

## **Investigations for Bias, Offensiveness and Sensitivity**

### **Materials needed:**

1. Current copies of instrument(s) (to include all test forms), enough for each participant.
2. In the case of performance assessments and direct writing assessments, the prompts (i.e., questions and stimulus material) and associated scoring rubric.

### **Persons supplying data:**

Impacted class members (see Methods section which follows).

### **Method:**

1. The procedure recommended (and expected as a minimum requirement for any Approval level) is referred to as a “logical review.”
2. Form a panel of impacted class members (membership based on ethnicity, culture, gender, disability, language, etc.); **majority class members are not to be involved in this review**. Typically 2 to 3 impacted class members at a minimum should represent each group. Participation is limited to impacted/minority group members.
3. In the case of ESL/ELL assessments, a panel made up of members of the language/linguistic groups typically sitting for the college’s assessment into ESL courses should be assembled for the review.
4. It is better if a class/language member could supervise and conduct the panel and its review.
5. Discuss with panelists the need to review the assessment for fairness based on their ethnic, cultural, group membership. Ask each panelist to review/inspect test questions for evidence of bias, offensiveness or insensitivity. Notes associated with “problem” items should be made by the individual. Following independent review, the panel discusses findings (item-by-item consideration).
6. A record needs to be kept as to test questions that a group identifies as problematic or in need of attention. Once the item review is completed, it is advised that each person be asked (via a survey) if the test, with appropriate adjustments would be a fair assessment, i.e., is acceptable and if not, why not. Findings and input needs to be considered by the college using the test and a final determination made as to usability, modification, adjustments, etc.

**Other Guidelines:**

1. *Persons who developed the assessment should not be the ones to review the test and decide whether it is fair and free from bias.* This approach is too subjective and the conflict of interest inescapable. Developers, even impacted class member developers can miss problems with their questions. The procedure needs to be carried out by an independent reviewing group.
2. *If a College does not have a sufficient number of faculty needed to do an independent review, consider using students who have already completed the course and who represent the impacted groups or use individuals from the local community.* Reviews need not be faculty who represent the group.
3. *As a general rule, when a group constitutes 2 to 3 percent of the student population served, that group should be represented in the review.*
4. *If multiple forms of the assessment are in place, all must be reviewed.* Any test has the potential to be unfair. As such, panels should review all active and available test forms. In the case of writing assessments, all active prompts must be reviewed. A good resolve to issues of bias in writing assessments is to provide students with “choice” prompts (i.e., respond to either prompt (a), (b), (c), etc.).
5. *Majority rule should not be used to make final decisions.* Panels should discuss their individual findings and provide recommendations to the users. When a fault is found, the test needs to be corrected. It should not require a “majority” to cause change. If a panel had two persons representing a group and those members see a problem, they should not be “out voted” by others who do not consider the issue a problem. Reports need to address what was found, and the course of action followed.

## **Common Deficiencies in Test Bias Studies Submitted by Local Colleges and Preliminary Report Comment Examples**

### **Common Errors or Deficiencies in Evidence Submitted**

- 1. The persons serving on the test bias review panel are not described in terms of the groups they represent.**
- 2. An ethnic group being served by a college is not represented on the test bias review panel.**
- 3. A class of students serves as the review panel, but the results are not reported for each of the affected groups, i.e., gender and ethnicity.**
- 4. The attention is focused on the overall group review and not on the review of individual members of the panel representing specific ethnic groups.**
- 5. The review is conducted and items/prompts identified as being bias, insensitive or offensive, but the results are ignored and/or a convincing rationale for continued use of the items/prompts is not presented.**
- 6. The college needlessly conducts their own test bias review study when they could have used the information reported in the test's Technical Manual.**
- 7. Inappropriate persons are used as members of the review panel.**

### **Comment Example 1: Writing Sample**

A review of prompts for test bias must be completed by members representing affected groups, i.e., females and ethnic minorities. While 23 reviews were reported there was no indication of the gender or ethnic group composition of the reviewers.

### **Comment Example 2: ESL Writing Sample**

The persons who created the prompts are not appropriate judges of the fairness of the prompts. It is strongly advised that other not involved with development (e.g., students who have exited the program) serve as reviewers to judge the fairness/freedom from bias, etc.

**Comment Example 3: Math Placement Test**

The use of students in classes to review the items and make judgments on the insensitivity, offensiveness and fairness/freedom from bias is appropriate, but these data need to be reported by affected group membership (i.e., females, ethnic groups). White male responses should not be included.

**Comment Example 4: Holistic Writing Sample**

An exact description of the ethnicity of the steering committee members who review the prompts is needed.

**Comment Example 5: English Placement Writing Sample (ENGL)**

If new prompts are being used, the college needs to submit evidence addressing their bias.

**Comment Example 6: Calculus Test**

The bias review committee was composed of the same persons who developed the test items. This group is not viewed as an appropriate review panel to provide an “objective” appraisal of bias. An independent group should be convened to review the items. Also, if the campus serves a sufficient number of African-American students, at least one African-American should be included in the panel.

**Comment Example 7: CASAS - IRCA**

New test bias evidence is required only if items on the instrument have changed. The current write-up is not sufficient as a test bias study. If no change has occurred, no new data is needed.

**Comment Example 8: Nelson-Denny**

The test bias information reported by the test publishers in the test’s technical manual may be referenced to satisfy the requirement for this area of the Standards.

**Test Bias Data Collection and Reporting Examples**

**Sample 1**

Investigations for test bias were done for both faculty or staff and students. In the faculty/staff test, the following question was asked:

DOES THIS PROMPT SEEM TO BE CONFUSING OR IN SOME WAY OFFENSIVE TO AN INDIVIDUAL FROM YOUR CULTURE? If SO, HOW MIGHT IT BE WORDED DIFFERENTLY SO THAT THIS INDIVIDUAL WILL NOT BE AT A DISADVANTAGE WHEN RESPONDING TO THIS PROMPT?

**Faculty Bias Study**

Six members of the faculty/staff responded to the request. Their responses are reflected below.

**BIAS STUDY FOR FACULTY/STAFF**

DEMOGRAPHIC	Panel Member					
	1	2	3	4	5	6
<b>GENDER</b>						
Male	X	X				
Female			X	X	X	X
<b>ETHNICITY</b>						
Asian		X	X			
Black	X					
Caucasian					X	
Hispanic				X		
Filipino						X
Evidence of Bias for Gender	N	N	N	N	N	N
Evidence of Bias for Age	N	N	N	N	N	N
Evidence of Bias for Ethnicity	N	N	N	N	N	N

**COMMENTS:**

I found nothing culturally offensive or confusing to Filipinos.

Language could be very difficult for ESL students, but nothing is offensive for ethnicity or gender or age.

No bias in the content for Asian ethnicity. Universal themes.

No bias for age, gender, or ethnicity was found by any staff or faculty member.

**Student Bias Study**

A selected group of students, representing various ethnicities, age groups, and genders were also asked to evaluate the prompt. They were asked to comment on any “Perceived examples of bias for age, gender, or ethnicity.” Ten students participated in this process.

The results of their analysis follow:

**BIAS STUDY FOR STUDENTS**

Demographics	Panel Member									
	1	2	3	4	5	6	7	8	9	10
<b>GENDER</b>										
Male	X		X	X		X				X
Female		X			X		X	X	X	
<b>AGE</b>										
<20 years old	X	X		X			X		X	X
21-30 years old					X	X		X		
>30 years old			X							
<b>ETHNICITY</b>										
Asian									X	X
Black			X	X			X			
White					X		X	X	X	
Hispanic	X	X				X				
Evidence of Bias for Gender	N	N	N	N	N	N	N	Y#	N	N
Evidence of Bias for Age	N	N	N	N	Y*	N	N	N	N	N
Evidence of Bias for Ethnicity	N	N	N	N	N	N	N	N	N	N

**COMMENTS:**

\* “The sample contains more advanced language not directed to younger readers. There is an age bias toward more mature readers.”

# “Some might feel that the continuous use of the word “man” is an example of gender bias.

“Author uses ‘human’ and ‘man’ interchangeably, so in this case ‘man’ encompasses male and female.”

Most respondents had no comments.

Overall, no real evidence of any bias was found by the panel of students. We did clarify the writing instructions to address the issue of the use of the word “man” as being universal for all humankind to address any concerns that might be construed as gender bias.

## Sample 2

### **Investigation of Bias, Offensiveness and Sensitivity Summary of Findings**

During Spring 2005, an investigation on the bias of COMPASS Mathematics instrument was conducted. A committee of seven members: four Mathematics faculty, one faculty-counselor, and two administrators were convened to validate the instrument. The instrument from ACT contained 79 actual questions that were used in their assessment.

In terms of ethnic background, the group comprised of 4 African-Americans, one Hispanic, one Asian and one White.

The instrument is used to place students into the following classes, Math 1 Pre-Calculus, Math 2 Pre-Calculus with Analytic Geometry, Math 3 Calculus, Math 13 Statistics, Math 15 Math for Liberal Arts, and Math 50 Trigonometry. The raters were instructed to check for bias and offensiveness towards students. Furthermore, they were instructed to look at whether the instrument is sensitive towards students of all color and culture, including Caucasians.

The instrument, COMPASS Mathematics Standard Item Sets comprised of 79 questions: of which 17 (Q1-Q17) are Pre-Algebra/Numerical Skills questions, 19 (Q18-Q36) are Algebra questions, 15 (Q37-Q51) are College Algebra questions, 16 (Q52-Q67) are Geometry questions, and 12 (Q68-Q79) are Trigonometry questions.

In terms of percent of questions that are rated as “not bias”, all 7 raters (100%) gave a rating of at least 90%. These range from a low of 92% to 100% that the items are not bias.

Out of 79 questions, 70 questions (89%) were rated as not bias while 9 questions (11%) were rated as possibly biased. Those questions were then analyzed.

Questions Not Biased	
70	88.6%
4	5.1%
4	5.1%
1	1.3%
0	0.0%

In analyzing those questions, the issues raised were not really bias issues, but issues of the differences in the use of units in the U.S. and the metric units used by the other countries. For example, temperature only U.S. uses Fahrenheit while the other countries use Celsius. Other examples are miles versus kilometers, miles per hour (mph) versus kilometers per hour (kph), feet versus meters, and pounds versus kilogram.

Another question raised by three raters to be possibly age biased is the term “record” which was used instead of “CDs” in a question. Younger students may not understand what a record is thereby being disadvantaged. Another rater raised the use of the term “altitude” instead of “length”. This may distract the testing of some Math concept if a student’s vocabulary is somewhat lacking.

As for sensitivity towards students of different cultures, the faculty were satisfied that there were no questions that would be a problem. By and large, the whole test instrument was deemed to be relatively bias-free by all the raters.

### Sample 3

ESL English Language Proficiency Test Listening Item:

**Mark the word that is spelled correctly.**

**numbr**      nombre      number      numbor