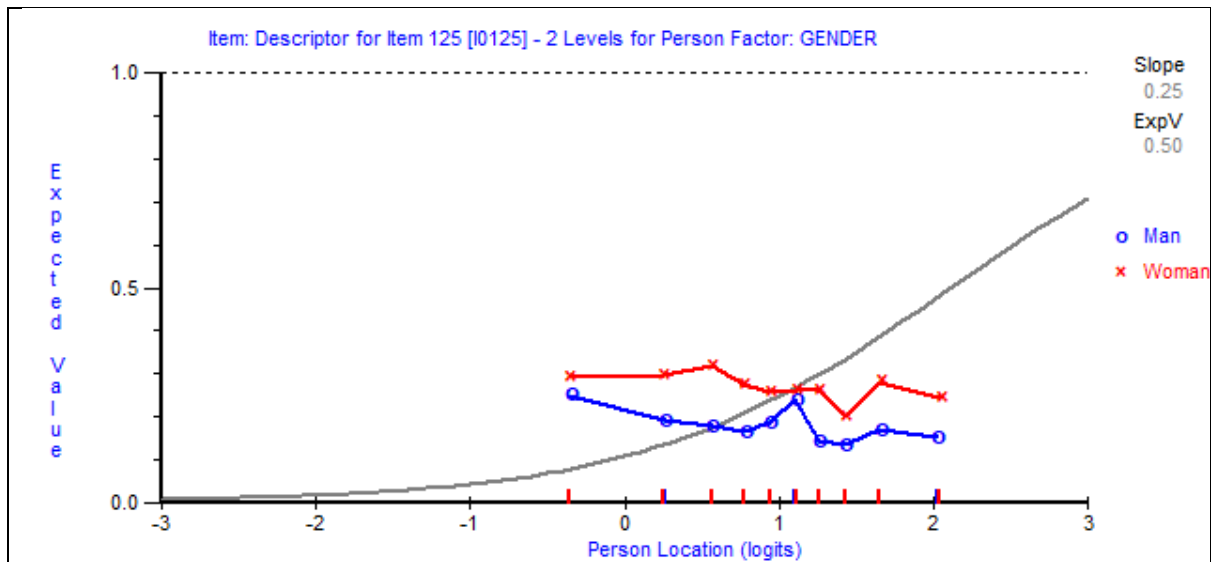
<sup>4</sup> There are  $3 \times 178 = 534$  separate significance tests in each DIF analysis.

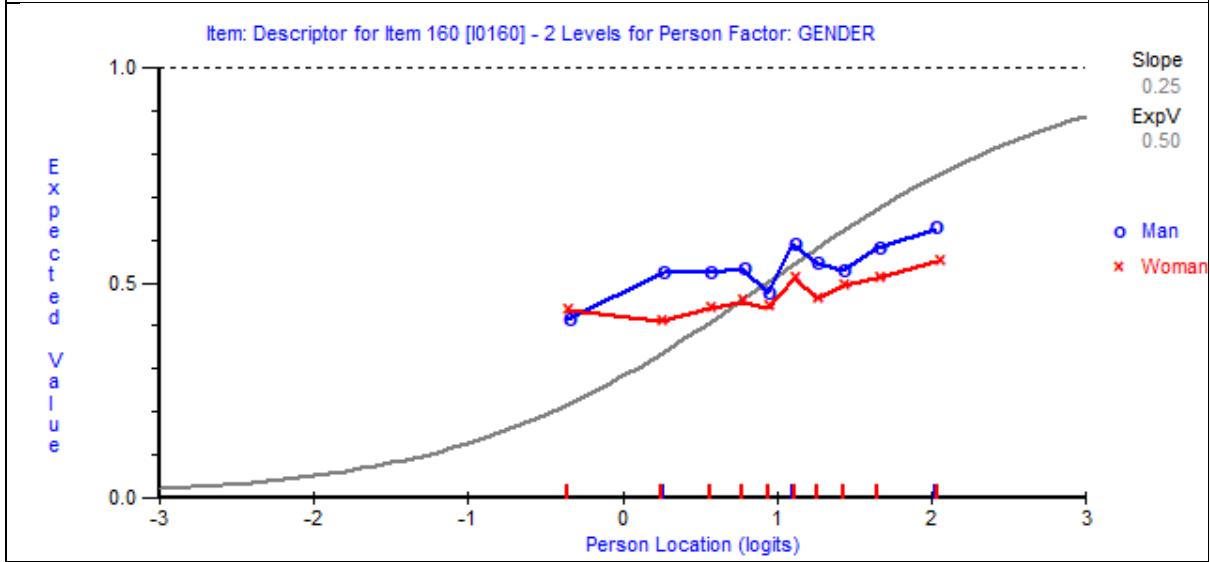
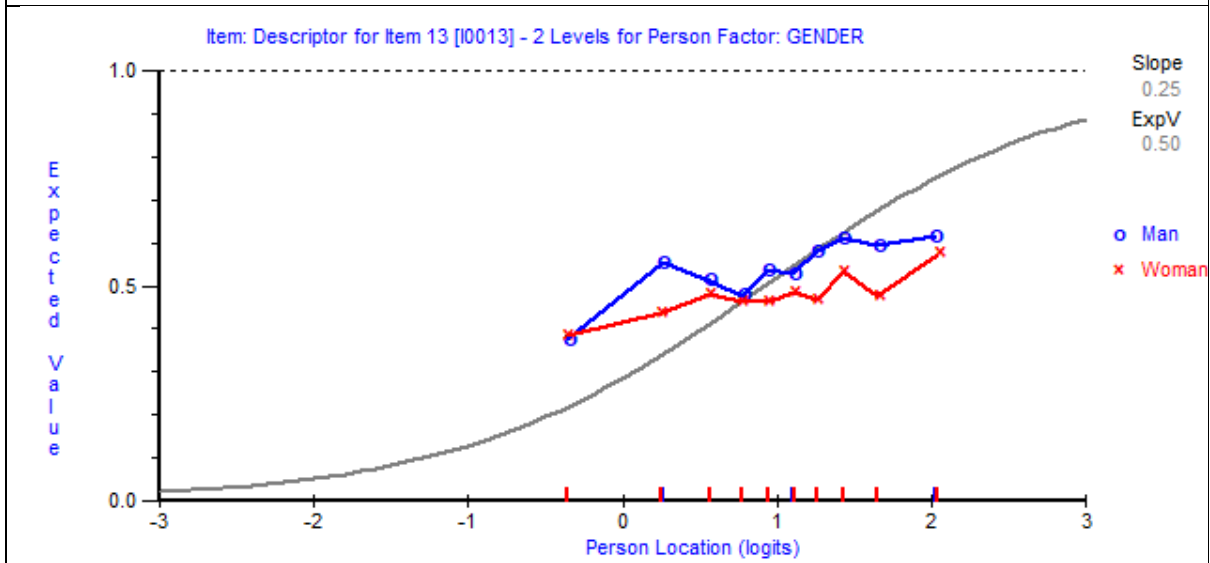
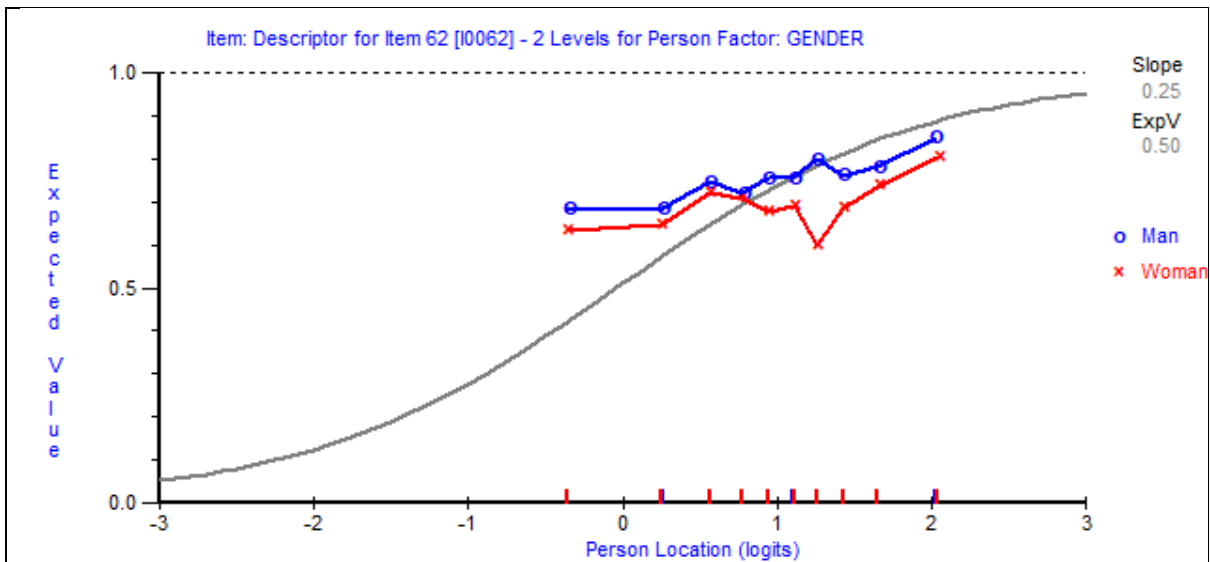
<sup>5</sup> DIF by ethnicity for item 132 from May 2024 (Table 9)

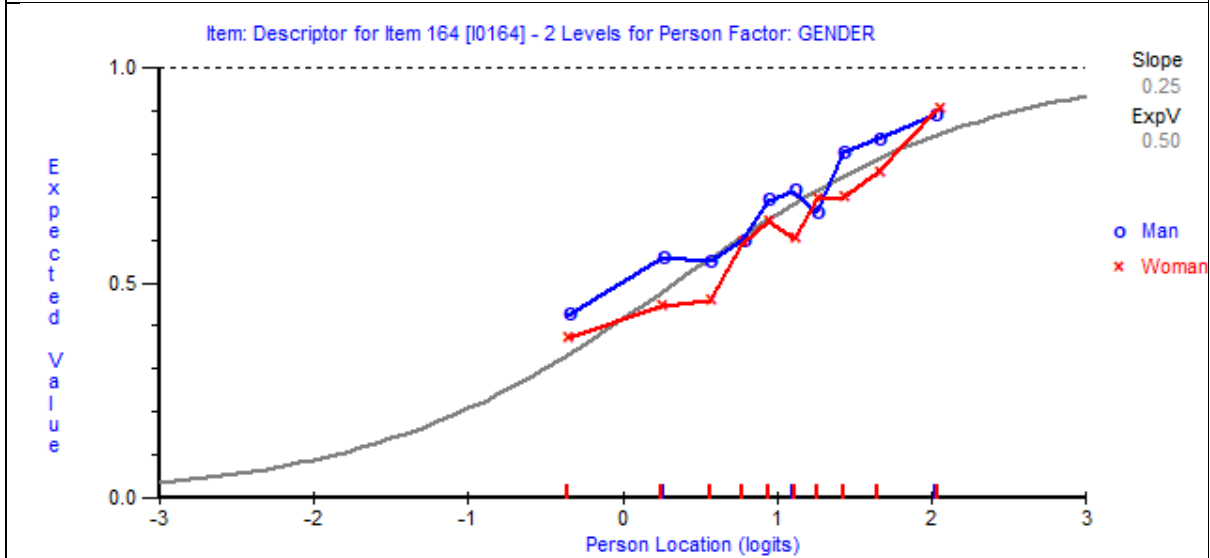
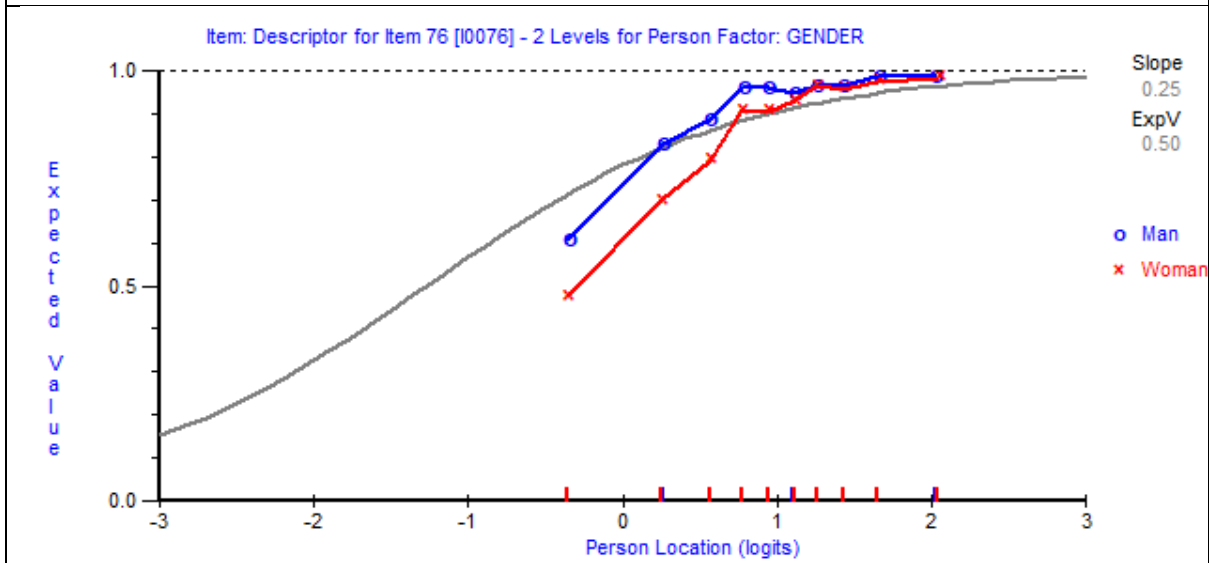
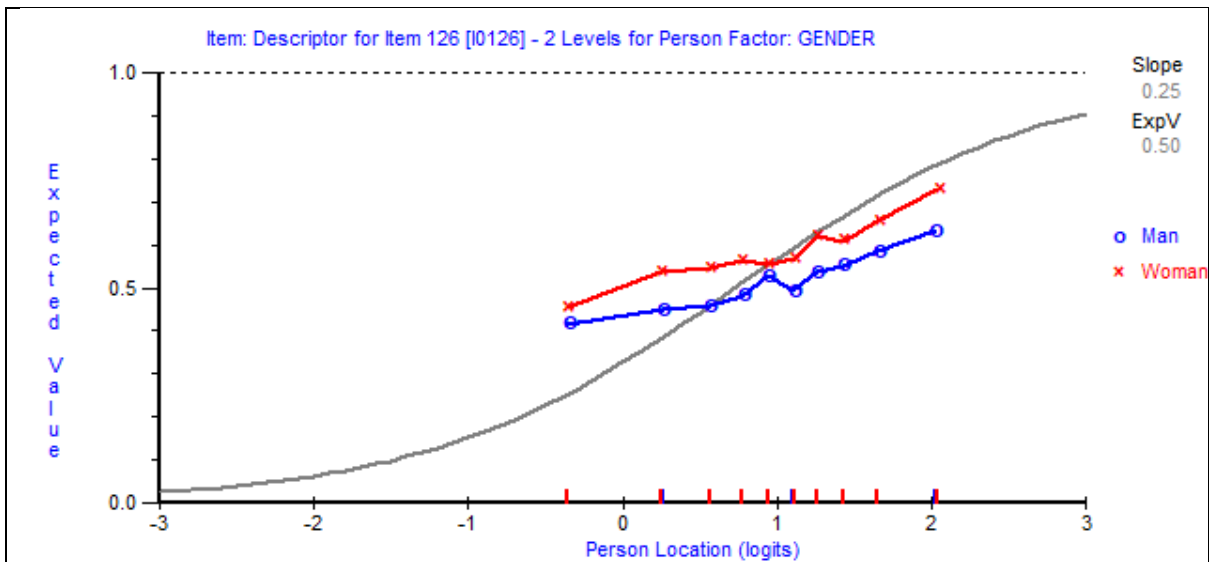
Rasch item fit					DIF main				DIF interaction			
Item	MS	F	DF	Prob	MS	F	DF	Prob	MS	F	DF	Prob
I0098	10.0354	10.0462	9	0.000000	59.70933	59.77361	1	0.000000	1.64796	1.649740	9	0.095589
I0125	75.29086	65.53257	9	0.000104	48.47754	42.19446	1	0.000000	2.12786	1.852070	9	0.054466
I0085	4.094	4.56647	9	0.000049	40.56728	45.24903	1	0.000000	0.49473	0.551830	9	0.837138
I0128	4.61524	4.41276	9	0.000000	38.11903	36.4467	1	0.000000	0.69389	0.663440	9	0.742760
I0062	30.06705	26.56036	9	0.000127	34.20169	30.21278	1	0.000000	1.63625	1.445410	9	0.162555
I0013	33.88225	30.48369	9	0.000041	28.75762	25.87309	1	0.000000	0.36665	0.329870	9	0.965469
I0160	39.63407	35.45509	9	0.000000	26.13569	23.37996	1	0.000000	0.41603	0.372160	9	0.948767
I0126	22.67219	20.63342	9	0.000000	24.75251	22.52667	1	0.000000	0.81311	0.739990	9	0.672469
I0076	15.38673	20.93237	9	0.000000	21.7391	29.57424	1	0.000000	2.3863	3.246360	9	0.000559
I0164	3.90988	4.00749	9	0.000000	18.58175	19.04566	1	0.000000	1.79604	1.840880	9	0.056154
I0172	3.02835	3.15793	9	0.000744	18.55976	19.35392	1	0.000000	1.13752	1.186190	9	0.298858
I0034	27.38729	24.76847	9	0.000105	17.36234	15.70212	1	0.000000	2.32361	2.101430	9	0.026123
I0072	3.80094	3.7266	9	0.000211	16.08736	15.77271	1	0.000000	0.73758	0.723160	9	0.688103
I0152	18.48081	26.76005	9	0.000000	11.29929	16.36127	1	0.000000	1.08657	1.573340	9	0.117045
I0103	8.55719	9.41407	9	0.000097	13.96775	15.36643	1	0.000030	0.36298	0.399330	9	0.935994
I0047	3.84285	3.66478	9	0.000033	22.68778	21.63649	1	0.000039	1.00105	0.954660	9	0.475913
I0001	5.29133	5.6941	9	0.000000	18.47726	19.88374	1	0.000040	0.98796	1.06317	9	0.386773
I0180	6.55974	7.29169	9	0.000000	14.70968	16.35104	1	0.000044	1.23267	1.370210	9	0.195594
I0063	5.44085	5.47649	9	0.000000	28.60882	28.7962	1	0.000063	0.6665	0.670860	9	0.736082
I0130	8.29911	10.19826	9	0.000134	12.60361	15.48779	1	0.000071	1.0988	1.350250	9	0.205181

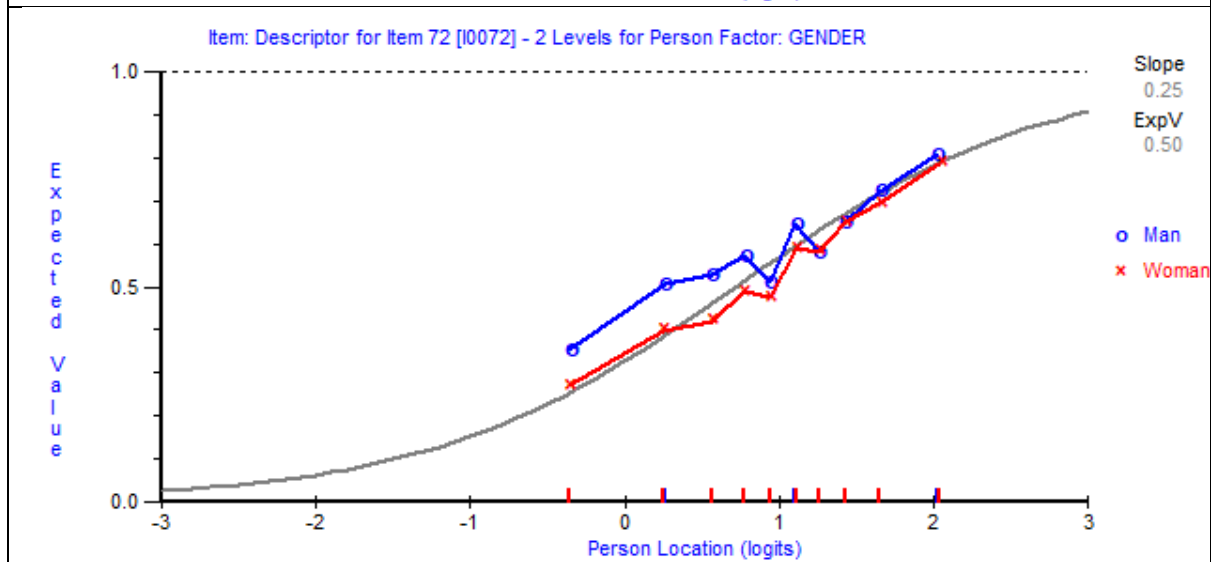
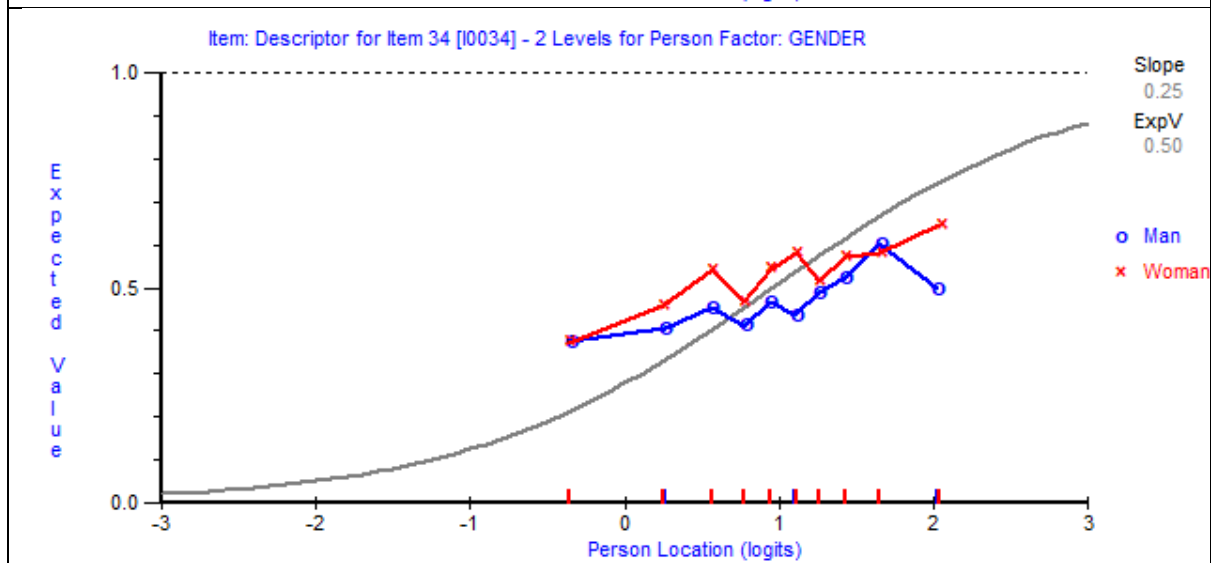
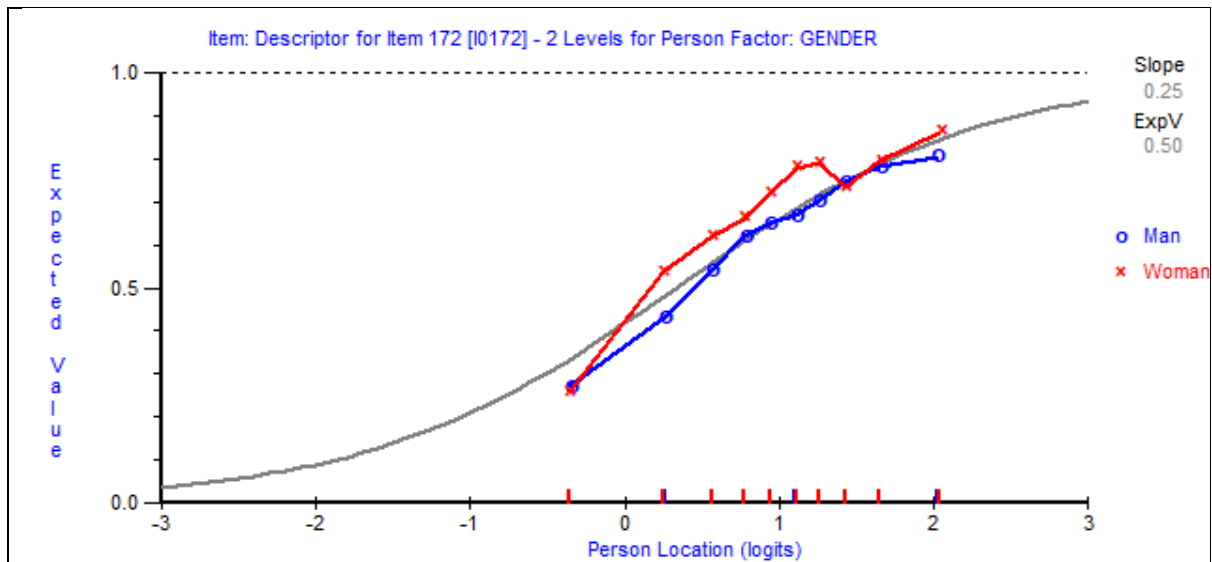
Table 7: Summary DIF data by sex for February 2024

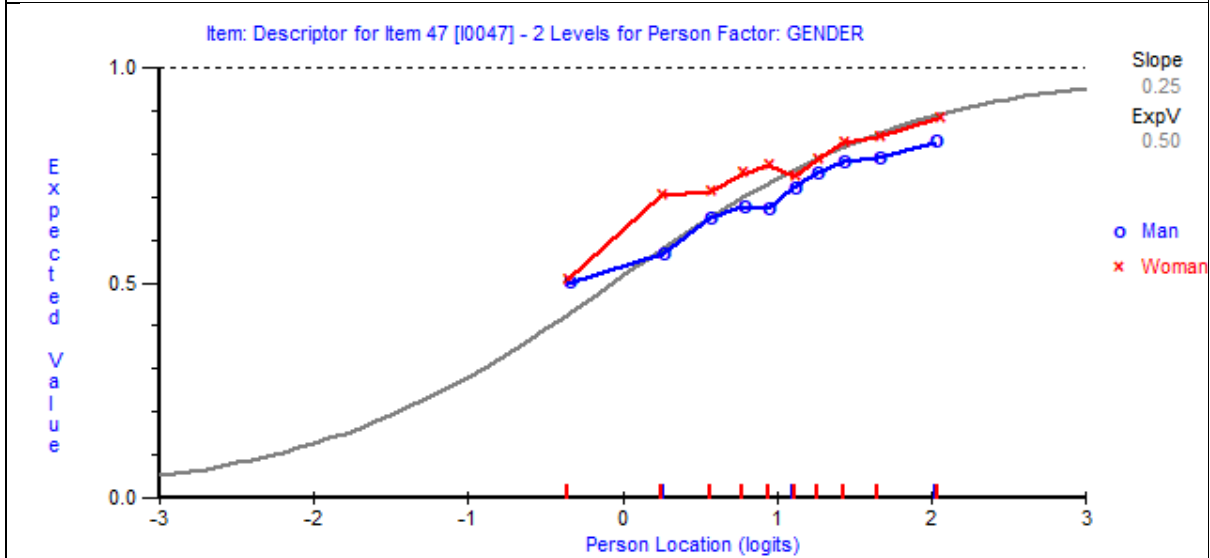
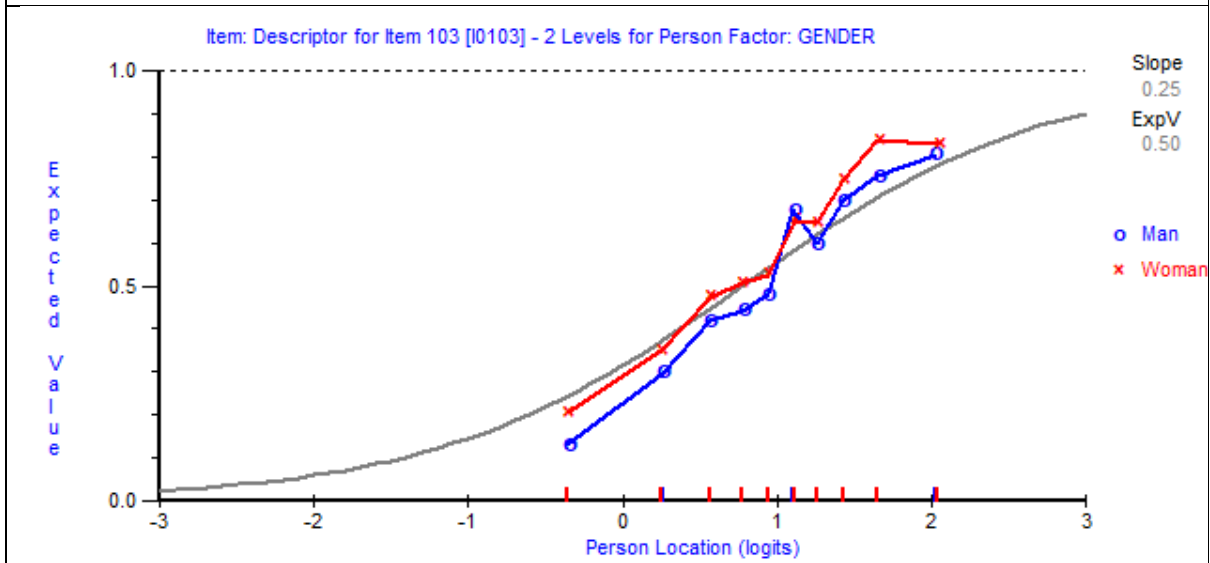
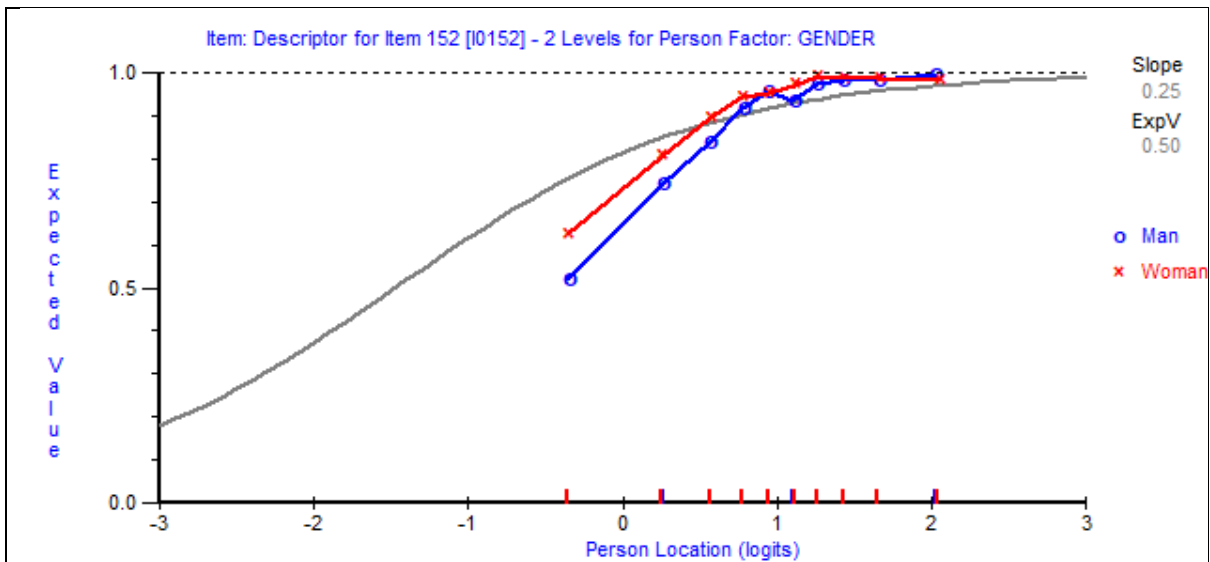


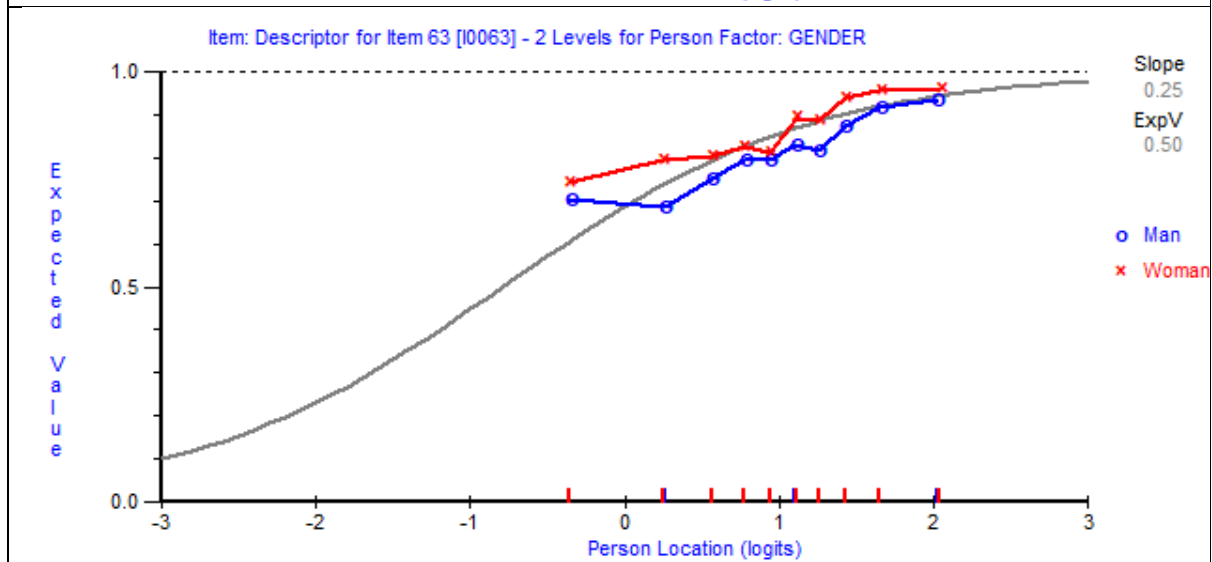
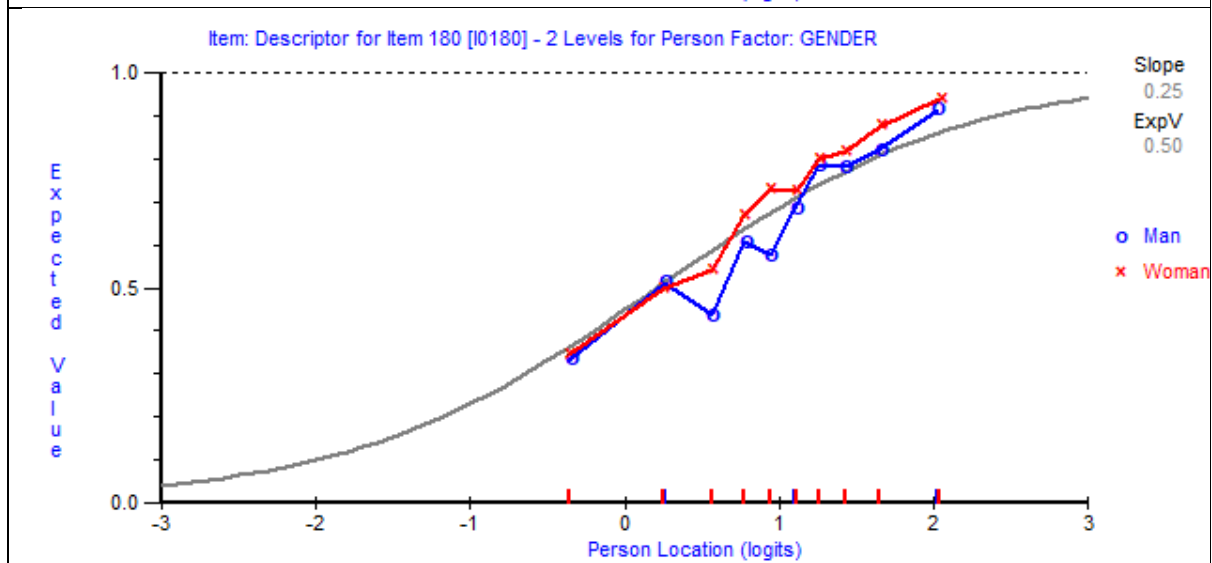
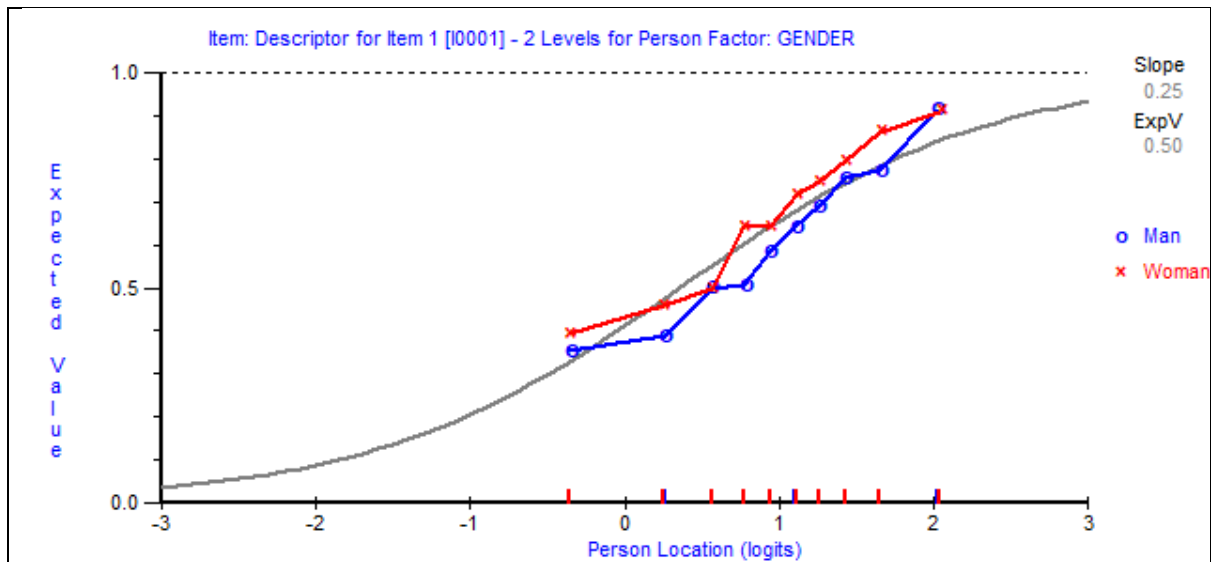












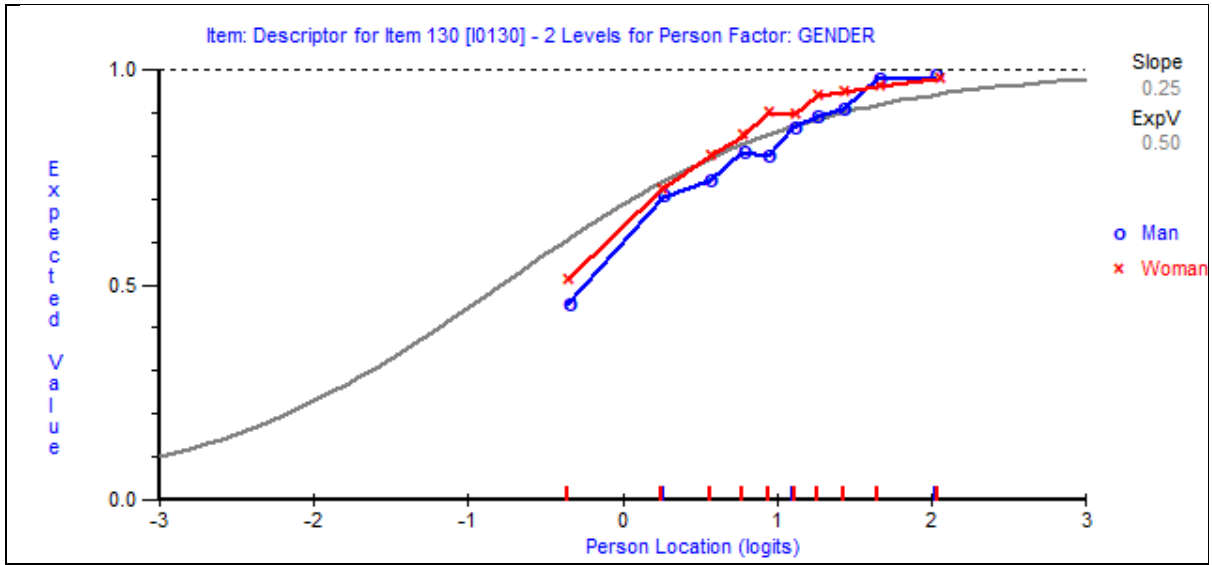


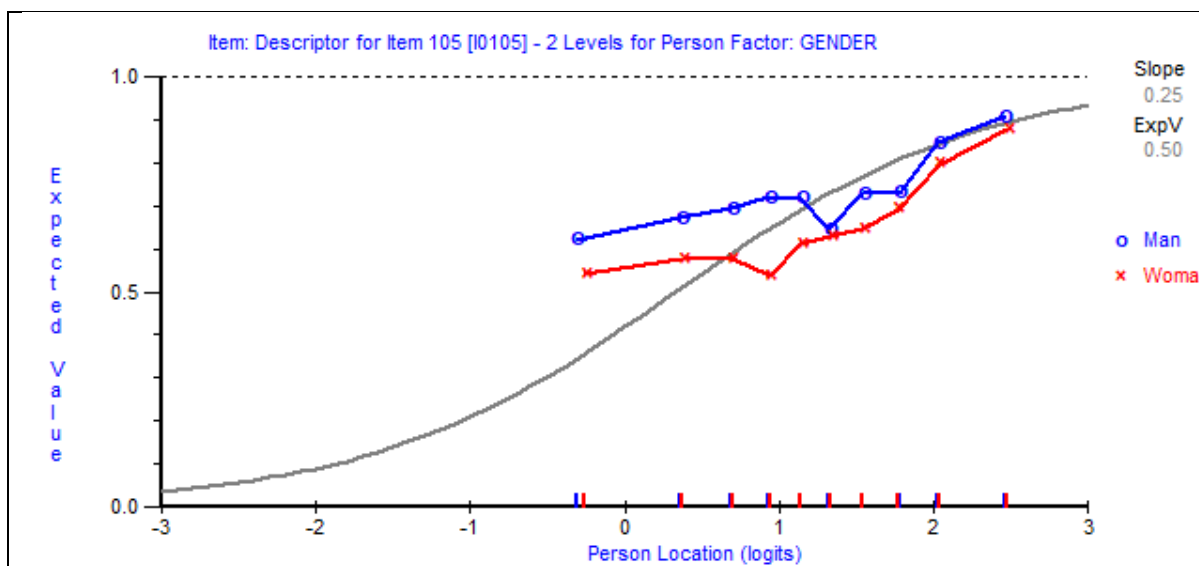
Figure 12: ICCs by sex for February 2024

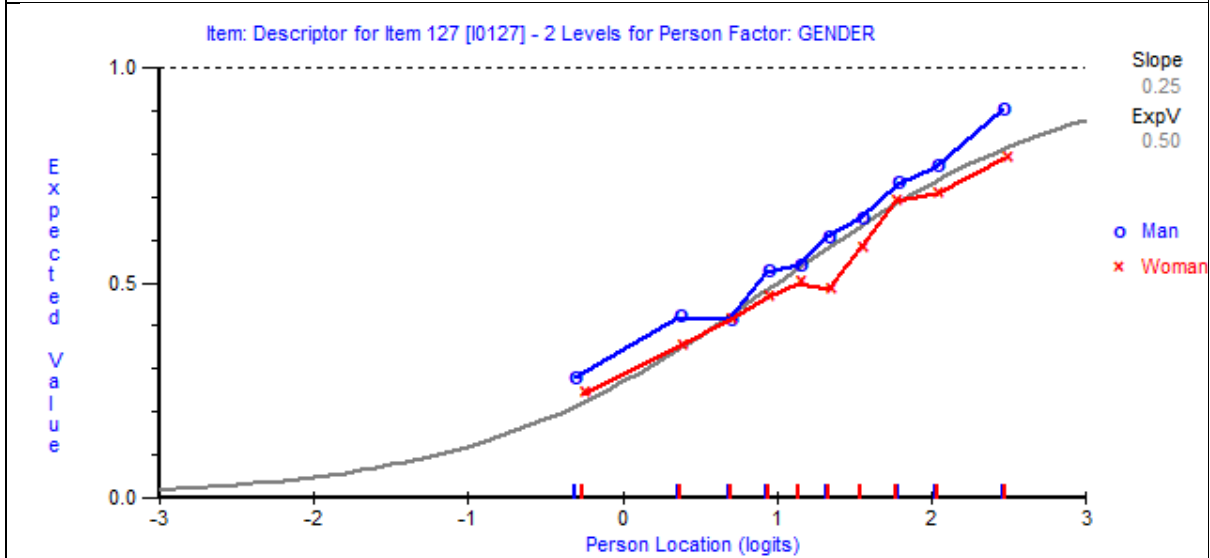
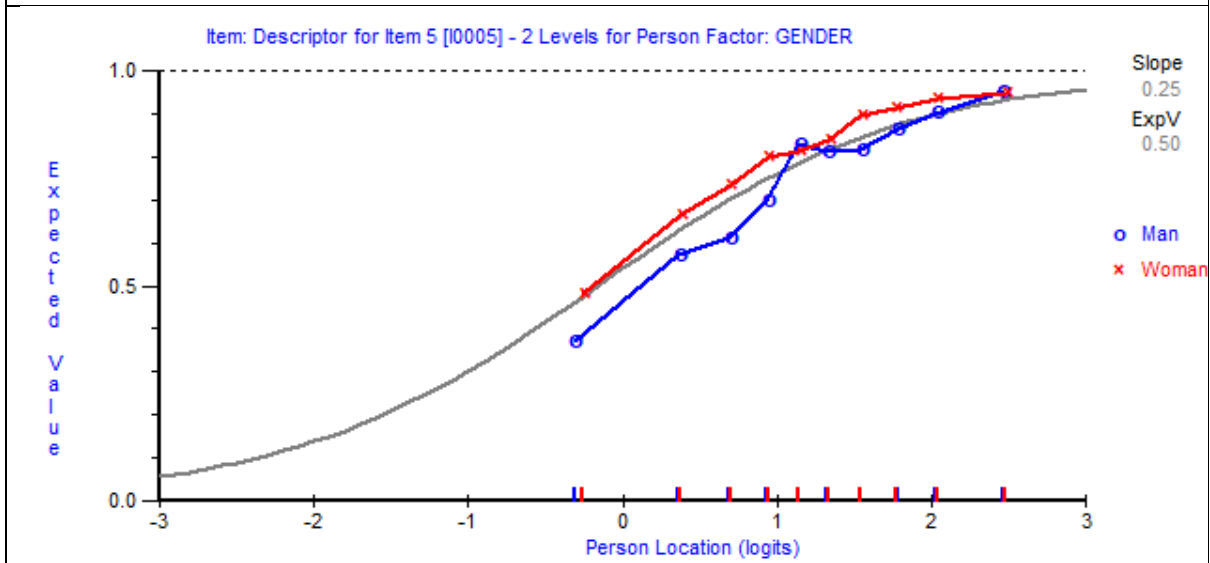
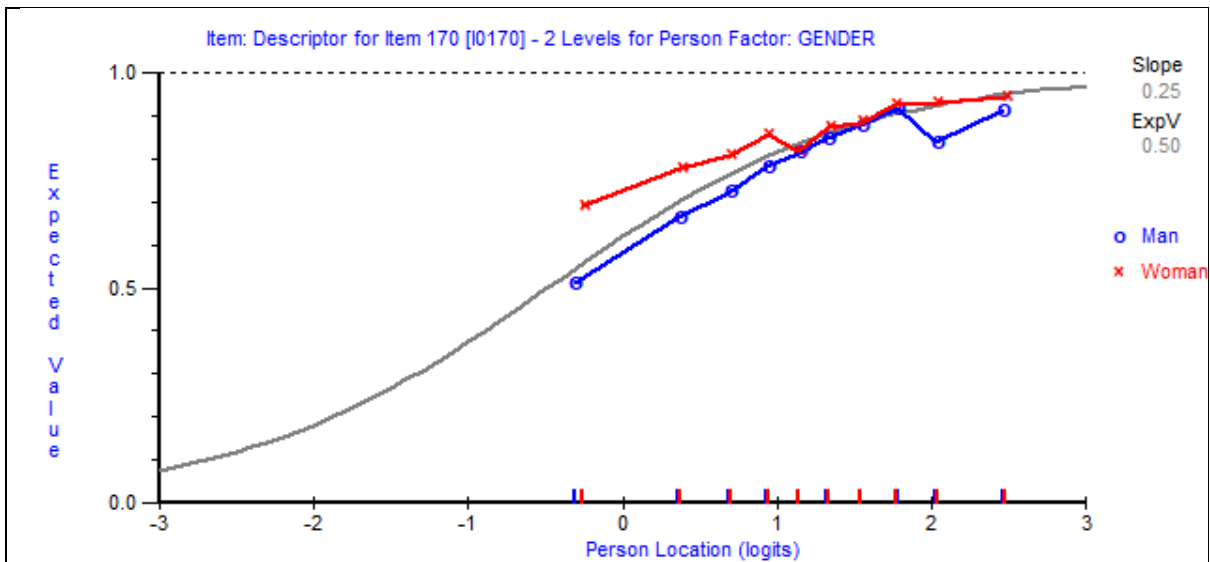
### DIF analysis May 2024 by sex full details

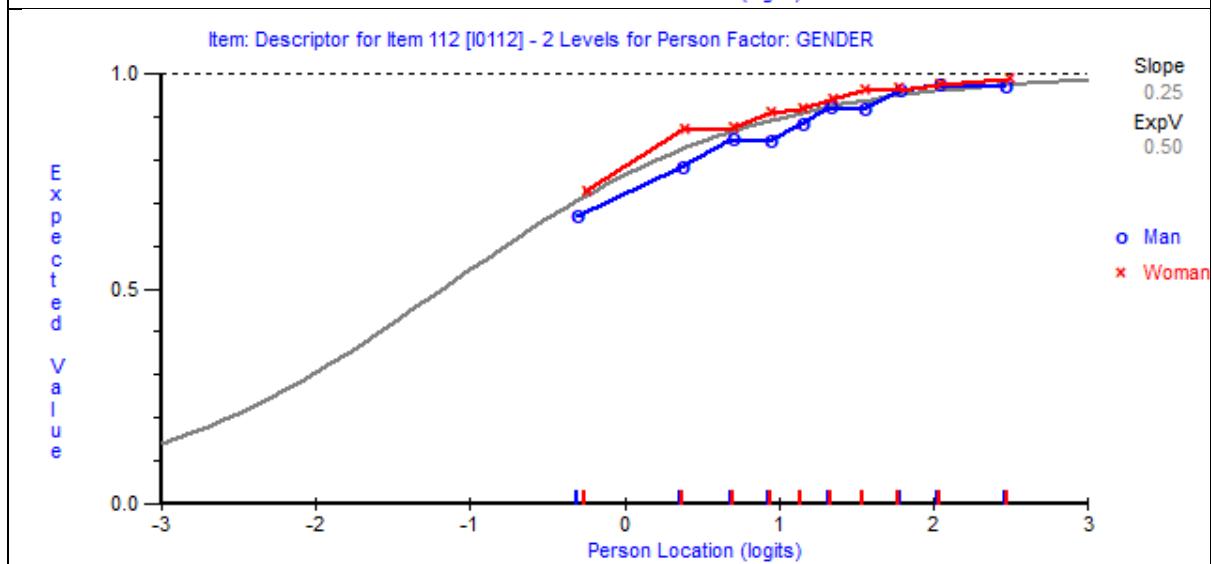
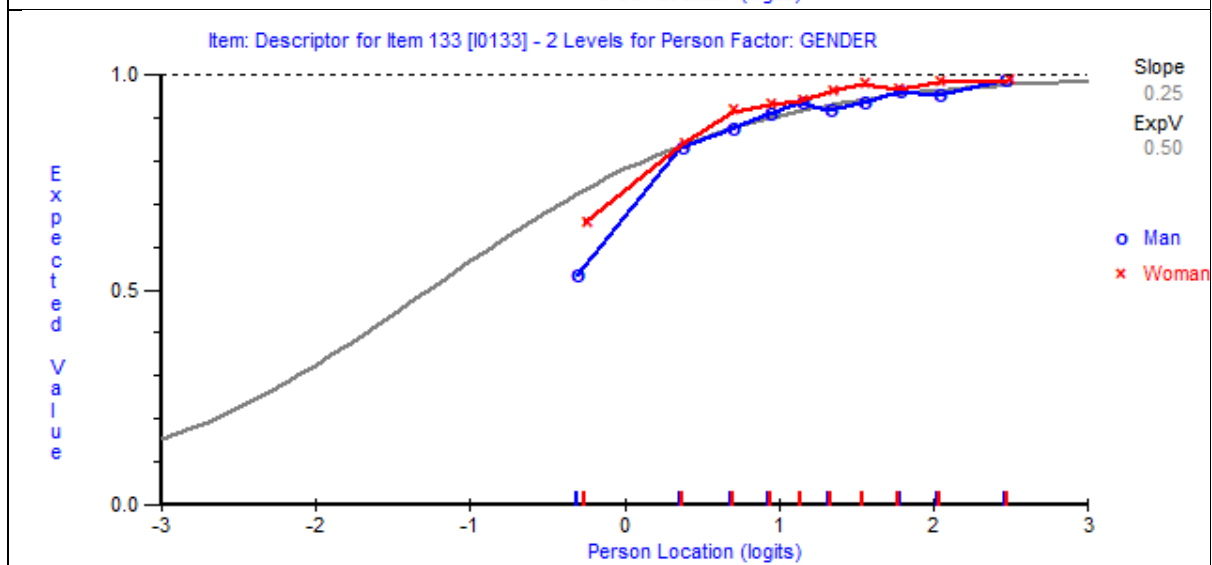
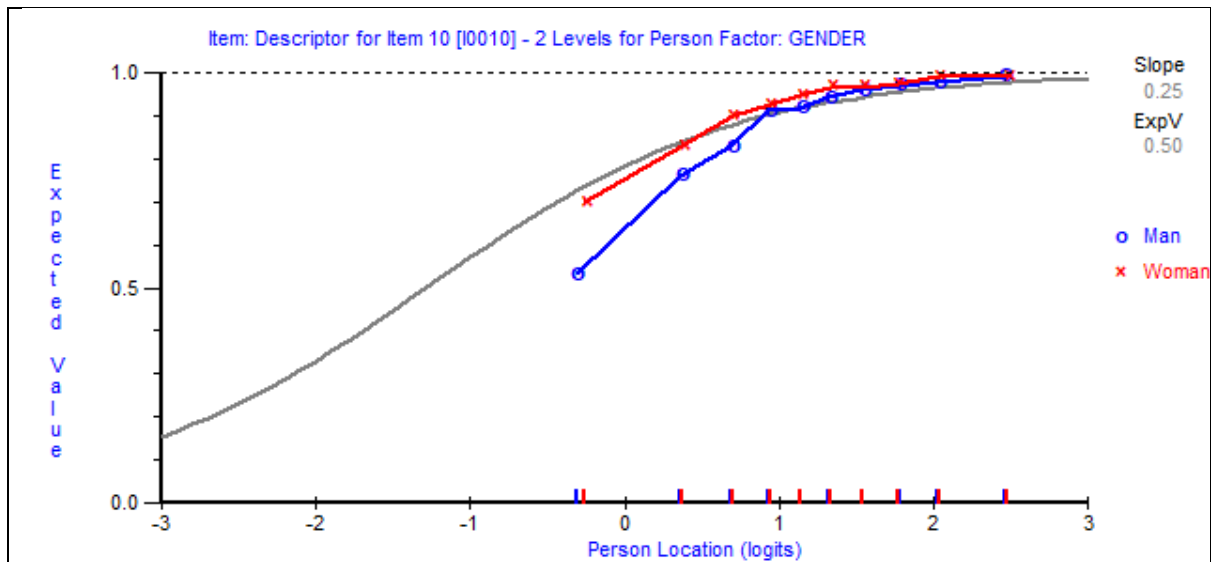
Table 8 and Figure 13 give the full tabular and graphical results of the DIF analysis by sex for May 2024.

Item	Rasch item fit				DIF main				DIF interaction			
	MS	F	DF	Prob	MS	F	DF	Prob	MS	F	DF	Prob
I0105	25.82642	23.72145	9	0.000013	45.09094	41.41582	1	0.000000	1.07667	0.98891	9	0.446666
I0170	2.78234	2.6845	9	0.004113	30.97088	29.88184	1	0.000000	2.23198	2.1535	9	0.022295
I0005	1.68762	1.85125	9	0.054622	25.38131	27.84225	1	0.000000	1.29409	1.41957	9	0.173327
I0127	1.66043	1.6572	9	0.093601	23.13973	23.0947	1	0.000000	0.82598	0.82437	9	0.593548
I0010	6.93313	9.10344	9	0.000082	16.86154	22.13978	1	0.000000	1.44952	1.90327	9	0.046959
I0133	5.89796	7.14902	9	0.000092	15.22472	18.45416	1	0.000000	0.99094	1.20113	9	0.289262
I0112	0.6764	0.74224	9	0.670349	15.04076	16.50468	1	0.000003	0.79744	0.87505	9	0.546800
I0078	25.63297	22.36226	9	0.000126	18.75362	16.36069	1	0.000014	1.13846	0.99319	9	0.443103
I0033	3.15789	3.05604	9	0.001192	21.2483	20.56297	1	0.000024	1.09657	1.0612	9	0.388248
I0143	1.2381	1.32322	9	0.218832	19.86749	21.23335	1	0.000075	1.0119	1.08146	9	0.372731

**Table 8: Summary DIF data by sex for May 2024**







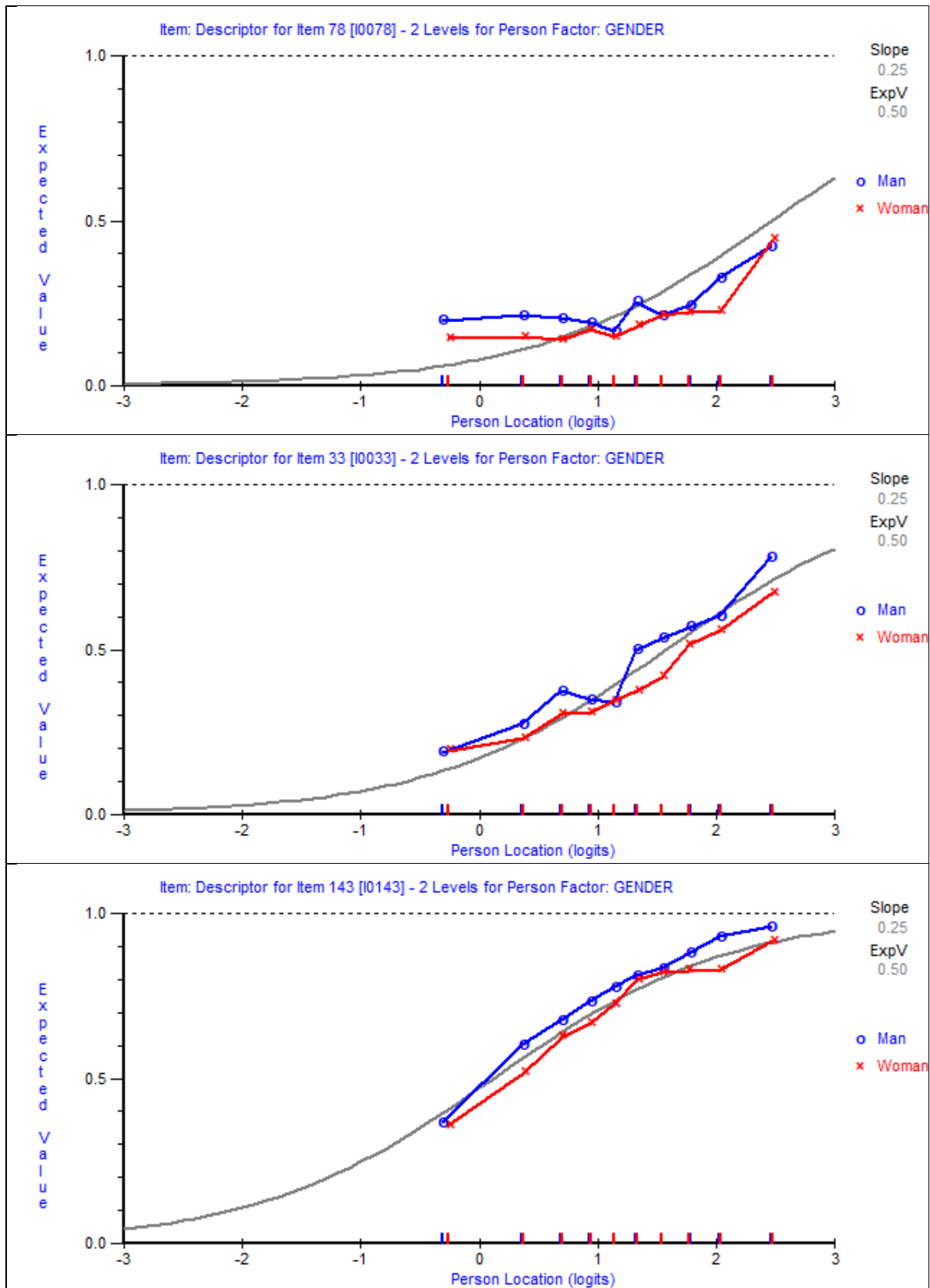


Figure 13: ICCs by sex for May 2024

### The collapsing of ethnicity into three categories

Figure 14 compares two DIF analyses by *ethnicity* for the same item (007 from February 2024). This is one of the items from that PLAB1 exam that shows statistically significant DIF by *ethnicity*.

In the upper graph, with ethnicity in six categories, there is considerable variability in the patterns of DIF estimates for some of these ethnic groups. However, once the four smaller categories are merged into one, the variability is largely smoothed out and the DIF analysis becomes more easily interpretable and meaningful.

Notice that the red and blue lines are exactly the same in both graphs – these two larger groups of candidates have not changed and so the estimations of likelihood of answering the item correctly are exactly the same across both analyses.

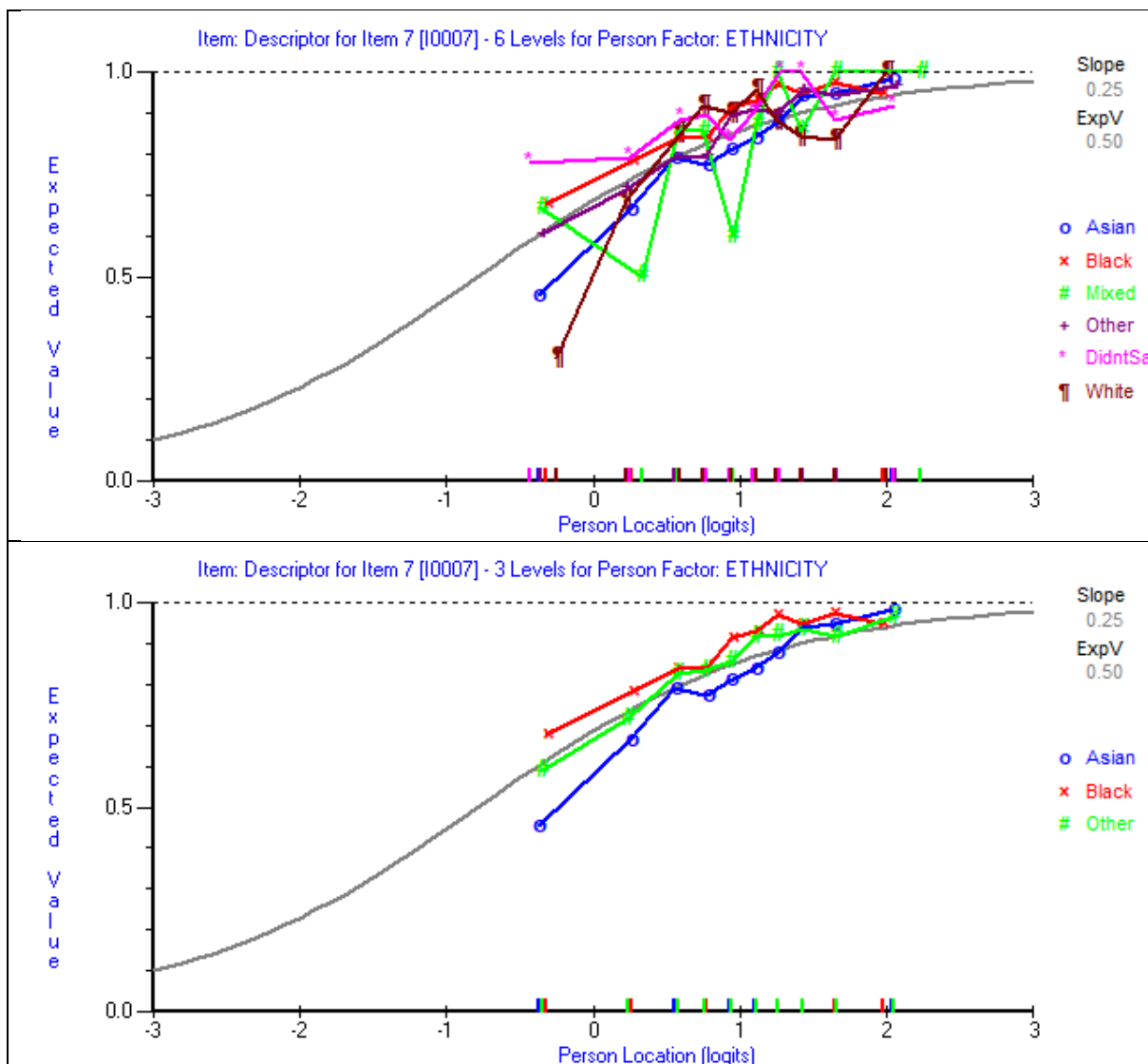


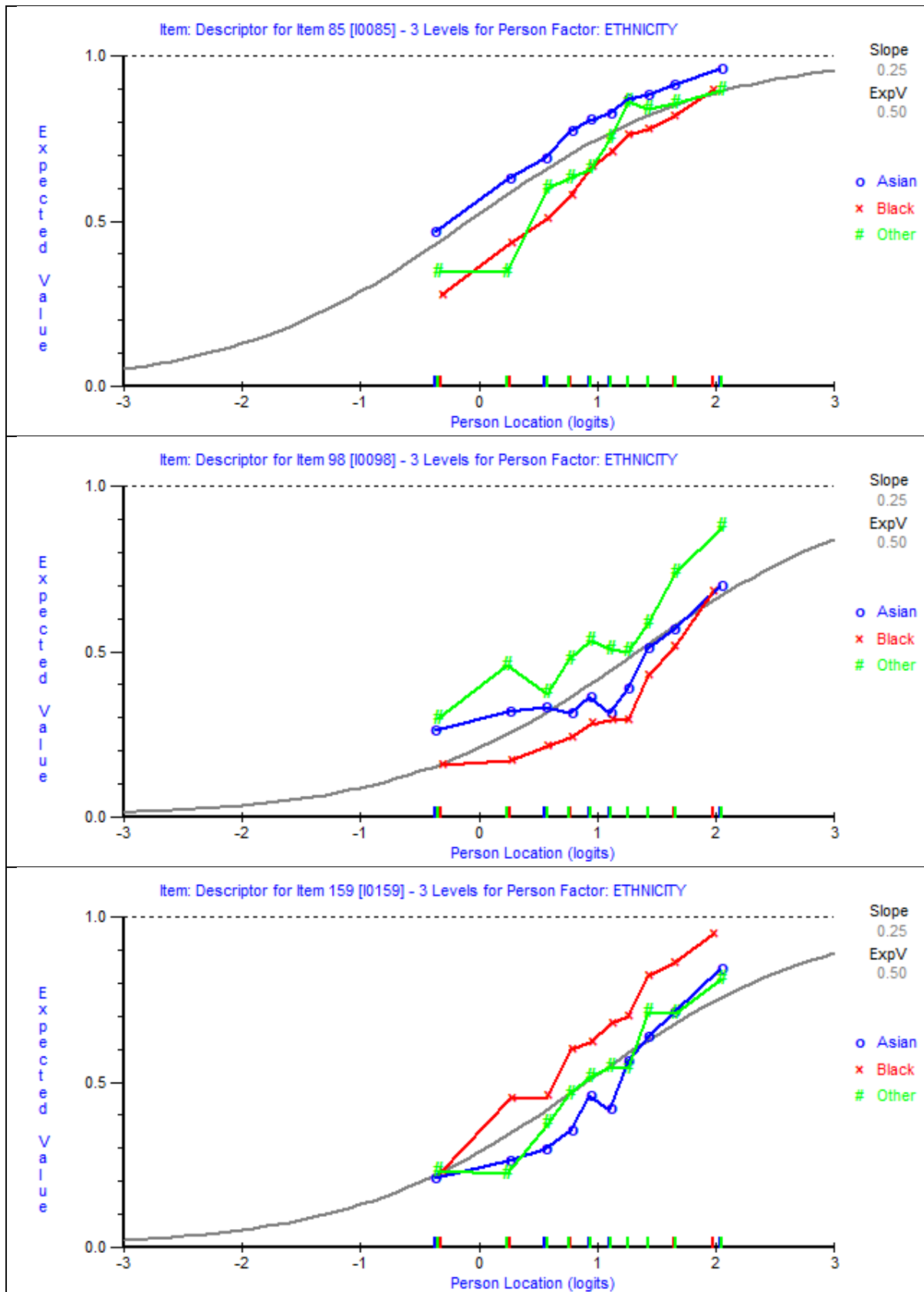
Figure 14: ICCs by ethnicity for Item 007 from February 2024

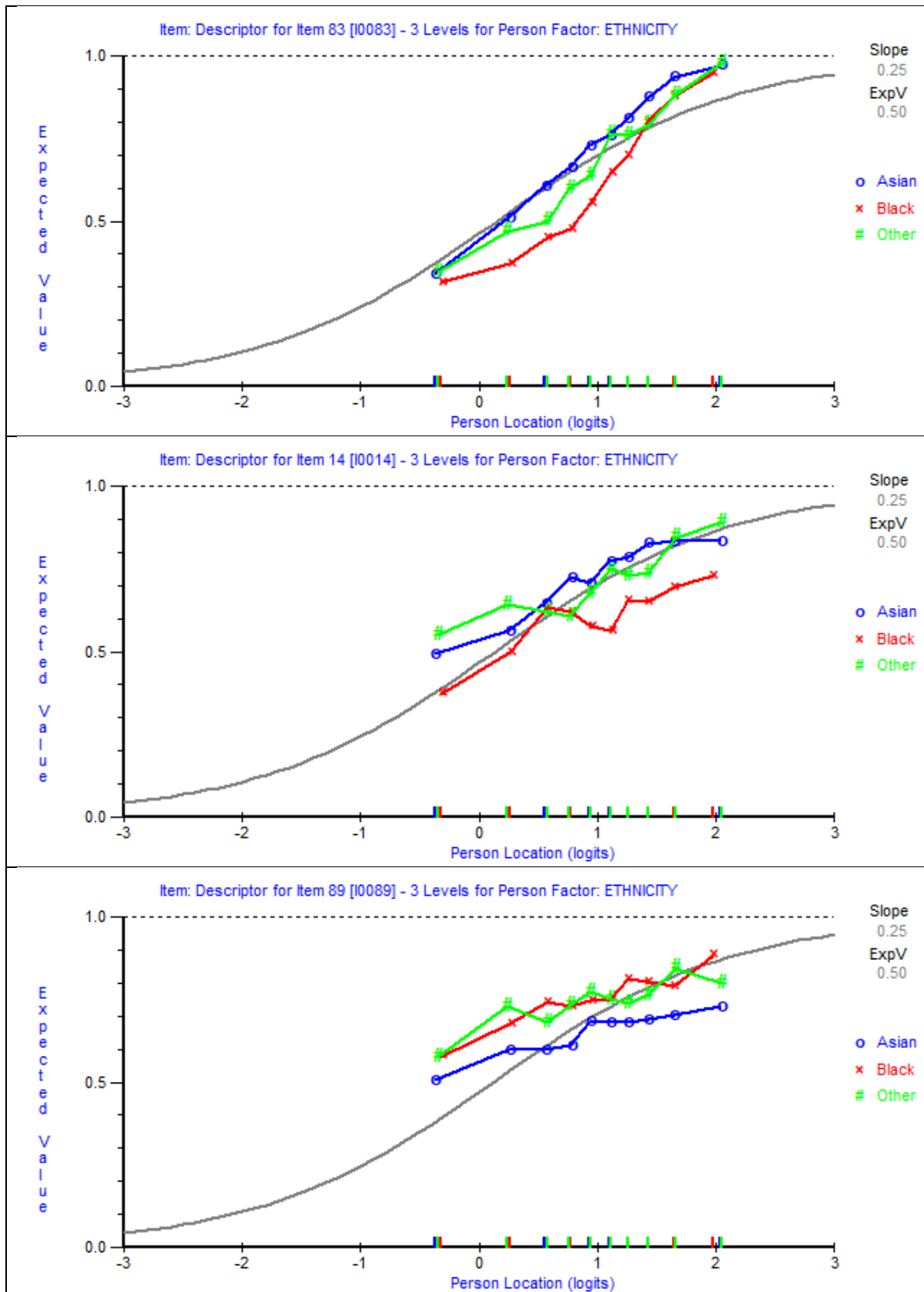
### **DIF analysis February 2024 by *ethnicity* full details**

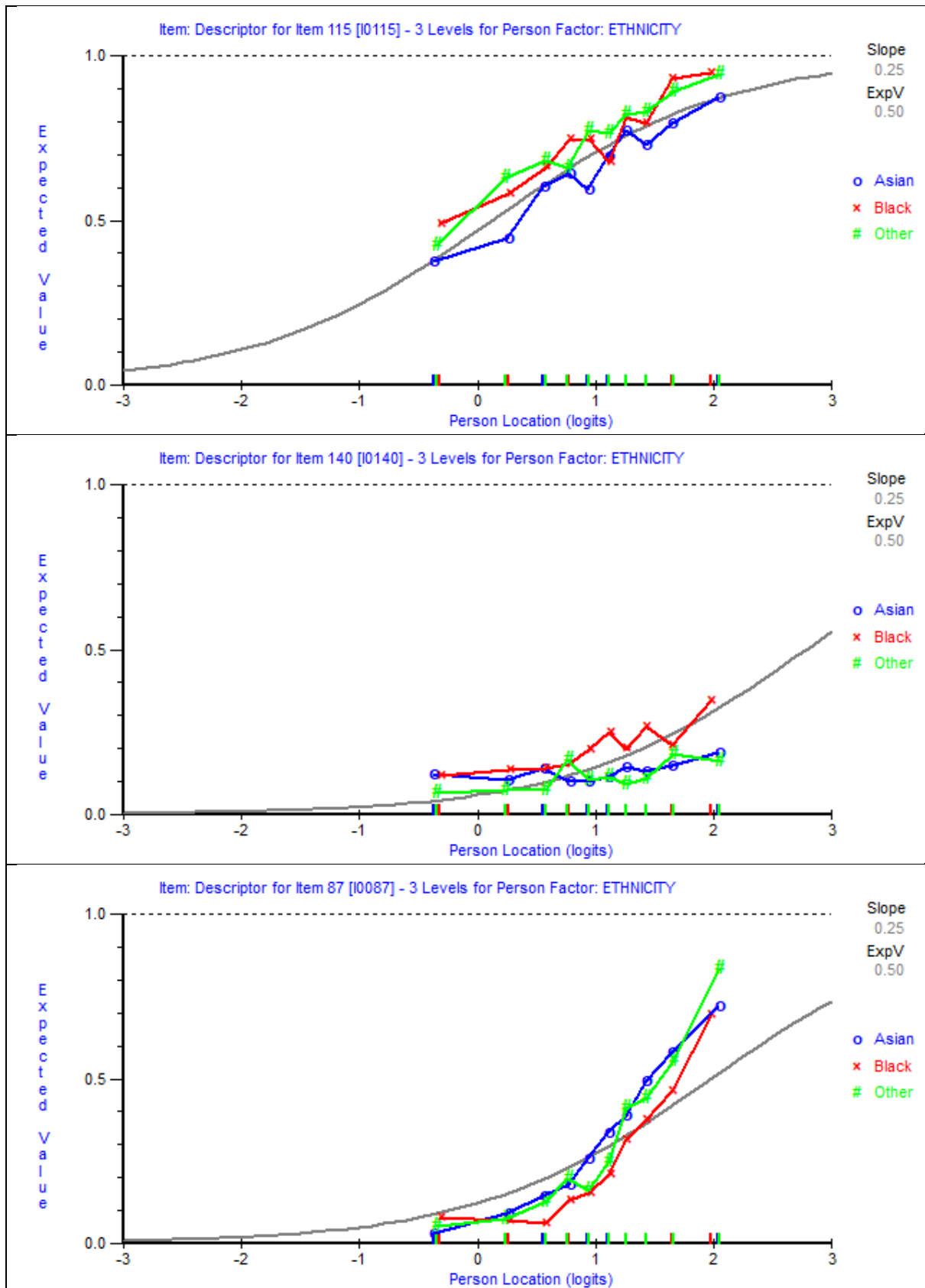
Table 9 and Figure 15 give the tabular and graphical results of the DIF analysis by *ethnicity* for February 2024. The table contains full details for all items with statistically significant DIF but given their prevalence (n=45) I have only included associated graphs for the 'worst' 20 items in this regard. Beyond this level of DIF, the graphs become more difficult to meaningfully interpret.

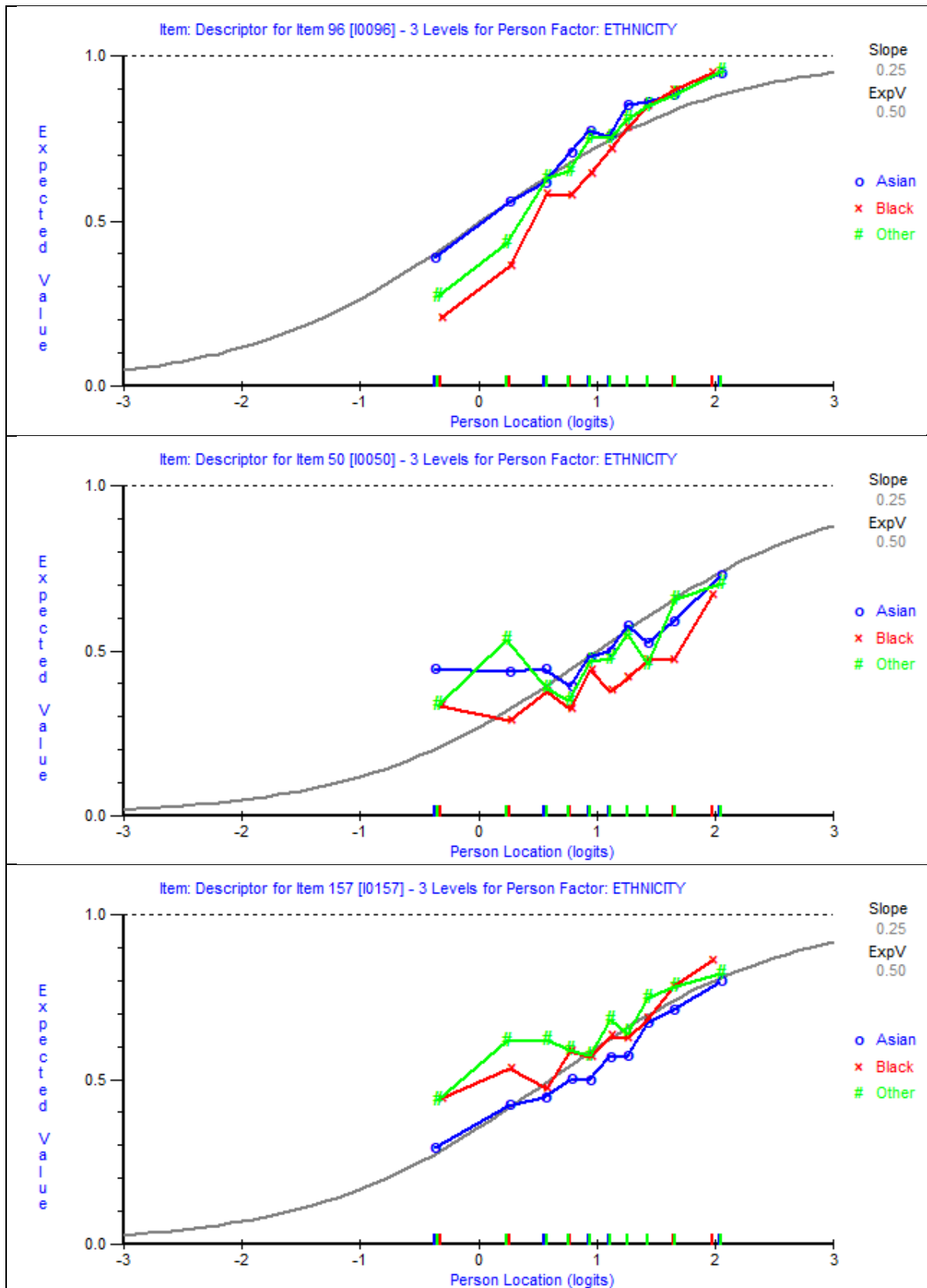
Rasch item fit					DIF main				DIF interaction			
Item	MS	F	DF	Prob	MS	F	DF	Prob	MS	F	DF	Prob
I0085	4.094	4.63592	9	0.000000	56.25943	63.70647	2	0.000000	0.61365	0.69487	18	0.819721
I0098	10.0354	10.13234	9	0.000091	56.7351	57.28315	2	0.000000	0.83639	0.84446	18	0.648090
I0159	7.34926	8.06509	9	0.000067	39.9222	43.8107	2	0.000000	2.00956	2.20529	18	0.002395
I0083	13.87122	16.73528	9	0.000056	32.04988	38.66737	2	0.000000	0.19095	0.23038	18	0.999690
I0014	3.89839	3.7685	9	0.000017	30.83729	29.80986	2	0.000000	1.82737	1.76648	18	0.023483
I0089	19.20297	17.20395	9	0.000000	33.15911	29.70726	2	0.000000	0.4048	0.36266	18	0.993504
I0115	1.90885	1.9912	9	0.036252	23.11008	24.10707	2	0.000000	1.1631	1.21327	18	0.239810
I0140	15.7881	14.14356	9	0.000000	23.08373	20.67926	2	0.000000	1.02562	0.91879	18	0.555106
I0087	28.9441	35.06345	9	0.000000	15.62449	18.92782	2	0.000000	0.58031	0.703	18	0.811470
I0096	8.35304	9.74462	9	0.000000	16.03857	18.71054	2	0.000000	0.99763	1.16384	18	0.282482
I0050	23.21191	21.7054	9	0.000000	19.35596	18.09971	2	0.000000	0.9814	0.9177	18	0.556442
I0157	3.88511	3.8064	9	0.000047	17.68113	17.32293	2	0.000000	0.94387	0.92475	18	0.547662
I0101	4.98534	6.00216	9	0.000000	13.99213	16.84597	2	0.000000	0.72066	0.86765	18	0.619162
I0007	5.27997	6.20421	9	0.000120	13.65418	16.04429	2	0.000000	1.43639	1.68783	18	0.034249
I0158	8.75135	9.64018	9	0.000000	13.99918	15.42101	2	0.000000	1.23703	1.36267	18	0.139050
I0009	1.90014	1.92425	9	0.044266	14.71572	14.90248	2	0.000000	1.59276	1.61298	18	0.048472
I0064	8.74404	8.18243	9	0.000002	15.18855	14.21302	2	0.000000	0.89564	0.83812	18	0.655961
I0099	5.25887	5.02837	9	0.000121	14.81736	14.16791	2	0.000000	1.29314	1.23646	18	0.221398
I0125	75.29087	65.0948	9	0.000000	16.37275	14.15551	2	0.000000	0.35217	0.30447	18	0.997875
I0162	5.54807	6.31961	9	0.000120	12.23829	13.9402	2	0.000000	0.52402	0.5969	18	0.904679
I0063	5.44085	5.46799	9	0.000051	13.44793	13.51499	2	0.000000	0.53564	0.53832	18	0.941388
I0016	19.50613	26.99827	9	0.000108	9.52152	13.17865	2	0.000000	0.8804	1.21855	18	0.235465
I0081	2.44325	2.80986	9	0.002787	10.67185	12.27314	2	0.000000	1.0232	1.17673	18	0.270873
I0133	6.74086	8.01983	9	0.000000	10.09769	12.01357	2	0.000000	1.02296	1.21705	18	0.236692
I0149	16.47785	22.1455	9	0.000000	8.6677	11.649	2	0.000000	1.24981	1.67969	18	0.035614
I0170	13.54013	12.3527	9	0.000104	12.36316	11.27894	2	0.000000	1.1899	1.08555	18	0.359718
I0019	5.21502	5.90113	9	0.000011	9.8909	11.19219	2	0.000000	0.99666	1.12779	18	0.316651
I0106	11.38008	13.78938	9	0.000000	8.87149	10.7497	2	0.000000	0.93532	1.13334	18	0.311219
I0079	5.70295	6.05506	9	0.000110	9.88106	10.49113	2	0.000000	1.73557	1.84273	18	0.016186
I0130	8.2991	10.19251	9	0.000000	8.53521	10.48248	2	0.000000	0.62128	0.76303	18	0.746028
I0048	1.51401	1.52622	9	0.132353	15.51707	15.64222	2	0.000006	0.80516	0.81165	18	0.688426
I0059	12.02036	11.21969	9	0.000012	25.04634	23.37802	2	0.000022	2.15362	2.01017	18	0.006873
I0022	4.04735	3.91249	9	0.000024	23.09233	22.32284	2	0.000024	0.6067	0.58648	18	0.912008
I0036	9.16153	8.49985	9	0.000000	15.25653	14.15464	2	0.000027	1.2649	1.17354	18	0.273767
I0090	6.03237	7.34423	9	0.000035	17.05416	20.76295	2	0.000033	0.79919	0.97298	18	0.488253
I0060	1.25365	1.26486	9	0.250667	18.45106	18.61609	2	0.000036	1.90624	1.92329	18	0.010678
I0117	33.91628	30.24608	9	0.000000	11.14206	9.93634	2	0.000036	0.87288	0.77842	18	0.728170
I0021	4.08807	4.42762	9	0.000000	25.65228	27.7829	2	0.000045	2.17177	2.35216	18	0.001053
I0067	1.07275	1.11536	9	0.347634	15.08534	15.68461	2	0.000047	0.69038	0.71781	18	0.796038
I0151	78.2207	66.48229	9	0.000011	15.52939	13.19893	2	0.000049	1.92465	1.63582	18	0.043682
I0137	34.97913	53.51481	9	0.000004	10.92836	16.71937	2	0.000054	0.81783	1.25121	18	0.210248
I0174	45.74099	40.56411	9	0.000000	12.65168	11.21978	2	0.000064	1.23339	1.0938	18	0.351024
I0119	3.12762	3.36735	9	0.000369	10.09984	10.874	2	0.000071	0.74011	0.79684	18	0.706315
I0029	3.22907	3.62139	9	0.000137	9.22125	10.34159	2	0.000077	1.45673	1.63371	18	0.044021
I0076	15.38673	20.82129	9	0.000098	8.59279	11.62774	2	0.000081	0.72578	0.98212	18	0.477224

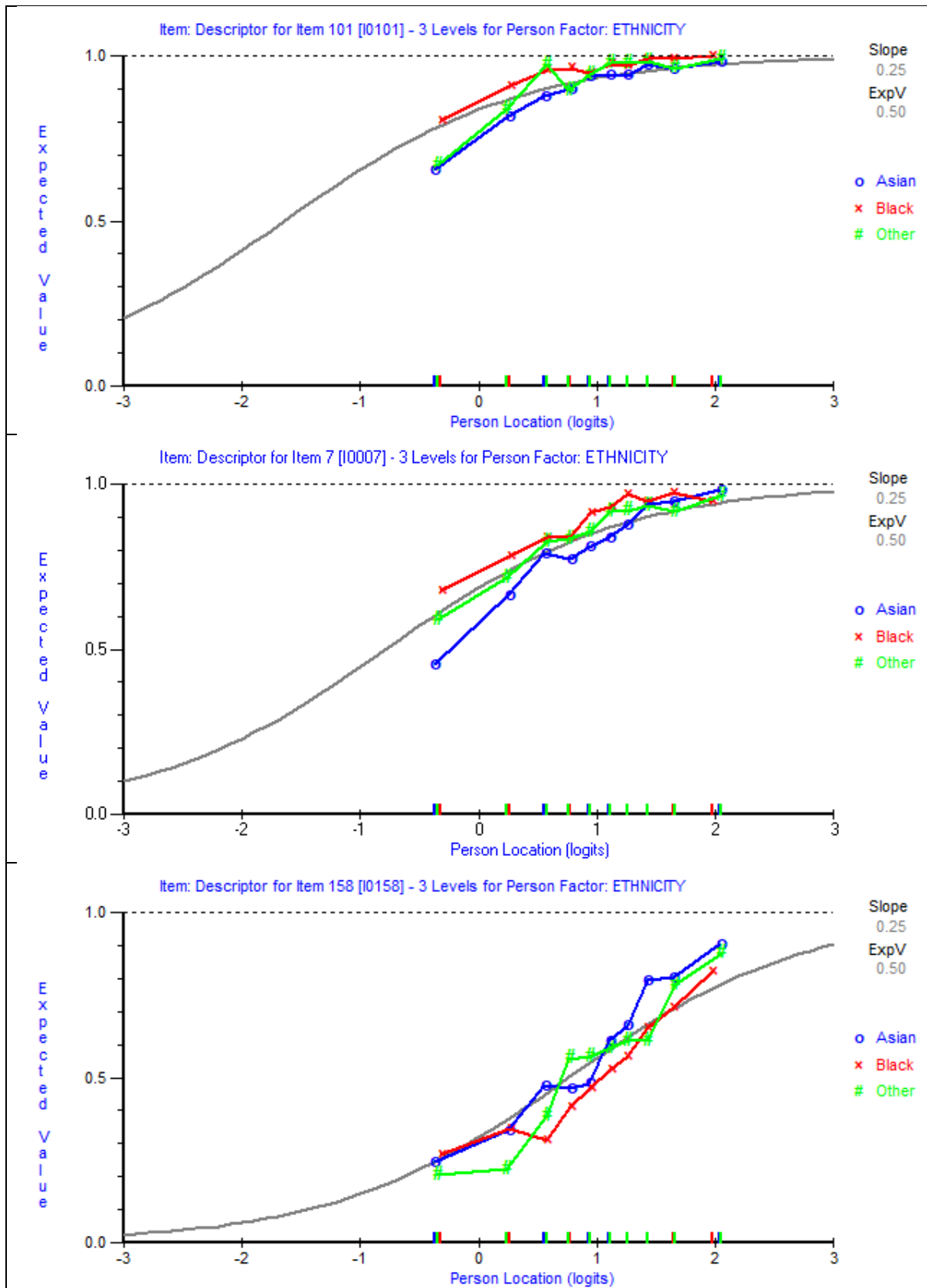
**Table 9: Summary DIF data by ethnicity for February 2024**

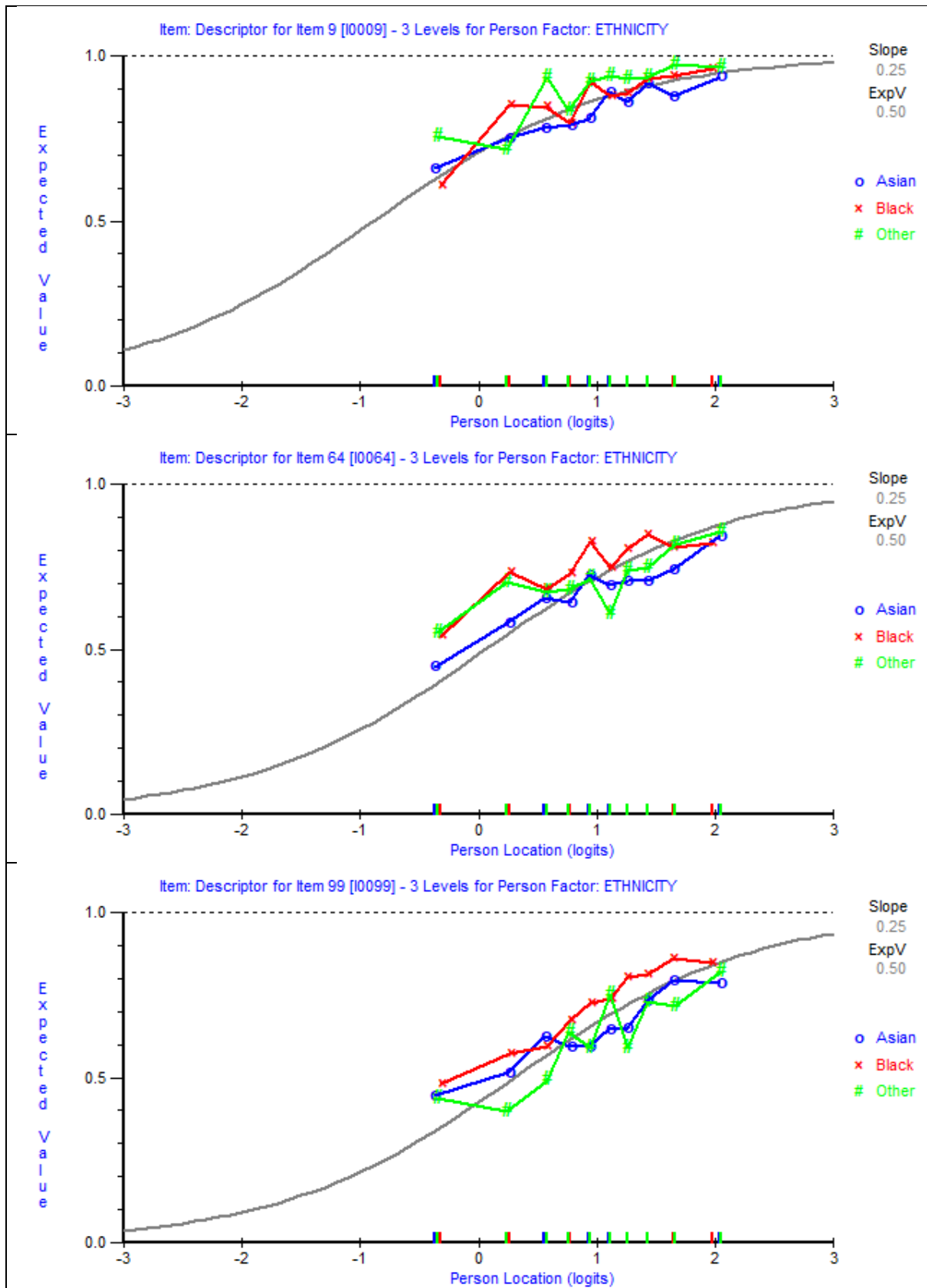












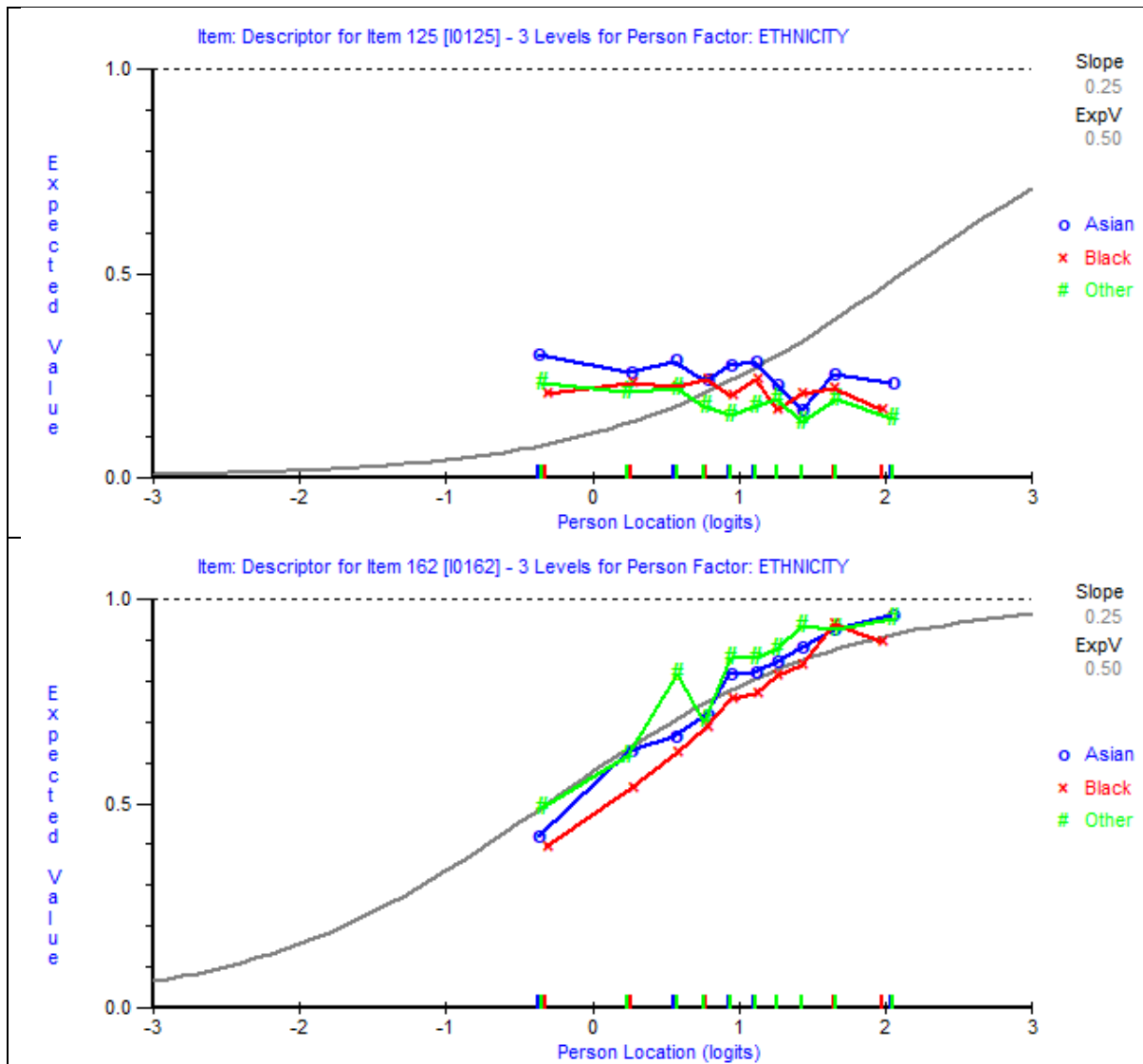


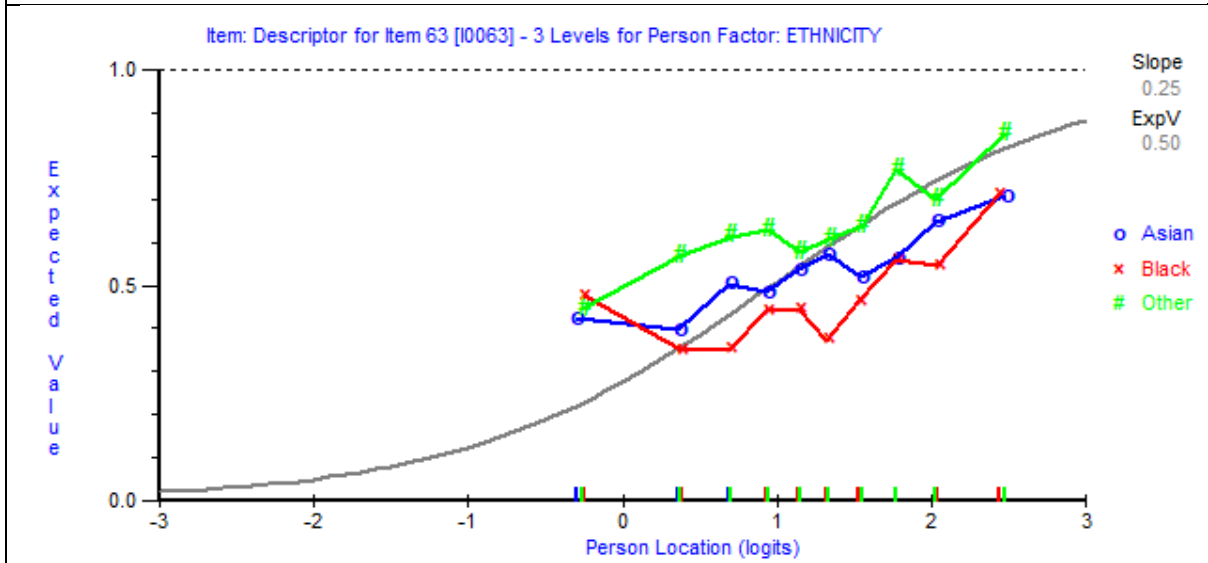
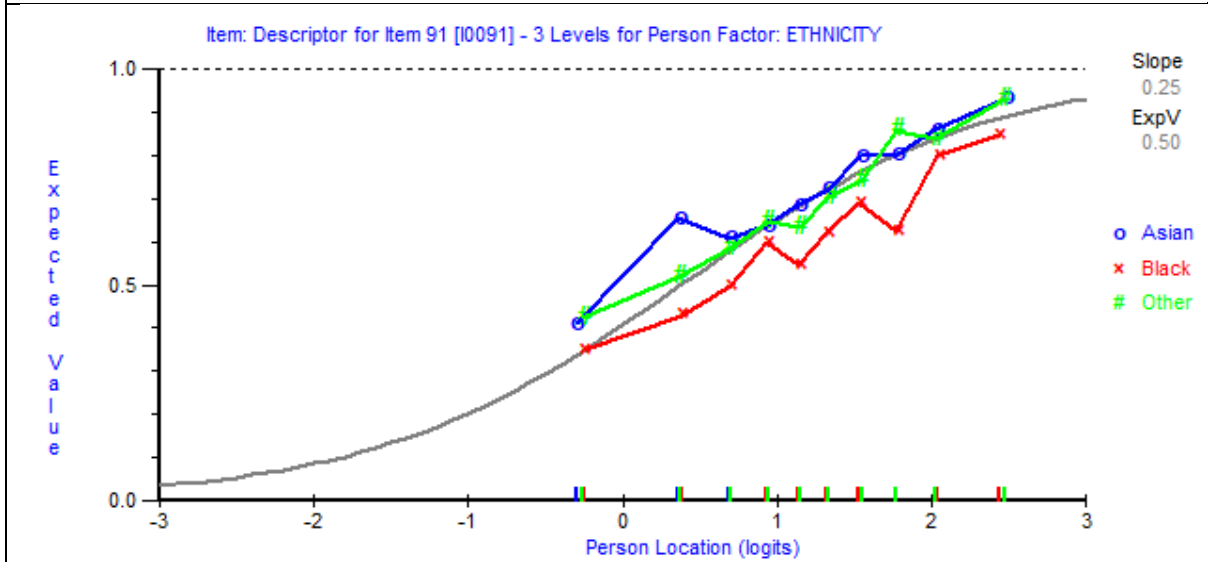
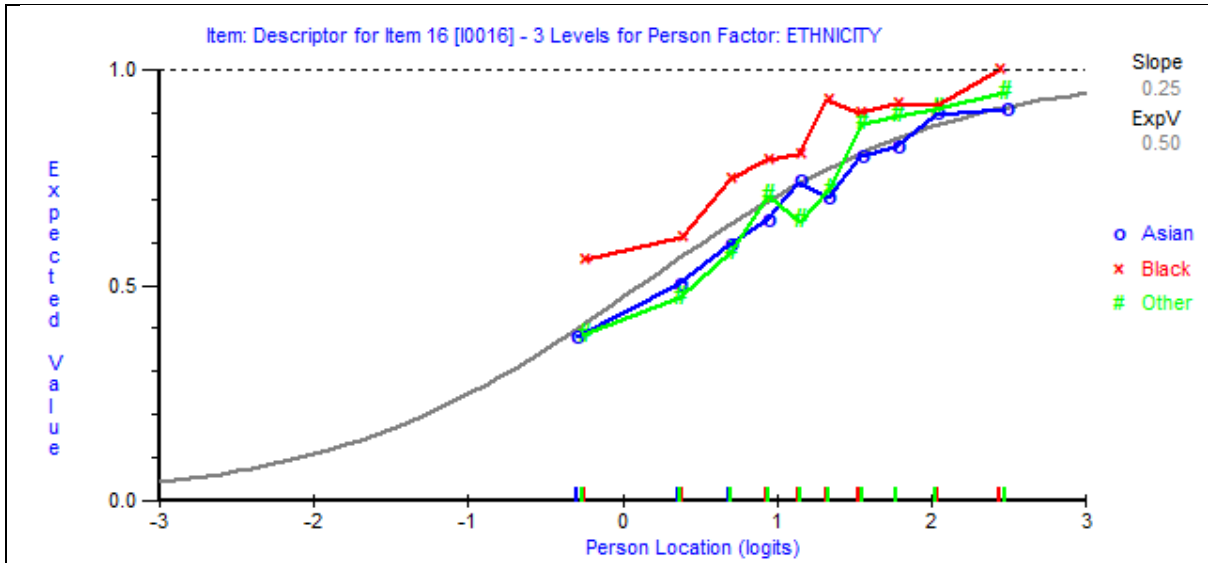
Figure 15: ICCs by ethnicity for February 2024

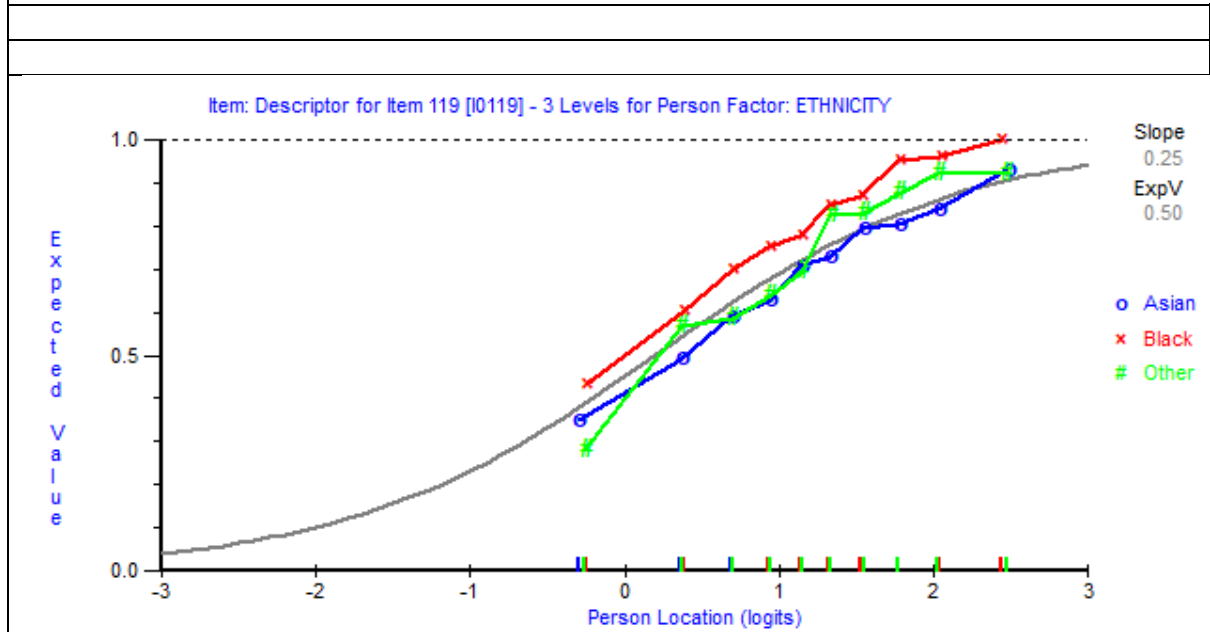
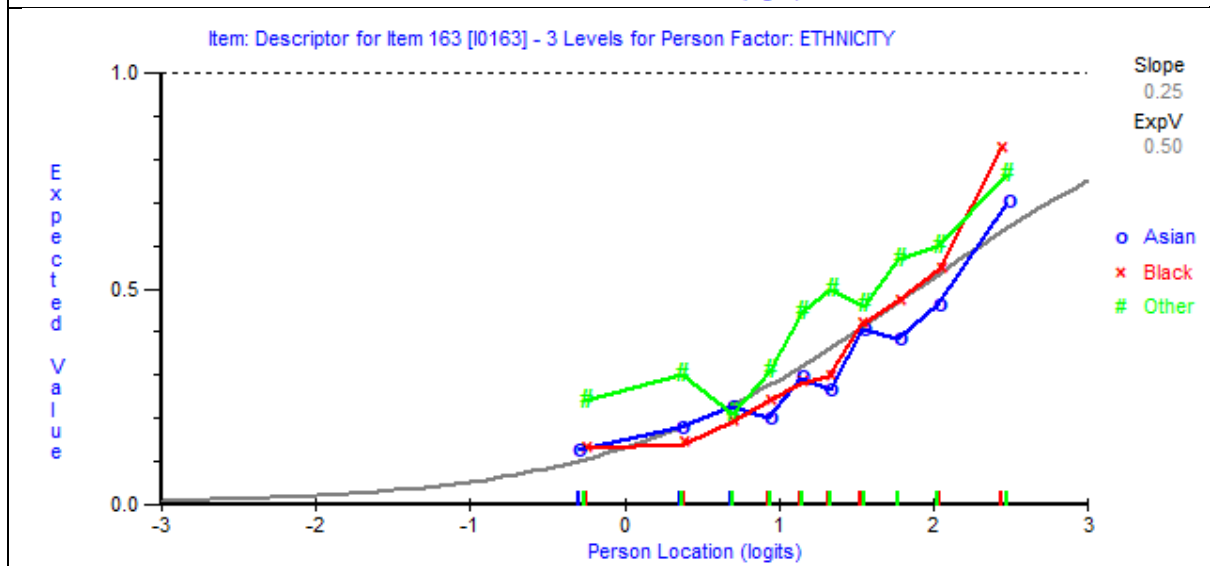
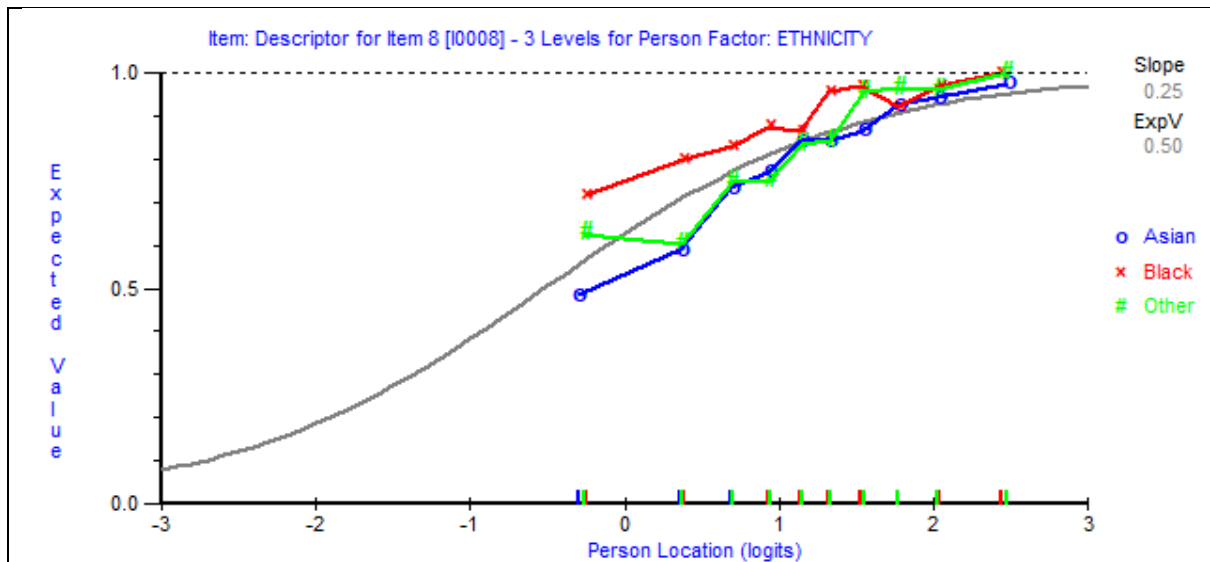
**DIF analysis May 2024 by ethnicity full details**

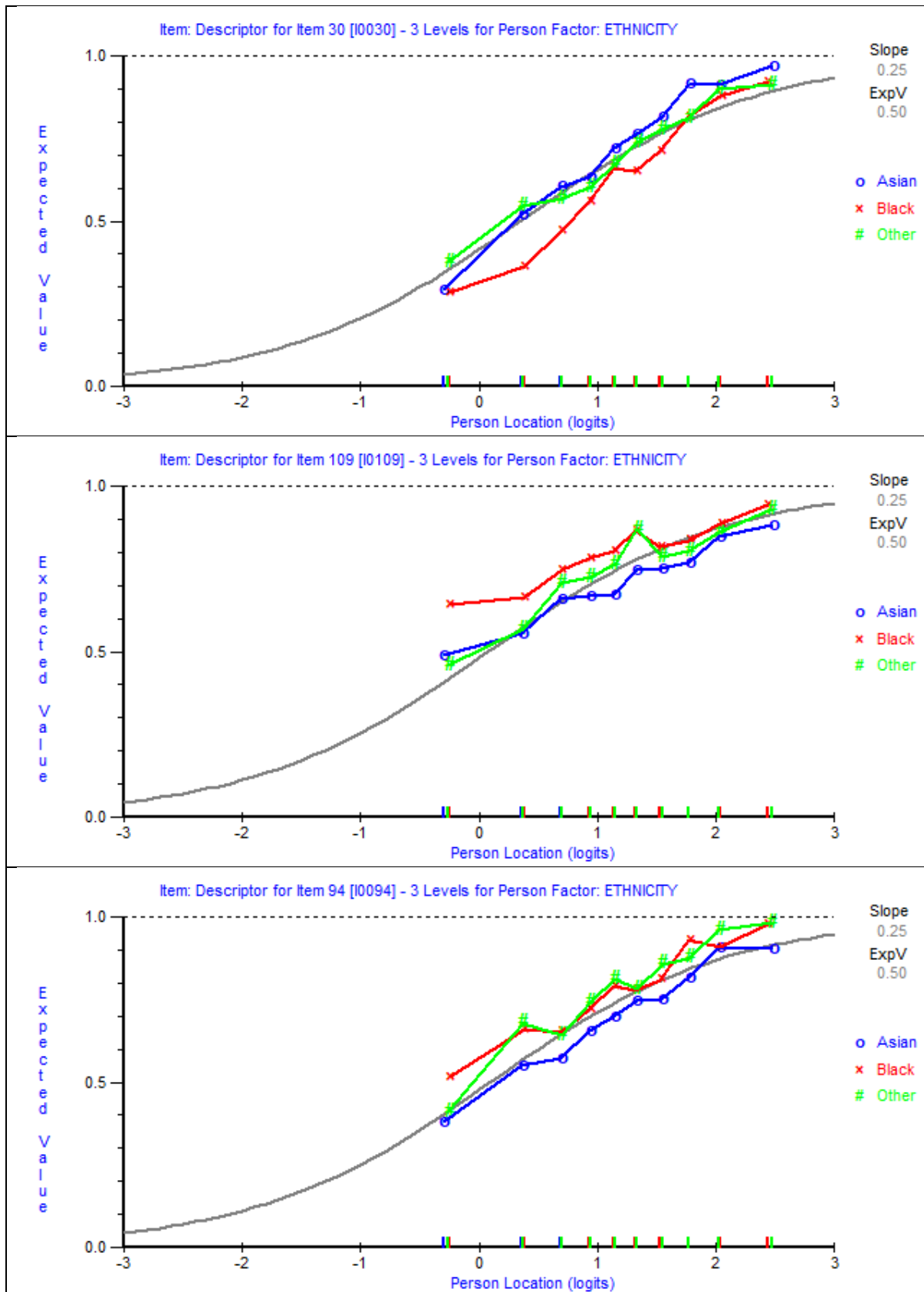
Table 10 and Figure 16 give the tabular and graphical results of the DIF analysis by ethnicity for May 2024. The table contains full details for all items with statistically significant DIF but given their prevalence (n=40) I have again only included graphs for the 'worst' 20 items in this regard.

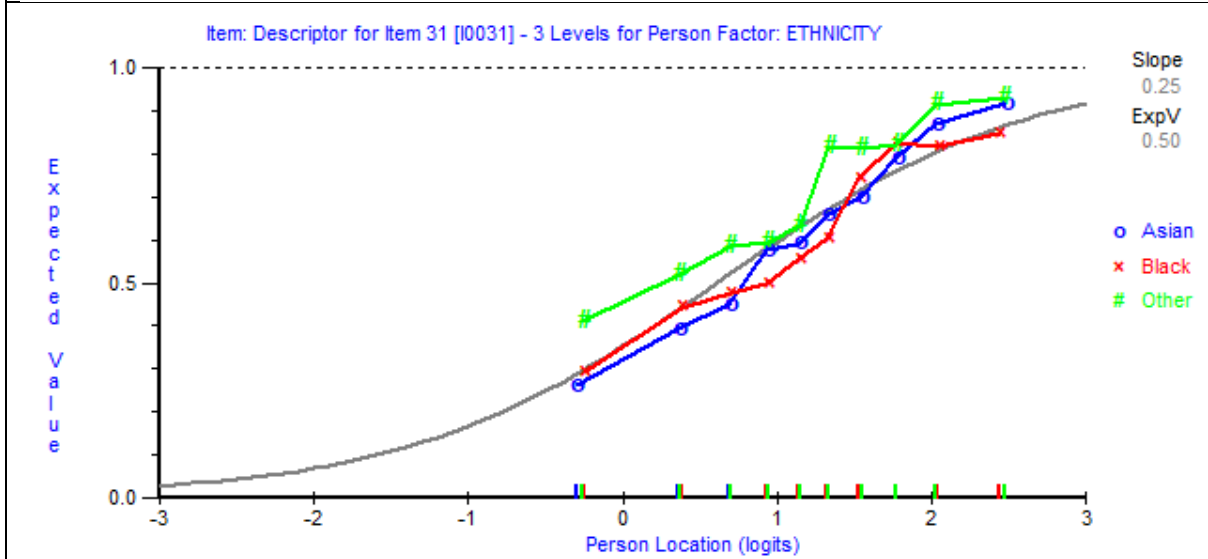
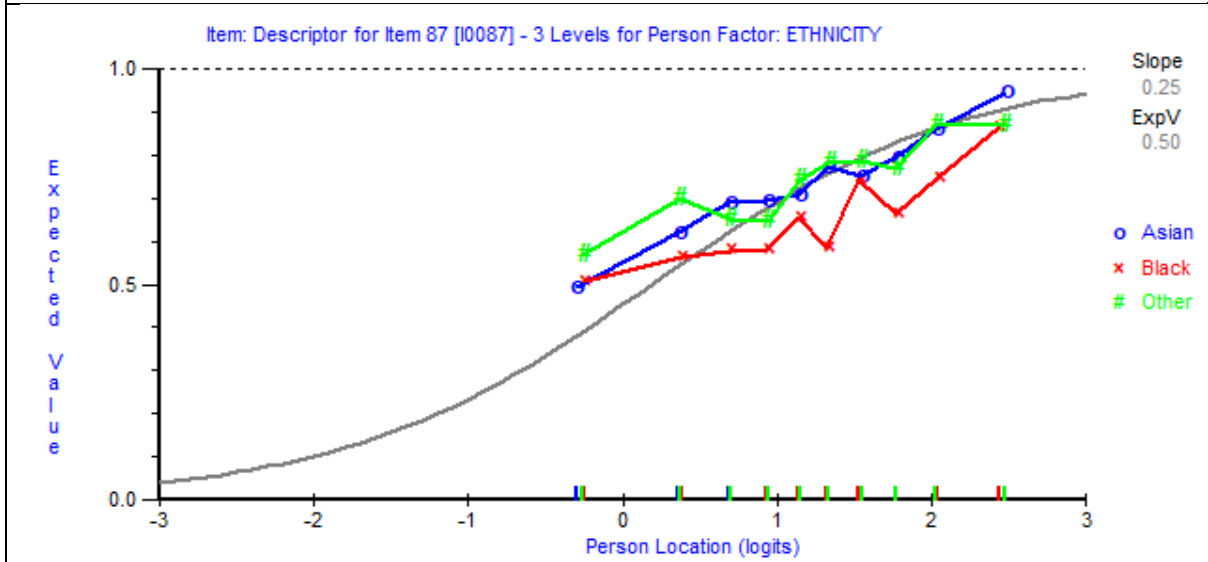
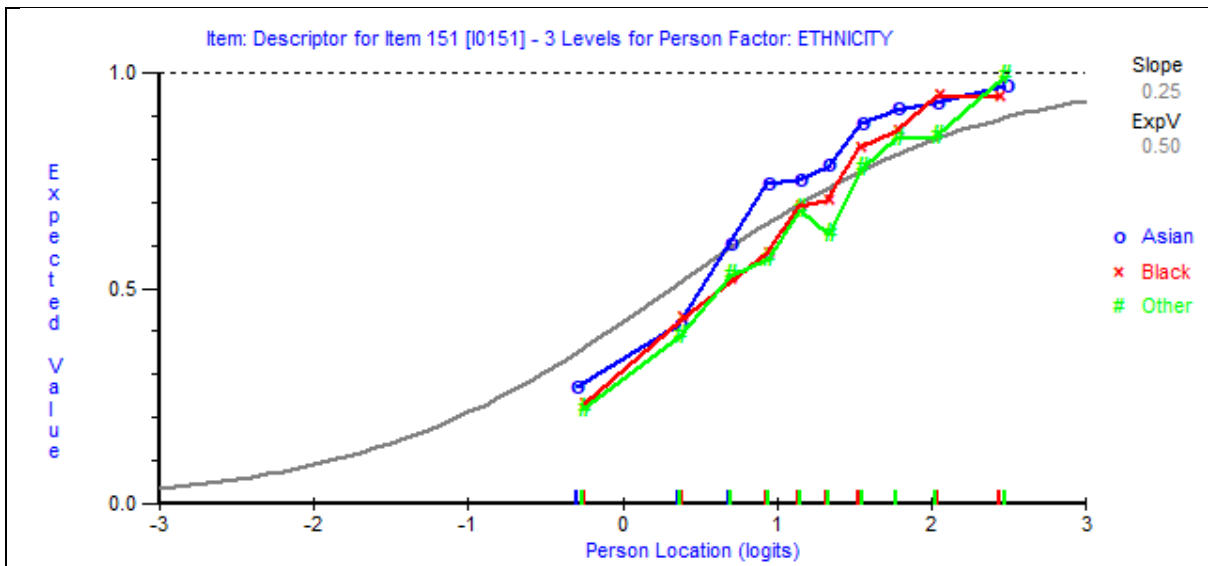
Rasch item fit					DIF main				DIF interaction			
Item	MS	F	DF	Prob	MS	F	DF	Prob	MS	F	DF	Prob
I0016	1.46507	1.57715	9	0.115856	30.59866	32.9396	2	0.000000	1.64087	1.76641	18	0.023513
I0091	3.37603	3.40809	9	0.000431	28.99285	29.26814	2	0.000000	0.93807	0.94698	18	0.520088
I0063	28.01407	25.21826	9	0.000016	31.36372	28.23362	2	0.000000	2.52562	2.27357	18	0.001570
I0008	3.25662	3.7711	9	0.000000	21.77959	25.22032	2	0.000000	2.05084	2.37483	18	0.000906
I0163	4.34706	4.38261	9	0.000000	24.895	25.09858	2	0.000000	1.55403	1.56674	18	0.059558
I0119	1.49293	1.62577	9	0.101942	22.52512	24.52937	2	0.000000	0.98314	1.07062	18	0.375648
I0030	5.96788	6.84299	9	0.000000	18.91544	21.68915	2	0.000000	0.51812	0.5941	18	0.906682
I0109	5.63067	5.35909	9	0.000000	22.58708	21.49763	2	0.000000	0.48503	0.46164	18	0.973557
I0094	2.38497	2.53508	9	0.006633	19.96367	21.22025	2	0.000000	0.70114	0.74527	18	0.766152
I0151	14.71704	18.56738	9	0.000000	15.6948	19.80096	2	0.000000	1.03745	1.30888	18	0.170572
I0087	8.08129	7.71267	9	0.000014	17.71226	16.90433	2	0.000000	1.52268	1.45322	18	0.096726
I0031	3.86558	4.21692	9	0.000000	15.48023	16.88722	2	0.000000	0.90999	0.9927	18	0.464560
I0055	2.38934	2.61404	9	0.005250	15.40138	16.84978	2	0.000000	1.74721	1.91152	18	0.011399
I0123	9.56461	8.88097	9	0.000000	18.05622	16.76563	2	0.000000	0.7286	0.67652	18	0.837697
I0092	5.99602	6.71144	9	0.000000	13.3214	14.91085	2	0.000000	1.15345	1.29108	18	0.182136
I0153	7.97249	9.60304	9	0.000000	11.94637	14.38967	2	0.000000	1.53254	1.84598	18	0.015925
I0049	9.04062	8.51809	9	0.000086	14.38687	13.55533	2	0.000000	2.01848	1.90181	18	0.011948
I0079	4.60642	5.48019	9	0.000110	9.30531	11.07041	2	0.000000	0.92805	1.10409	18	0.340395
I0103	3.55733	3.94152	9	0.000067	9.69036	10.73692	2	0.000000	0.86774	0.96146	18	0.502298
I0078	25.61777	22.40418	9	0.000000	12.20332	10.67249	2	0.000000	1.70084	1.48748	18	0.083840
I0084	1.9029	2.12626	9	0.024234	9.11719	10.18735	2	0.000000	1.12194	1.25363	18	0.208398
I0145	25.32163	22.53446	9	0.000000	11.33022	10.08309	2	0.000000	0.34915	0.31072	18	0.997571
I0080	1.22089	1.24893	9	0.259856	9.04568	9.25344	2	0.000000	1.9239	1.96809	18	0.008515
I0168	2.73776	2.94926	9	0.001672	10.98671	11.83548	2	0.000003	1.32001	1.42199	18	0.109904
I0102	4.30692	4.88065	9	0.000048	13.05314	14.79195	2	0.000004	1.08829	1.23326	18	0.223891
I0035	2.56166	2.40697	9	0.010137	41.15163	38.66663	2	0.000007	0.47555	0.44683	18	0.977895
I0099	1.57435	1.63511	9	0.099332	25.01418	25.97949	2	0.000011	0.93088	0.9668	18	0.495777
I0096	7.92087	10.61997	9	0.000128	8.96656	12.02199	2	0.000012	1.85946	2.49309	18	0.000418
I0141	0.74873	0.74809	9	0.664906	12.7964	12.78548	2	0.000037	0.52913	0.52868	18	0.946397
I0005	1.68482	1.84782	9	0.055166	9.42364	10.33533	2	0.000037	1.47891	1.62199	18	0.046522
I0125	16.53448	20.37671	9	0.000000	16.2947	20.08121	2	0.000040	0.96361	1.18752	18	0.261442
I0018	5.35943	6.7141	9	0.000000	8.97914	11.24874	2	0.000042	1.66943	2.0914	18	0.004475
I0106	11.90601	11.50125	9	0.000111	24.36965	23.54117	2	0.000047	1.2379	1.19582	18	0.254330
I0134	3.52882	4.15294	9	0.000000	49.62271	58.39924	2	0.000053	1.51565	1.78371	18	0.021634
I0132	16.3367	23.47152	9	0.000113	19.16802	27.53938	2	0.000056	2.21669	3.18479	18	0.000039
I0064	38.53787	32.82561	9	0.000089	15.53786	13.23477	2	0.000056	1.27111	1.0827	18	0.362731
I0001	24.80129	22.22331	9	0.000038	21.57932	19.33625	2	0.000060	1.75799	1.57526	18	0.057320
I0060	1.07046	1.11269	9	0.349600	11.4428	11.89416	2	0.000065	0.79066	0.82184	18	0.676004
I0133	5.89402	7.14098	9	0.000102	7.23928	8.77086	2	0.000076	0.89962	1.08995	18	0.355060
I0043	2.40514	2.86996	9	0.002166	9.35667	11.16495	2	0.000085	1.05468	1.25851	18	0.204844

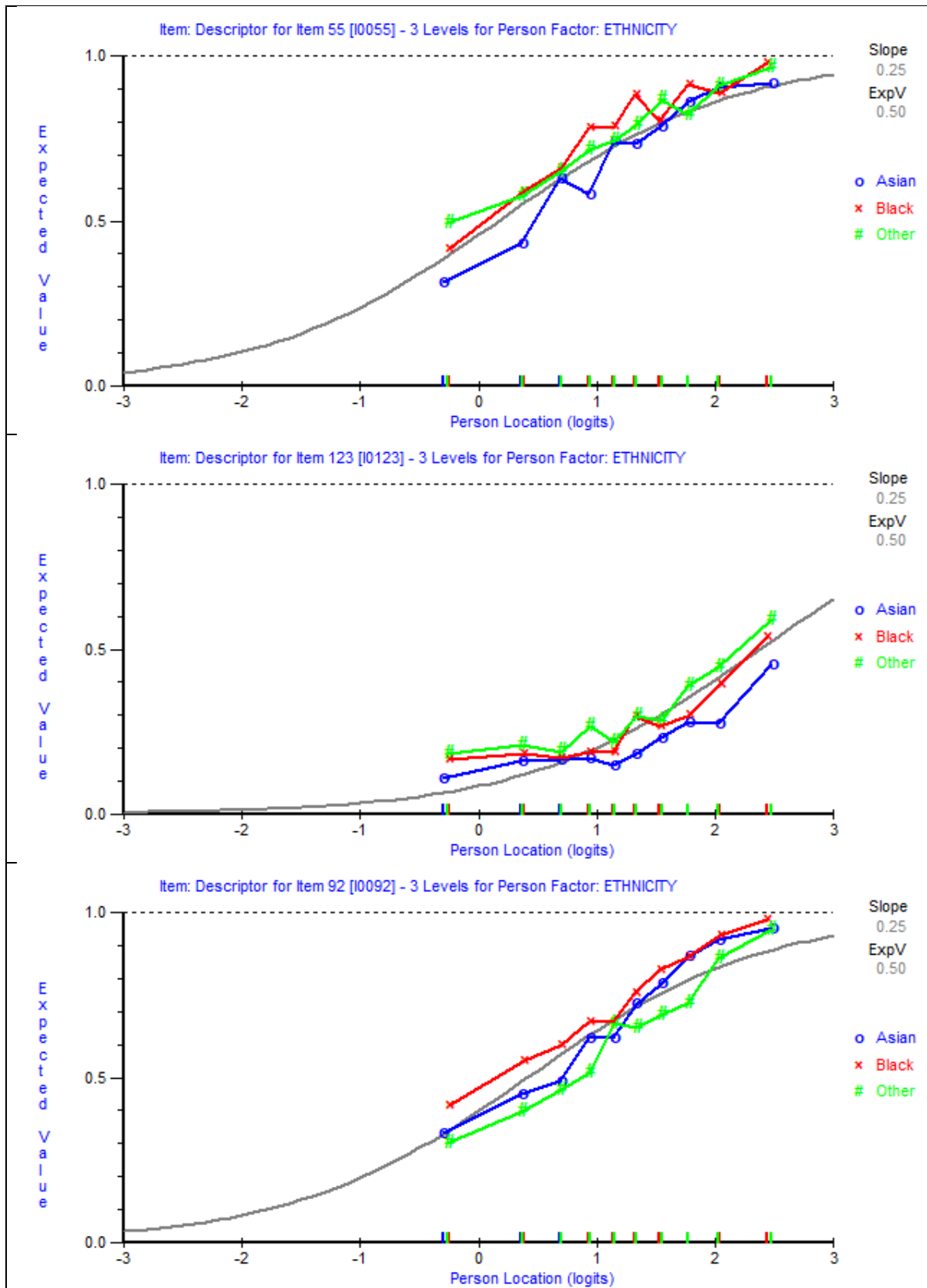
**Table 10: Summary DIF data by ethnicity for May 2024**

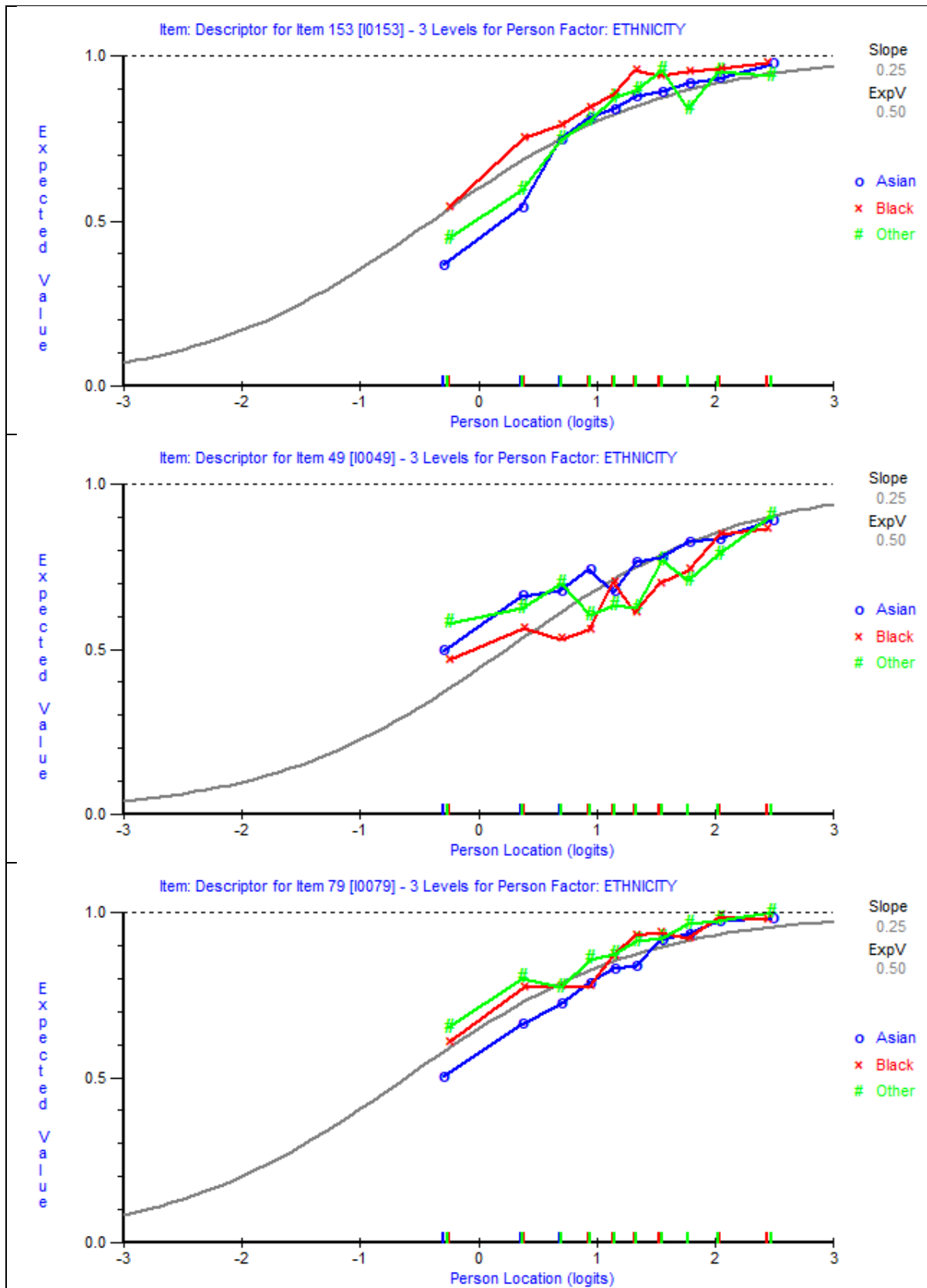












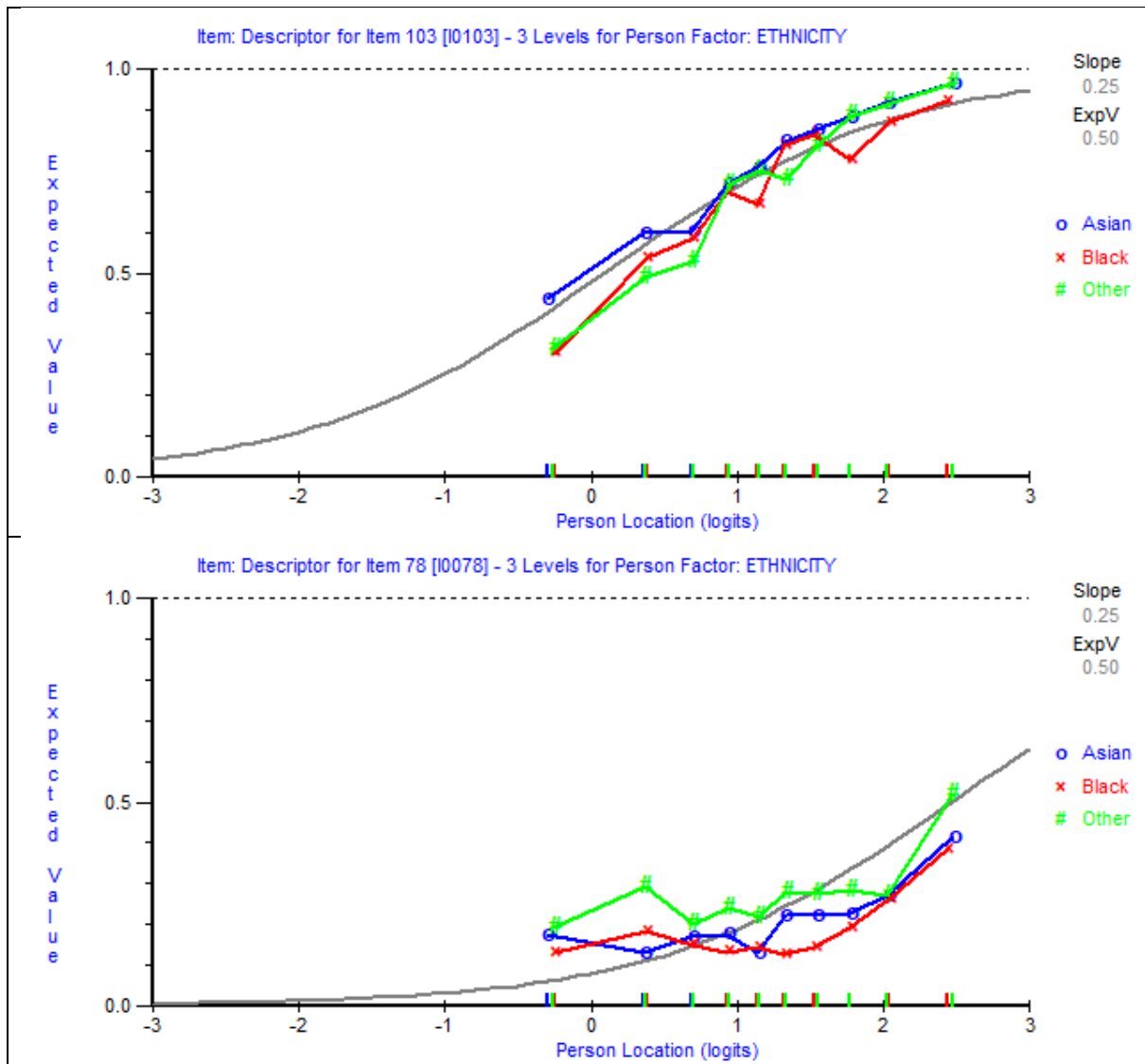


Figure 16: ICCs by ethnicity for May 2024