The DIF-Free-Then-DIF Strategy for the Assessment of Differential Item Functioning

Three methods to establish common metrics

1. EMD method: the mean item difficulties of the test for the two groups are constrained to be equal.

Software: ConQuest, BILOG-MG PARSCALE

EMD is true when (a) no DIF item, or (b) multiple DIF items but some favor the reference group and others favor focal group, and the DIF amount are identical. i.e., ASA ≒ 0. When ASA >>0, EMD is severely violated.

When applied EMD method, overall, the DIF amount in a test is forced to be balanced between groups. Therefore, half items will be classified as favoring the reference group while another half as favoring the focal group. i.e., DTF=0.

1. AOI method: all but the studied item in the test are treated as DIF-free and served as a matching variable

Software: Winsteps, IRTLRDIF, CFA-based DIF detection

AOI is true when (a) the test is perfect, i.e., no any DIF item, or (b) the studied item is the only DIF item in the test

1. CI method (CI): select a subset of items to serve as anchor to establish a common metric for DIF assessment

The CI method yields appropriate DIF assessment even when the test contains 40% DIF items. In addition, the longer the anchors, the higher the power rate of detecting DIF. Moreover, four anchors are generally enough to yield a high power rate.

The superiority of the CI method is verified in the CFA-based DIF assessment (MIMIC method) and IRT-based DIF detection technique.

1. Scale purification procedure: the purpose of purification is to remove DIF items from matching variables

Three major steps:

1. calibrate item parameters separately for each group, compute the linking coefficients (a slope and intercept) to place the reference & focal groups on a common metric and assess each item for DIF.
2. re-link group matric using only those items not identified as having DIF in the previous step, and assess each item for DIF.
3. repeat step b until the same items are identified as having DIF in consecutive iterations.

When the characteristic curve method is used to link matric in the Rasch model, the slope will be equal to unity, and the intercept will be equal to the mean item difficulty of the anchored items for the reference group minus the focal group.

DIF assessment method with scale purification procedure provide a substantial improvement over EMD and AOI method. But it will come worse when they are high percentage DIF items (>20%, constant DIF)

1. Woods’ Rank-based (RB) method to select DIF-free items
2. assess item 1 for DIF using all other items as anchors, and repeat this step until the last item is assessed.
3. compute the ratio of the likelihood ratio statistic (LR) to the number of free parameters for each item (*df*), i.e., LR/*df*.
4. rank order the items based on the ratio LR/*df*.
5. designated the desired numbers of items (generally 4 items) that have the smallest ratios.

Findings:

1. RB method has very high rate of accuracy in selecting clean anchors in large DIF items and large sample size. 🡪 unknown on small sample size
2. the rate of accuracy decreased when the percentage of DIF items increased. 🡪 need purification procedure
3. the designated anchors yielded more accurate DIF assessment than the AOI method in terms of Type I error rate and power rates. 🡪 support DFTD strategy.
4. Rank-based scale purification method (RB-S): to add a scale purification procedure to the RB method.
5. assess item 1 for DIF using all other items as anchors and repeat this step until the last item is assessed.
6. remove those items that are identified as having DIF from the anchors and re-assess all items for DIF using the new anchors;
7. repeat step b until the same set of items are identified as having DIF in two consecutive iterations
8. designed the desired numbers of items that have the smallest ratios (LR/*df*).

**Summary:**

RB-S method: use RB-S method to select four anchor and then assess the other items for DIF with the designated anchors (purification procedure, DFTD).

RB method: use RB method to select four anchor and then assess the other items for DIF with the designated anchors (no purification procedure, DFTD).

AOI method: all other items were treated as anchors when assessing the studied item for DIF (no purification, no DFTD)

AOI-S method: the AOI method with scale purification procedure (purification procedure, no DFTD)

PA: pure anchor method, in which four clean anchors were selected by design and then the other items were assessed for DIF.

Conclusion & Discussion:

Limitation of IRTLRDIF in implement CI method: when the CI method is implemented with IRTLRDIF, the test is treated as if it consisted of the anchored items and the studied item. For example, when assessing item 1 for DIF with the four anchor, IRTLRDIF treated the test as if it consisted of the four anchors and item. And, if there is only one item was used as anchor, the test was treated as consisting of only two items. As a consequent, the result is very poor. Hence IRTLRDIF will reduce the power rate of the CI method, and thus the DFTD strategy.