Buros Center for Testing - CCC Advisory Assessment Committee Training (September 2017)
Sample example created for training purposes based on fictitious data

CCC Standards: Fairness Empirical Review

- Sample a representative and sufficient number of students from CCC protected classes
- Data collected in last 3 years
- Methodology and results described
- Detail follow-up of flagged items (removal, revision, or retain)
- Mantel-Haenszel Statistic
    - Groups: Focal (protected) and Reference (non-protected) (Hispanic)
    - Directions: Negative (item more difficult for focal) and Positive (item more difficult for reference)
    - Magnitude (3 levels):
        - A – little to no difference (typically retained)
        - B – Small to moderate difference (removal/retention depends)
        - C – greatest difference (item is dropped or revised unless content is critical)

Other review considerations

a. Was an empirical fairness study of test data or item performance conducted reasonably recently (within the last 3-5 years)?
b. Did the study sample include enough students from protected classes served by CCC?
c. Were the focal and reference groups relevant to the ESL student groups served by CCC?
d. Were the follow-up investigations of any items flagged as potentially unfair appropriately conducted?
e. Were all items, prompts, tasks (if applicable) included in the studies?

Description of the Example

A Differential Item Functioning (DIF) study was conducted among all CCC students who took the test in general academic contexts in the 2018 – 2019 academic year (including fall, spring, and summer semesters). Samples from 16 CCCs were combined for the study. Sample demographics are shown in the table below. For each focal group under consideration (i.e., white, black, Hispanic, Asian), the comparison group was all other students in the study sample.

N	White	Hispanic	Asian	Black	American Indian/Alaskan Native	Filipino	Pacific Islander/ Hawaiian Native
14,452	12.40%	60.80%	21.30%	3.60%	0.90%	0.70%	0.30%

The method used for the DIF analysis was the Mantel Haenszel (MH) procedure. (See prior section for formula and description of the calculation.) Results were classified into three categories using standard thresholds. Items with delta absolute values greater than or equal to

1.5 were considered C-level and were evaluated further for possible bias. Results of the analysis are as follow in the table.

	Number of Items in MH Category A or B	Number of Items in MH Category C
Hispanic/non-Hispanic	57	3
Asian/non-Asian	56	4
Black/non-Black	58	2

Following completion of the study, the item writers reviewed all items flagged as C-level DIF and determined the content was necessary for the measurement of the target constructs for each respective content area of the specifications.