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

CCC Standards: Fairness Logical Review

- Panel qualifications/background described (represent CCC protected classes 2% or more)
- 2 panelists from each protected group
- Independent of item writers and test developer
- Review training, guidelines, and procedures
- Results of panel decision & response (removal, revision, or retention of items)

Other review considerations

a. Did the review panel members appropriately represent key demographic groups in the CCC ESL student population (e.g., based on age, gender, race/ethnicity, cultural/linguistic background)?
b. Were the panel members' qualifications and background appropriate for representing the CCC ESL student populations (or groups)?
c. Would any of the panel members have any potential conflicts of interest?
d. Was the review conducted reasonably recently (within last 3-5 years)?
e. Were the training, review process, and results described in enough detail to indicate if the review was appropriately conducted for CCC ESL students?
f. Did the fairness review involve more than just the test items but also a review of the instructions for administrators and test takers and/or the scoring rubrics/algorithms (if applicable)? Are you satisfied with the follow-up decisions made about items that were identified as potentially unfair by the panel?
g. If computer-based or computer-adapted tests were reviewed, was there supportive evidence that lack of familiarity with the technology or item-type would not unfairly impact students' ability to demonstrate their English language skills?

Description of the Example

17 panelists participated in the fairness review. All panelists had doctoral degrees. Panelists were recruited from among those working with ESL students at the community college level, and the recruitment incorporated consideration of gender balance and ethnic/racial group diversity. At least two members of the panel represented each of the ethnic/racial groups that represent more than 2% of the CCC student population: Filipino, African American, Asian, Hispanic, and multi-ethnicity. Nine of the panelists were female, and eight of the panelists were male. None of the panelists had been involved in writing the test items. Review protocols were conducted using company's standardized guidelines for bias and fairness review, which have undergone extensive evaluation from professionals in the field to ensure they align with best practices. Prior to the day of the review panel, panelists began by taking online training to introduce them to the review process prior to the in-person review. Next, at the in-person review, the presenter reviewed the guidelines and facilitated a discussion among panelists to ensure all the panelists understood the guidelines. The day proceeded in blocks of 10

questions, in which the reviewers independently rated each item according to the fairness guidelines and then the panel discussed all items as a group. If panelists' views on the items changed as a result of the discussion, they could indicate a revised rating, but the initial rating was also kept in the records. This process was completed throughout the day until all items were independently rated and discussed. Guidelines for the fairness review are included at the end of this section. In the ratings, panelists indicated if they thought the question had no problems, had a problem that could be fixed, or had needed to be removed completely. 130 multiple-choice items were reviewed. Of these, the panelists determined 112 had no problems, 16 had problems that could be fixed, and three needed to be removed completely.

Relevant Excerpts from Guidelines

- Topics to avoid - Highly sensitive or controversial topics among various population groups, generally not required by the constructs being measured by the tests, should typically be avoided.
- Portrayal - All population groups should be portrayed fairly, authentically, and with respect. No group should be depicted in a demeaning (including self-demeaning) way by test materials.
- Stereotyping - Instances of stereotyping should be avoided, whether the stereotype is negative or "positive." Examples of stereotyping include portraying all women as homemakers and cooks; girls as ballerinas and boys as athletes; African Americans as people who live in the inner city or who excel exclusively in sports; Asian Americans as particularly gifted in math; Native Americans as particularly attuned to nature; or older people as feeble, slow, incompetent, or dependent.
- Language - Foreign words and phrases, slang and dialect, and idiomatic expressions should be avoided in testing materials unless their use is pertinent to the construct.
- Ethnocentrism - Test materials shouldn't treat aspects of U.S. or Western political systems, society, cultures, or values as universal when they are specific to those regions.
- Regionalisms - Regional differences are often subtle. Referring to "yard sales" or "garage sales," for example, could be confusing for urban test takers. Generic terms are, therefore, generally preferable to region-specific ones. For example, a term such as "soft drink" is usually preferred over "pop," "soda," or "Coke."
- Highly emotional content - Special care should be taken to avoid content and contexts that seem likely to trigger or add stress to the testing context. Highly controversial topics are often inappropriate for a testing context, during which students are timed and have no emotional outlet.