

Annual DIF Review Process

- A. With regard to items that were flagged for DIF on the operational test, the WVDE staff accepts the DIF flag does not confirm or prove bias, but alerts the state that the item is functioning differently for members of different subgroups. The state determined not to suppress the DIF flagged items from scoring in the first year, but used the Delta flag to study all Level B (moderate) and Level C (strong) items.
- 1) The research staff reviewed the values placed as per the Mantel-Haenszel metric for interpretive purposes for the DIF alerted items (see pages 568-569 of Technical Document). The flagged items pages are presented on pages 573-586 of Technical Document.
 - 2) The content staff reviewed each DIF alerted item to determine if the item exhibited any type of content bias.
- B. Each item that exhibited a Level B or Level C DIF flag was placed on a **DIF Item Alert List** for future review on subsequent WESTEST 2 administrations.
- C. C. If future operational test data flags the same item or items as exhibiting bias, the repeat DIF flagged items will be studied for effect size* and reliability to determine if the item/s should be removed from WESTEST 2 forms.
- 1) If the item is determined to exhibit bias in addition to the statistical parameters previously mentioned, WVDE will replaced the item/s as per the refresh clause provided for in contract. WVDE will require any replaced will be required to exhibit the same item statistics so as to ensure the item aligns with the reference curves for each grade level and content area.
 - 2) If item is determined not to exhibit bias beyond the statistical parameters previously mentioned, WVDE will elect to keep the item/s in future WESTEST 2 forms.
- *WVDE believes that fairness is an issue whenever the measured ability differences between subgroups are overly large. However, if the criterion for large difference is lacking, another way to evaluate the magnitude of the differences is to calculate the effect size. WVDE would use the Cohen's d is used to calculate the effect size. Cohen's d , then, expresses the difference in group means as a function of standard deviations.
- D. If future operational data does not indicate the item to be a repeat item that exhibits bias, item will be kept in forms.