

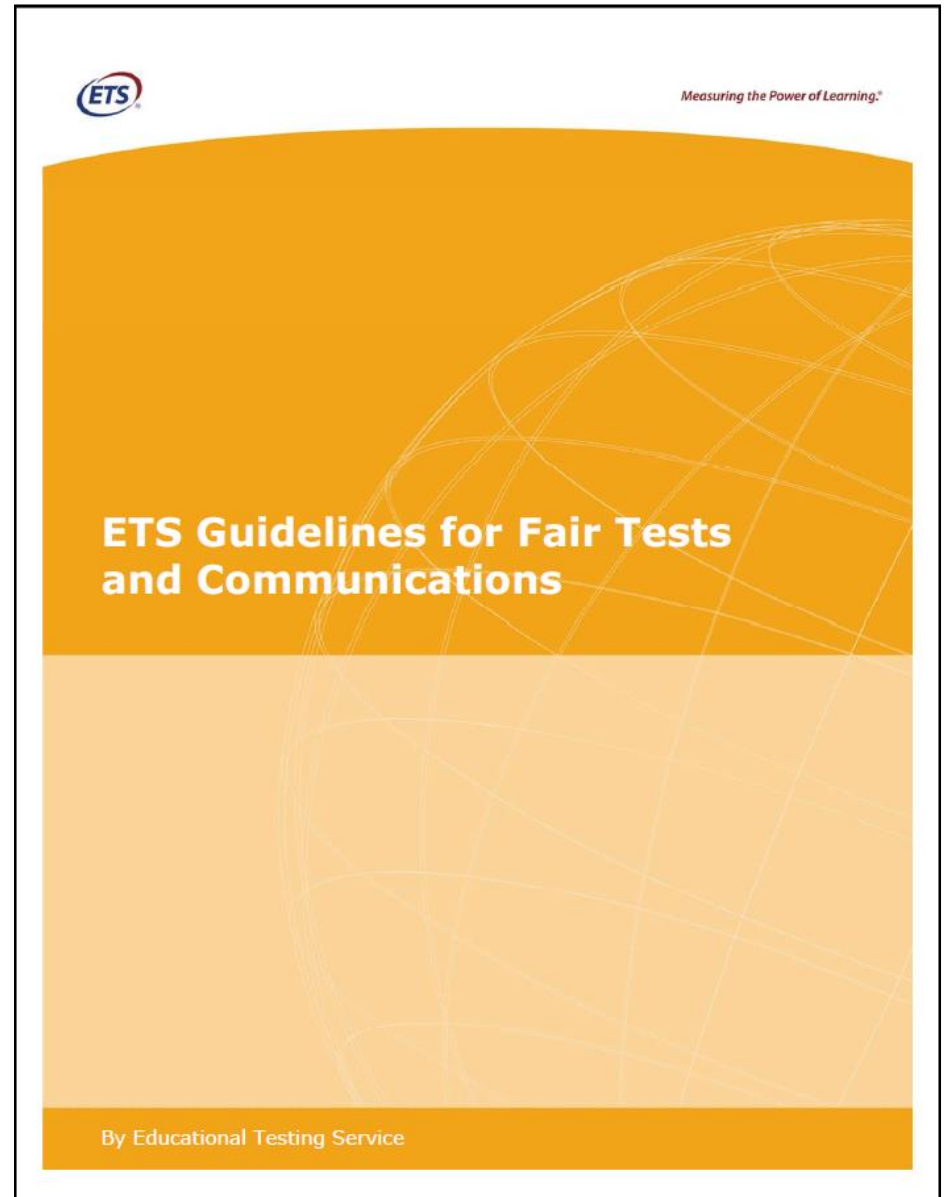
ETS Guidelines for Fairness Review

William Monaghan
Research Fairness Review
Coordinator



Purpose

The primary purpose of the *ETS Guidelines for Fair Tests and Communications* is to enhance the fairness of items, tests, and communications created by ETS.



Purpose

The guidelines are intended to help the people who design, develop, and review ETS materials to:

- Obtain a better understanding of fairness
- Take fairness into account as materials are designed,
- Avoid the inclusion of unfair content or images as materials are developed,
- Find and eliminate any unfair content or images as materials are reviewed,
- Represent diverse people in materials, and
- Help reduce subjective differences in decisions about fairness.

Fairness as a Question of Validity

For test designers, developers, and reviewers it is useful to think of fairness in assessment and even other materials in terms of validity.

Messick (1990) defined validity as "an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the *adequacy* and *appropriateness* of inferences and *actions* based on test scores."

Messick, S. (1990, August). Validity of test interpretation and use (ETS Research Report No. RR-90-11). Educational Testing Service, Princeton, NJ, p.1. Available at: <http://files.eric.ed.gov/fulltext/ED395031.pdf>

Fairness as a Question of Validity

Validity is the extent to which a test measures the knowledge, skills, or other attributes (KSAs) in a way that provides scores that function appropriately.

KSAs are referred to as the *construct* a test is intended to measure. KSAs that are part of the construct is *construct relevant*; KSAs not part of the construct are *construct irrelevant*.

Measuring KSAs in the “right way” requires measuring a suitable sample of the construct-relevant KSAs while minimizing the inadvertent measurement of any construct-irrelevant sources of score differences.

Fairness as a Question of Validity

The extent to which communications meet their intended purposes for their intended audiences is somewhat analogous to validity in tests.

Material identified as inappropriate may interfere with the ability of a publication to meet its intended purpose for intended audience.

Language that is more difficult than necessary to meet the purpose of a test makes the test less fair. Needlessly difficult language used in a publication or some other form of communication also will make the material less effective.

Fairness as a Question of Validity

Designing and developing fair tests and communications requires an understanding of what is meant by "fair" in those contexts. Discussions of fairness in communications have generally focused on:

- the avoidance of sexist language, offensive content, and stereotypes;
- the use of correct terminology for groups of people; and
- the representation of diversity.

Fairness as a Question of Validity

The fairness guidelines are not an exercise in political correctness. Their use is way to increase validity because fairness is essential for validity.

Shepard (1987) defined bias as “invalidity.” Practices that reduce bias, such as using the guidelines to help reduce irrelevant and potentially offensive material, will increase validity in context of tests and effectiveness in the context of communications.

Shepard, L. A. (1987). The case for bias in tests of achievement and scholastic aptitude. In S. Modgil and C. Modgil (Eds.), *Arthur Jensen: Consensus and controversy* (pp. 210-226). London: Falmer Press.

Fairness as a Question of Validity

An Additional Note

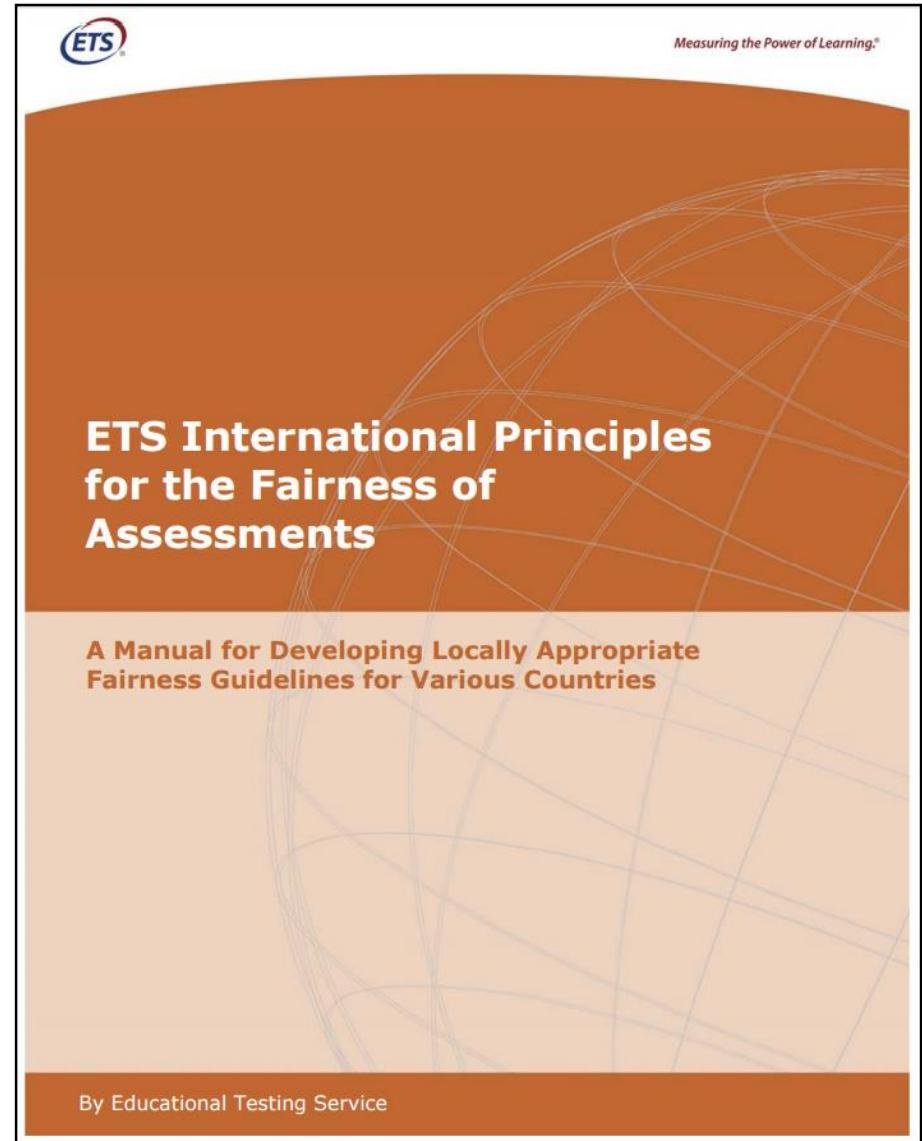
Many people believe that an item that is harder for some group than it is for other groups is not fair for the lower-scoring group. Such items are fair, however, if the differences in difficulty are caused by construct-relevant aspects of the item.

For example: For an item intended to measure reading comprehension in Spanish, score differences between native speakers of Spanish and Spanish-language learners would be fair and valid if the two groups really differed in their ability to comprehend Spanish text.

Application of the Guidelines

ETS Guidelines for Fair Tests and Communications designed for use in the United States and materials used worldwide.

Materials for specific countries will likely require modifications to one or more of the guidelines.



Application of the Guidelines

It is appropriate for ETS to share its opinions about fairness with its clients.

If, however, a client disagrees with some aspects of the guidelines, the fairness requirements of the client should be followed unless they would result in a violation of fairness that ETS could not accept in a test made under its auspices.

Example: ETS would not allow construct-irrelevant material that incites hatred or contempt for people on the basis of disability, gender, race, religion, sexual orientation, or similar factors.

Fairness Review Essential Features

- Mandatory for all ETS tests and communications
- Performed by trained reviewers
- Written documentation
- Arbitration in cases of severe disagreements

Note that fairness reviews also should be held for:

- Materials that are reused once every five years
- Materials used for a purpose or population substantively different from that for which they were initially reviewed

Procedures in Research

All materials go through a technical review

- Technical reviewers also check for fairness

Technical reviewer requirements:

- Must have been trained in the Fairness Review process
- Should not have a stake in the material being reviewed

Procedures in Research

If there is a challenge, reviewer must cite specific guidelines that have been violated

- Challenged material cannot be released outside of ETS until the issue has been resolved

Author and reviewer discuss issue(s) and find mutually acceptable resolution

- Resolution and reviewer's acceptance must be documented

Procedures in Research

Steps for handling a continuing dispute:

1. Fairness Review Coordinator – Discusses issue with author and reviewer; provides solution
2. Center for Fairness in Assessment (CFA) in General Counsel's Office – Discusses issue with author and reviewer; provides solution
3. Three-member subset of the Fairness Review Steering Committee – Discusses issue with author and reviewer; provides solution
4. General Counsel – Reviews documentation; issues ruling

What is the Job of a Reviewer?

- Learn the guidelines
- Use common sense
- Avoid extremism

Important to remember:

Judgments are made from the point of view of the test taker or reader

Questions

ETS Guidelines for Fair Tests and Communications

- Guidelines Regarding:
 - Construct-Irrelevant Cognitive Barriers
 - Construct-Irrelevant Affective Barriers
 - Construct-Irrelevant Physical Barriers
 - Appropriate Terminology for Groups
 - Representation of Diversity

Construct-Irrelevant Cognitive Barriers

ETS communications should be at the appropriate levels of difficulty for the target audiences of those communications

- If construct-irrelevant knowledge or skill is required to answer an item and the knowledge or skills is not equally distributed across groups, then the fairness of the item is diminished

What to Avoid / What to be Aware of

- Unnecessarily Difficult Language
- Specialized Knowledge / Unfamiliar Topics
 - Do not assume audience has specialized knowledge in subjects such as agriculture, computers, finance, law, machinery, politics, science, sports, technology, tools, or transportation
- Inappropriate Contexts
 - It is not appropriate to assume that all test takers have had the same experiences.
- Regionalisms
- Religion

What to Avoid / What to be Aware of

- United States Culture
 - Best not to require knowledge of topics specific to the United States, including product brands, customs, geography, government, history, holidays, institutions, laws, measurement systems, plants, politicians, political systems, public figures, slang, sports, television shows, wildlife found only in the United States

Construct-Irrelevant Affective Barriers

Avoid having language and images that are unnecessarily contemptuous, exclusionary, insulting, or the like in ETS test and communications

- Emotional reactions caused by inappropriate, construct-irrelevant test content that may affect people in different groups in different ways

Issues of Concern

- Accidents, Illnesses, or Natural Disasters
- Advocacy
 - Items and stimulus should be neutral and balanced whenever possible
 - Test takers who have opposing views may be disadvantaged by the need to set aside their beliefs to respond in accordance with the point of view taken in the stimulus material
 - Communications other than tests may advocate for those causes on which ETS has taken a position

Issues of Concern

- Biographical Passages
 - Avoid passages that focus on individuals who are readily associated with offensive or controversial topics, unless important for valid measurement
- Group Differences
 - Do not state or imply that any groups are superior or inferior to other groups with respect to such traits as caring for others, courage, honesty, trustworthiness, physical attractiveness, and quality of culture
- Humor, Irony, and Satire

Issues of Concern

- Images
 - Avoid images that depict content described in other guidelines as material to avoid
 - Avoid construct-irrelevant images of objects or actions that have become controversial or offensive themselves
- Luxuries
- Personal Questions
- Profanity
- Religion

Issues of Concern

- Sexual Behavior
- Slavery
- Stereotypes
- Substance Abuse
- Suicide or Other Self-Destructive Behavior
- Unstated Assumptions
 - Avoid material based on underlying assumptions that are false or that would be inappropriate if the assumptions had been stated
- Violence and Suffering

Construct-Irrelevant Physical Barriers

ETS communications should avoid aspects that interfere with audience's ability to attend to, see, hear, or otherwise sense the material

What to Avoid

- Unnecessary drawings of three-dimensional solids
- Images more complex, cluttered, or crowded than necessary
- Fine distinctions in shading or color to indicate important differences
- Text that does not contrast with the background
- Lines of text that are vertical, slanted, curved, or any other than horizontal
- Hard-to-read fonts
- Uppercase and lowercase versions of the same letter used to identify different things in the same image or segment of text

What to Avoid

- Images in the middle of paragraphs
- Letters that sound alike (e.g., s, x) used as labels for different things in the same image or segment of text
- Letters that look alike (e.g., O, Q) used as labels for different things in the same image or segment of text
- Special symbols or non-English alphabets
- Numbers 1-10 and letters A-J used as labels in the same image or segment of text (because the same symbols are used for those numbers and letters in braille)

Appropriate Terminology for Groups

- Avoid using unnecessary labels for people
 - If group identification is necessary, use the terms a people call themselves whenever possible and note: preferred terms change
- Use group names as adjectives rather than nouns, especially for first references
- Recognize diversity within groups and refer to individual groups whenever possible rather than to use broadly inclusive terms
 - Cuban American or Mexican American students as opposed to Hispanic American students

Appropriate Terminology for Groups

How to refer to people by population group
(Terms should begin with an uppercase letter)

- Black or African American
- Asian American, Pacific Island American, Asian/Pacific Island American
- Hispanic American, Latino American, Latina American
- Native American, American Indian
- White, Caucasian, European American

Appropriate Terminology for Groups

How to refer to people with disabilities

- Refer to the person first, disabling condition second
- “Disabled” preferred over “handicapped”
- Avoid terms that have negative connotations or that reinforce judgments (e.g., “crippled,” “inflicted,” “victim,” “unfortunate,” “confined”)
- Avoid patronizing terms (e.g. “special,” “inspirational,” “courageous,” “overcoming a disability”)
- Avoid references such as “invalid,” “patient,” “sickly”

Appropriate Terminology for Groups

How to refer to people over the age of 65:

- Use a specific age: “people age 65 and above”
- Use “older” as an adjective: “older citizens”
- Avoid “elderly” as a noun
- Minimize the use of euphemisms such as “senior citizens” or “seniors”
- Do not categorize older people as sickly, fragile, or sedentary
- Avoid stereotypical language

Appropriate Terminology for Groups

How to refer to gender differences

- Refer to women and men in parallel terms (“ladies and gentlemen,” “daughters and sons,” “wives and husbands”)
- Females and males over the age of 18 are “women” and “men”
- Do not use the generic “he,” “she,” or “man” to refer to all people
- Generic terms (e.g., “doctor,” “firefighter,” “poet,” “executive”) include both men and women

Appropriate Terminology for Groups

How to refer to people who are gay, lesbian, bisexual, or transgendered

- Avoid issues of sexuality unless required
- “Gay,” “lesbian,” “bisexual,” or “transgendered” are all acceptable
 - Avoid “homosexual” or “queer”
- Use “sexual orientation” rather than “sexual preference”

Representation of Diversity

Strive for gender, racial, and ethnic balance in the depiction of people in tests and publications whenever possible

- Whenever possible, the status of the groups depicted should be equivalent

Questions

The guidelines are available on ETS's external site so they can be shared with clients and colleagues outside of the organization.

The direct link to the Fairness page is:

<http://www.ets.org/about/fairness/>

If you have any questions, feel free to reach out to me:

Bill Monaghan, wmonaghan@ets.org

Example

When President Woodrow Wilson promised to make the world “safe for democracy,” his words stirred the emotions of all Americans, white and black. Black spokesmen predicted a landmark event in black American history. Such hope was based on the premise that black men freely fighting and dying for their country would force the white majority in the United States to finally bestow on blacks all the rights and privileges of citizenship.

According to the passage, black Americans considered fighting in the First World War

- (A) a burden
- (B) a mystery
- (C) a duty
- (D) an opportunity

Issues in the Item

When President Woodrow Wilson promised to make the world “safe for democracy,” his words stirred the emotions of all Americans, white and black. Black spokesmen predicted a landmark event in black American history. Such hope was based on the premise that black men freely fighting and dying for their country would force the white majority in the United States to finally bestow on blacks all the rights and privileges of citizenship.

According to the passage, black Americans considered fighting in the First World War

- (A) a burden
- (B) a mystery
- (C) a duty
- (D) an opportunity

Corrected Item

When President Woodrow Wilson promised to make the world “safe for democracy,” his words stirred the emotions of all Americans. Black community leaders predicted a landmark event in American history. Such hope was based on the premise that Black men freely fighting and dying for their country would force the United States to finally bestow on Black Americans all the rights and privileges of citizenship.

According to the passage, many Black Americans considered fighting in the First World War

- (A) a burden
- (B) a mystery
- (C) a duty
- (D) an opportunity

Proposal Development Code

Project: 01591 – Proposal Preparation and Support

Product: 00021

Activity: 001

Resource: 35

End of Presentation