2009 | 2010

Fairness Report

for the ACT® Tests

ACT®

2009 | 2010

Fairness Report

for the ACT® Tests

ACT endorses the *Code of Fair Testing Practices in Education* and the *Code of Professional Responsibilities in Educational Measurement,* guides to the conduct of those involved in educational testing. ACT is committed to ensuring that each of its testing programs upholds the guidelines in each Code. A copy of each Code may be obtained free of charge from ACT Customer Services (68), P.O. Box 1008, Iowa City, IA 52243-1008, 319/337-1429.

Visit ACT's website at: **www.act.org**

© 2010 by ACT, Inc. All rights reserved.

15813

Contents

Introduction 1

Developmental Procedures for the ACT Tests 2
- Test Specifications 2
- Item Development 3
  - Item Writing 3
  - Item Reviewing 4
  - Pretesting of Items 5
  - Operational Forms Construction 5

Test Form Analyses 11
- Equating 11
- Differential Item Functioning (DIF) 11

Conclusion 13

Appendix
- Guidelines for Group Representation 14
- Guidelines for Gender Representation 16

Introduction

The purpose of this report is to describe the procedures ACT followed when preparing the multiple-choice ACT® test forms that were administered in 2009–2010 to help ensure that these tests are as fair as possible to all examinees who take them. It is ACT's goal to accurately assess what students can do with what they know in the content areas covered by ACT's testing programs. If we were to allow factors other than the academic skills and knowledge in those content areas to intrude, we would provide a less accurate picture of what students know and can do and would risk subjecting students to situations in which their performance might be adversely affected by language or contexts that are perceived to be unfair. ACT is therefore committed to fairness both in principle and in the interest of accuracy in all its tests.

As a testing organization, ACT endorses the *Code of Fair Testing Practices in Education,* a statement of guidelines for those who develop, administer, and use educational tests and data. The *Code* sets forth criteria for fairness in four areas: developing and selecting appropriate tests, administering and scoring tests, reporting and interpreting test results, and informing test takers. According to the *Code,* test developers should provide "tests that are fair to all test takers regardless of age, gender, disability, race, ethnicity, national origin, religion, sexual orientation, linguistic background, or other personal characteristics." Test developers should "avoid potentially insensitive content or language," and "evaluate the evidence to ensure that differences in performance are related to the skills being assessed." ACT is committed to ensuring that each of its testing programs upholds the *Code*'s standards for appropriate test development practice and use.

ACT also endorses the *Code of Professional Responsibilities in Educational Measurement*, which numbers among test developers' responsibilities "to develop assessment products and services that are as free as possible from bias due to characteristics irrelevant to the construct being measured, such as gender, ethnicity, race, socioeconomic status, disability, religion, age, or national origin."

To ensure fairness in a test is a critically important goal. Unfairness must be detected, eliminated, and prevented at all stages of test development, test administration, and test scoring. The work of ensuring test fairness starts with the design of the test and test specifications. It then continues through every stage of the test development process, including item (test question) writing and review, item pretesting, item selection and forms construction, and forms review. ACT makes every effort to see that all ACT tests are fair to the populations for which the tests are intended.

Developmental Procedures for the ACT Tests

Carefully designed procedures to ensure the fairness of the ACT tests were closely followed in every stage of the developmental process. These procedures are described below.

Test Specifications

Test specifications are the blueprints that are used to construct a test. They are composed of two basic parts: (1) a description of both the content and the cognitive level of the skills to be measured by the test, and (2) a description of the statistical characteristics of the items in the test. To make the tests valid and fair, ACT assures that the specifications include only knowledge and skills actually required to meet the purposes of the test. To include anything else in the specifications is to invite various sources of test invalidity, including unfairness, to affect the students' scores. The content specifications of ACT tests describing the specific knowledge and skills selected for evaluation were determined through a detailed analysis of three sources of information. First, the objectives for instruction for grades seven through twelve were obtained for all states in the United States that had published such objectives. Second, textbooks on state-approved lists for courses in grades seven through twelve were reviewed. Third, educators at the secondary and postsecondary levels were consulted to determine the knowledge and skills taught in grades seven through twelve prerequisite to successful performance in postsecondary courses. These three sources of information were analyzed to define a scope and sequence for each of the areas measured by the ACT Program.

While care is taken to ensure that the basic content structure of the ACT tests remains the same from year to year so that the scores are comparable, the specific allocation of test items to the tests is "fine-tuned" on a yearly basis. Each year, consultant panels are convened to review the new forms of the ACT tests in order to verify their content accuracy and the match of test content to both the content emphasis of the secondary curriculum and the prerequisites for postsecondary course work. Further, curriculum surveys are sent periodically to educators throughout the nation to monitor shifts in the secondary school curriculum and in prerequisites for postsecondary course work. Thus, current information is used when items are selected each year for inclusion in the test.

The statistical specifications for the ACT tests indicate the level of item difficulty and the level of item discrimination of the test items to be used. The level of item difficulty for the items in the ACT tests has been selected to help ensure that the tests will effectively measure achievement across a broad range of ability. The

standard of item quality for the ACT tests requires that each item effectively differentiate between students who are well prepared and those who are less well prepared.

Item Development

Item Writing

Once the test specifications are complete, item development can begin. ACT contracts with item writers for the test items used for ACT's tests. Identification and selection of item writers is designed to create a pool of item writers whose particular qualifications fit the content and grade levels of the ACT tests. Most item writers are actively engaged in teaching at the high school or the university level. They represent a variety of schools, from small private institutions to large public ones. ACT makes every attempt to include item writers who represent the diversity of the population of the United States with respect to ethnic background, gender, and geographic location. Broad representation among item writers helps ensure that varying perspectives are included and that members of both genders and various ethnicities are involved in the process and have the opportunity to participate in the development of the test product.

Item writers are provided with a detailed item writer's guide to assist them in developing test materials. The guide includes specific information on fairness concerns, item content and scope, item types, item skill levels, expected item difficulties, and so on. Each item writer, after submitting a sample that meets ACT's specifications, is given an assignment to produce a small number of items, usually 15 to 30. The small size of the assignment ensures that materials on a wide variety of topics will be produced and that the security of the testing program will be maintained, since any item writer will have knowledge of only a very small proportion of the items produced. The item writers work closely with ACT test development associates through correspondence and telephone conversations in producing items of high quality that are designed to meet the test specifications, represent diversity, and be fair to all examinees.

ACT test development associates alert all item writers to issues regarding the way in which various groups of the population are portrayed. It is generally recognized that unless dictated by the test specifications, which are curriculum-based, material likely to be less familiar to one group than another should be excluded from the test content. Other criteria include the degree to which representatives of various groups are depicted in active versus passive circumstances, as exhibiting stereotypic mental or physical characteristics or tendencies, or as engaged in particular occupations or roles. In addition, the

construction of fair test items requires sensitivity to the changing circumstances of our society: increased variation in family structures; the multiethnic composition of the population; and a wide range of socioeconomic and urban, suburban, and rural lifestyles.

ACT test development associates carefully review all test items submitted to ACT. Every new test item and passage is comprehensively reviewed for fairness, interest, and appropriateness to the grade level for which the test is designed; adherence to specifications; soundness and defensibility of passages and items; grammatical accuracy; and sound measurement characteristics. In terms of fairness issues, all items and stimulus materials are reviewed to make sure that they do not contain any language, roles, situations, or contexts that could be considered offensive or demeaning to any population group. Also, no one group should be overrepresented in the pool of items for the ACT; the pool should be balanced in multicultural and gender representation. All groups should be portrayed accurately and fairly, and stereotypes of persons or roles with regard to gender, age, race, ethnicity, religion, physical condition, or geographic origin should be avoided. Presentations of cultural or ethnic differences neither explicitly nor implicitly should rely on stereotypes or make moral judgments. If any problems that detract from the quality of the test passages or items are detected at this early stage, the materials are immediately returned to the item writers for revision. Detailed explanations for the requested revisions are provided to the item writers to help them improve their understanding of basic item-writing principles, techniques, and fairness concerns. Upon receipt of the requested revisions, ACT test development associates reevaluate the items and passages in accordance with the criteria mentioned previously.

Item Reviewing

After items and passages pass the initial screening, they are edited to match ACT's stylistic conventions, content requirements, professional standards, and the requirements of the test forms. As soon as this editing of the 2009–2010 test materials was complete, the materials were sent by secure mail to two groups of external reviewers: content experts (including classroom teachers, college faculty, and curriculum specialists, representing diversity as to geographic region, ethnicity, and gender) and fairness reviewers, who are of diverse ethnicity, gender, and geographic background and are sensitive to issues of test and item fairness. ACT staff consulted with both groups to discuss and redress the reviewers' concerns and then carefully reviewed all the reviewers' comments and revised the test materials accordingly.

Pretesting of Items

Items that passed the item review for fairness, content, reflection of the curriculum, and grade-level appropriateness were pretested on a representative sample of the ACT examinee population in a national administration of the ACT. The purpose of pretesting items is to determine whether the items are technically sound and at the appropriate level of difficulty for the ACT examinee population. Statistical indices of item difficulty and discrimination, among other statistics, were compiled on the basis of pretest results. The items were evaluated according to their performance in the pretest. Those that performed acceptably on all criteria were included in the item pool from which preliminary forms of the ACT tests were constructed.

Operational Forms Construction

Preliminary forms of the ACT tests were constructed using the items that survived the pretest. Items were selected from the item pool to match the requirements of both the content and statistical specifications for the tests. The distributions of the items in each test form were also examined for fairness, variety, diversity, and balance. Each test form should be balanced in multicultural and gender representation. No one group should be overrepresented, nor should any groups be systematically underrepresented. While it is impossible, given the constraint of the limited amount of material in each test form, to represent every group in every form, a good-faith effort to represent diversity should be discernible in every final test form. Two strategies ACT uses to attain this diversity are ensuring the inclusion of culturally diverse passages within each form and ensuring that all passages depict universal themes applicable to all groups.

The preliminary versions of the 2009–2010 test forms were then subjected to several reviews to help ensure that the items were fair and accurate and that the overall test forms conformed to good test construction practices. The first review was performed by ACT's test development associates, editorial staff, and measurement staff. Items were checked for fairness, content accuracy, and conformity to statistical test specifications and to good testing practice. Any problems found in this review were corrected immediately.

The forms were then sent by secure mail to two groups of external reviewers: content experts (including classroom teachers, college faculty, and curriculum specialists, representing diversity as to geographic region, ethnicity, and gender) and fairness reviewers, who are of diverse ethnicity, gender, and geographic background and are sensitive to issues of test and item fairness. These groups of reviewers were not the same individuals who conducted the reviews prior to pretest. ACT staff

consulted with both groups to address the reviewers' concerns. The content review discussion focused on content accuracy, item classifications, skill levels, and grade-level appropriateness of the test items and forms. The fairness review discussion carefully examined any fairness-related issues in the test materials. The comments made by all these external reviewers were collated, and the items/passages identified as problematic replaced as necessary.

To help ensure the fairness of test materials to different examinee groups, ACT selected fairness reviewers from among African American, Asian American, Latino, American Indian, and Women consultants. ACT communicated with prominent, nationally recognized advocacy groups, listed in Table 1, to obtain nominees who could review the ACT test materials. From the recommendations of these groups, ACT selected fairness reviewers who had a history of active participation in promoting the concerns of the group within educational settings and beyond.

Table 1
Advocacy Groups Contacted for Fairness Consultants

Focal Group	Organization
African American	Council of the Great City Schools Florida Endowment Fund for Higher Education Lincoln Institute for Research and Education National Consortium for Graduate Degrees for Minorities in Engineering and Science (GEM) National Association for the Advancement of Colored People (NAACP) National Urban League The William Penn Foundation Quality Education for Minorities Network (QEM)
Asian American	Asian-American Legal Defense and Education Fund Asian Pacific American Heritage Council Japanese American Citizens League Organization of Chinese Americans
Latino	ASPIRA Association AVANCE Congressional Hispanic Caucus Institute League of United Latin American Citizens Ana G. Mendez Educational Foundation Mexican-American Legal Defense and Education Fund National Council of La Raza
American Indian	Association on American Indian Affairs Americans for Indian Opportunity Center of American Indian & Minority Health Indian Law Resource Center Indian Rights Association Maniilaq Association National Congress of American Indians Native American Community Board (NACB) U.S. Senate Select Committee on Indian Affairs
Women	Center for Women Policy Studies Leadership America National Council for Research on Women National Organization for Women (NOW) Women's Research and Education Institute

The fairness consultants reviewed all of the test items and related passages under secure review conditions after being provided with ACT's *Consultant's Guide for the Fairness Review of the ACT EPAS® Tests.* The reviewers attempted to detect any factors in the test forms that might be unfair to any ACT test examinee population group. The fairness consultants were asked to consider the following six major guidelines during their review.

1. **Offensiveness**

   Test materials should not contain any language, roles, situations, or contexts that could be considered offensive, prejudicial, or demeaning to any group.

2. **Fair Portrayal**

   All groups should be portrayed accurately and fairly without reference to stereotypes with regard to gender, age, race, ethnicity, religion, physical/mental condition, socioeconomic status, sexual orientation/identity, or geographic origin. Comparative descriptions of different population group attributes should be avoided. Presentations of cultural or ethnic differences should neither explicitly nor implicitly rely on stereotypes or make moral judgments.

3. **Diversity and Balance**

   Test materials should reflect the diversity of the world, with a particular emphasis on the diversity of the United States. Diverse experiences and points of view should be included in test materials in a manner that avoids problems with other aspects of fairness. For instance, stimuli should avoid ethnocentrism in the cause of promoting diversity.

   A test form, set of forms, or set of tryout units should contain a balanced representation of genders, races, and ethnic groups. No one group should be overrepresented. Test materials can also be offensive by virtue of omission if groups are systematically underrepresented. While it is impossible to represent every group in every set of test materials, a good-faith effort to represent diversity should be apparent. Two strategies to accomplish this diversity would be to ensure the inclusion of culturally diverse passages within each test form and to ensure that passages depict themes applicable to all groups.

   Where applicable, test materials should also be balanced in terms of the gender roles, settings (rural, urban, suburban), geographic regions, and income levels represented.

4. **Fairness in Language**

   The characterization of any group should be respectful. Slang and demeaning characterizations should not be used; representation of people using dialects or non-standard English usage should not be done in a way that belittles them or calls their intelligence into question. References to color, marital status, gender, and other characteristics should only be made when they are germane to the context. To help ensure fairness, gender-neutral terms are frequently used to replace terms that privilege one

gender over another (e.g., *humanity* for *mankind*), but keep in mind that neutralized test materials may sometimes seem bland and uninteresting, as well as forfeit opportunities for introducing multicultural perspectives and experiences into a particular test unit or form.

5. **Curriculum-Based Content and Skills**

No test material should contain elements extraneous to the curriculum-derived content and skills being assessed as a part of the test specifications. Such extraneous material could provide an unfair advantage or disadvantage to population groups. For example, certain groups may have or lack background knowledge about the rules of polo or the nature of a catechism.

More specifically, reviewers should consider the following:

Vocabulary should be suitable for the content area of the test and at a difficulty level appropriate for the age and educational attainment of the students taking the test. Difficult, specialized, or technical terms can be acceptable if the context makes clear their meaning, if they are defined in some way (e.g., in a Reading Test passage's advance organizer), or if they are commonly used in the content area being tested. Words and phrases in languages other than English can be acceptable if their meaning is obvious in context (e.g., *abuela* used as a second reference for "grandmother") or if translations are provided.

Concepts should be equally accessible to all examinee groups. Concepts should either already be known to students at a given age and educational attainment or derivable from the material itself. For example, a Science passage may deal with a topic not typically covered in secondary school science courses that is nonetheless accessible from the information provided in the test.

Experiences mentioned in or necessary to understand the content of stimuli and items should be equally familiar or unfamiliar to all examinee groups. No examinee group should be placed at an advantage or disadvantage due to experience (or lack thereof) with a topic that is not central to the content or skill being measured. For example, while it is fair to assess students on their skill with editing and revising an English Test essay, it would not be appropriate to assess their background knowledge about a topic in such an essay.

6. **Unfair Tests and Examinee Stress**

Consultants must consider the likely emotional reaction of examinees to unfair materials and call to ACT's attention any potential problems. Testing can create a stressful situation for the examinees, so extra care must be taken in the writing and selection of stimuli and items to eliminate any source of unfairness that would likely upset or distract examinees. For example, stimuli should not focus on testing, the college admissions process, or anxieties about going to or paying for college, as these topics could lead some examinees to consider their own situations rather than the test material. Similarly, graphic depictions of diseases or accidents are inappropriate because such depictions might stir up disturbing memories.

To ensure that all groups would be represented fairly, consultants were provided with guidelines for group representation in general and gender representation specifically. These guidelines are presented in the Appendix.

The fairness consultants provided ACT with written comments expressing the results of their individual reviews. All these comments were carefully read and considered by ACT. Any item identified as inappropriate or possibly unfair by any consultant was reviewed by ACT's test development associates and other staff. The consultants' comments and the results of the ACT review, together with recommendations for item acceptance, revision, or deletion, were considered for resolution. ACT's test development associates worked with the fairness and content consultants to prepare the final forms of the test for printing. In all, at least sixteen independent reviews were made of each test item before its appearance on a national form of the ACT, primarily to ensure that each student's level of achievement would be accurately and fairly evaluated.

Test Form Analyses

Equating

For the ACT Program, new test forms are developed each year. Even though each form is constructed to adhere to the same content and statistical specifications to make the forms of the ACT tests equivalent to one another, the forms may vary slightly in difficulty. To control and compensate for these slight differences, a statistical process of equating is used on all forms. The process of equating is used so that the scores reported, called scale scores, have virtually the same meaning regardless of the particular form administered to examinees.

Differential Item Functioning (DIF)

Fairness in the content of the items, however, does not prevent items from functioning statistically in different ways for different population subgroups. (Conversely, it is also true that items identified as "content unfair" in the judgmental review process would not necessarily function in a statistically different way for the population involved.) Differential item functioning (DIF) can be described as a statistical difference between the probability of a specific population group (the "focal" group) getting the item right and a comparison population group (the "base" group) getting the item right *given that both groups have the same level of expertise* with respect to the content being tested. In many instances one group will have, on the average, a higher probability of correctly answering an item, but these differences may be explained by differing levels of expertise between the groups (e.g., members of one group may typically take more rigorous course work or may be advantaged by attending better schools). DIF procedures take these background group differences into account and indicate whether an item may unfairly favor one group over another group.

To detect the existence of differential item functioning (DIF) for items in each test form, the response data from actual national and state administrations of each of the test forms were analyzed. The statistics used for detecting DIF were the standardized difference in proportion-correct (STD) and the Mantel-Haenszel common odds-ratio (MH).

In ACT's experience, the MH and STD procedures are useful statistics in detecting DIF. Both techniques are designed for use with multiple-choice items, and both require data from significant numbers of examinees to provide reliable results. For a description of these statistics and their performance overall in detecting DIF, see ACT Research Report No. 89-7, entitled *Performance of Three Conditional DIF Statistics in Detecting Differential Item Functioning on Simulated Tests,* by Judith A. Spray (1989).

After each national and state administration of a test form, large random samples representing examinee groups of interest, e.g., males and females, were selected from the total number of examinees taking the test. The groups compared were African Americans/Caucasians, Mexican Americans/Caucasians, Hispanics (other than Mexican origin)/Caucasians, Asian Americans or Pacific Islanders/ Caucasians, American Indians or Alaska Natives/Caucasians, and males/females. The samples of examinees' responses to each item on the test were analyzed using the STD and MH procedures. All items with MH and/or STD values exceeding a preestablished statistical tolerance level were flagged for further review. It should be noted that the act of flagging an item does not mean the item is necessarily unfair. Some items will be flagged as favoring one group over another simply by chance alone due to random fluctuations in population samples.

A total of 120 out of the 10,320 comparisons made on all the ACT items administered in 2009–2010 were flagged and further reviewed by content and measurement specialists for possible reasons for the unusual MH and/or STD results. The result of the review of the 120 items in 2009–2010 led to the conclusion that the items were fair and were flagged by chance alone. The fact that fewer items have been flagged by the DIF analyses than would have been expected by chance alone supports this conclusion. A summary of the DIF analysis results for the ACT forms administered in the 2009–2010 testing year is shown in Table 2.

Table 2
Summary of DIF Analysis Results for the ACT Forms Administered in the 2009–2010 Testing Year: Items Judged Not to Be Biased

Test Dates: Sept. 2009, Oct. 2009, Dec. 2009, Feb. 2010, Mar. 2010, April 2010, June 2010					
		Test			
		English	Mathematics	Reading	Science
Number of Comparisons Flagged by Either MH or STD or Both (Total number of comparisons = 10,320) for Favoring:	Males	2	2	2	
	Females			1	1
	African Americans	10		1	
	Caucasians	10	2	3	2
	Mexican Americans	7	2		
	Caucasians	14	1		
	Hispanics (other than Mexican origin)	5			
	Caucasians	10	2	1	
	Asian Americans or Pacific Islanders	12	4		
	Caucasians	16	9		1
	American Indians or Alaska Natives				
	Caucasians				

Conclusion

It has been and continues to be ACT's goal to accurately assess what students can do with what they know in the content areas covered by ACT's testing programs. To the degree that extraneous factors interfere, accuracy declines. Aware of the critical need for fair portrayal, fairness in language, and cultural diversity in its tests, ACT exerts major efforts to ensure those qualities. We offer this description of our attempts to eliminate irrelevant material and to include all the voices of our rich culture to illustrate ACT's commitment to fairness and accuracy in all its tests.

Appendix

Guidelines for Group Representation

1. All ethnic, racial, class, language, religious, and ability groups should be represented with dignity and respect. References to any group, whether implied or stated, must not be demeaning.

2. No jokes or clichés at any group's expense should appear.

3. Stereotypical references should not be made to any persons or to the characteristics or roles of any groups.

4. No group should be portrayed as superior to another in any way (in such terms as ability, culture, or achievement).

5. Portrayals of patronizing behaviors by some groups, and characterizations of other groups as dependent or helpless, must be avoided.

6. The achievements of members of groups throughout history should be portrayed accurately and positively. Such accomplishments should not be portrayed as unusual. The contributions should be portrayed as active rather than passive. No group should be characterized as inferior or subordinate to another group. When historical inequities are portrayed, there should be an accompanying historical context for the inequity.

7. No slang words should be used to refer to any group's heritage.

8. Modifiers that refer to an individual's race, ethnicity, or skin color (*Puerto Rican businessperson, Italian American actor, Diné rancher, Black lawyer, White journalist*) should be used only if they are germane to the context.

   When referring to groups, the most specific term available should be used. The terms *American Indian* or *Native American* should be avoided when a tribal designation (*Lakota, Cherokee*) is applicable. *Latino/a* should also be avoided if more specific terms such as *Puerto Rican* or *Mexican American* are accurate. The term *Americans* should be avoided when possible as a reference to people in the United States, since the term can accurately refer to people on two continents. (Note, however, that ACT's general policy is to retain such usage in copyrighted Reading Test passages.)

9. Persons with disabilities should not be defined by their physical or mental condition. Terms such as *the blind* and *the deaf* should be avoided (except where historically appropriate) in favor of terms such as *a person who is blind* and *a person who is deaf.*

10. The changing concept of what constitutes a family (multiracial family, only child, single parent, same-sex couples) and the changing roles of family members (working mother, father doing domestic chores) should be respectfully acknowledged.

(continued)

Guidelines for Group Representation *(continued)*

11. No demeaning portrayal of people based on their sexual orientation is to appear.

12. To portray the diversity of cultures in U.S. society, customs, events, beliefs, language, and religious practices unique to groups may be described as long as the descriptions do not hamper the examinee's ability to understand the stimulus or items and do not otherwise detract from the curriculum-based content and skills being measured.

Cultural differences should be respected and portrayed in a positive light. Such terms as *subgroup, minority/majority group, underdeveloped nation, culturally disadvantaged, deprived, primitive,* and *backward country* should be avoided.

Applicable primarily to sets of materials in the aggregate:

13. Groups should be portrayed primarily in active roles.

14. No group should be stereotyped in its behaviors or attitudes; rather, the members of all groups should be portrayed as having a range of behaviors and attitudes (e.g., self-esteem, ambition, anger).

15. No group should be portrayed merely in traditional, stereotypical activities.

16. All groups should be involved in various types of occupations and professions. Stereotypical characterizations of groups should be avoided, and groups should exhibit a wide range of aspirations, professions, and achievements.

17. All groups should be portrayed across the various types of community settings: urban, suburban, and rural.

18. All groups should be portrayed in a wide range of socioeconomic classes. No particular group should be identified solely with one class.

19. Although a person's name does not reliably identify that person's race/ethnicity, geographic origin, status, or even gender, names can be used in test materials to help examinees feel included. The balance of character and place names should represent the diversity in the United States and the world and ensure that no group is favored in representation. The use of character names should not give the impression that people act exclusively within their own groups (e.g., Latino/a characters interacting only with other Latino/a characters).

Guidelines for Gender Representation

1. Both women and men (as well as girls and boys) should be portrayed fairly; demeaning stereotypes of women or men and jokes and clichés at women's or men's expense are unacceptable.

2. Gender-neutral terms should generally be used. Inflected (feminine-ending) words such as *poetess* have traditionally connoted a lesser degree of talent, while words such as *mailman* deny women's abilities to perform those roles.

3. Women and men should be described according to the same attribute. For example, one group should not be described by physical qualities while another is described by academic or athletic accomplishments (e.g., *Henry is a handsome brunette* vs. *Ann is a brilliant lawyer*).

4. Women's capabilities should never be described using masculine references (*a man-size job* or *she can handle the job like a man*) or men's capabilities using feminine references (*he throws like a girl*).

5. References to marital status should be avoided unless they are germane to the context. In the Mathematics Test, where characters are created for the sake of story problems, there should be a mixture of titles used with women's names (e.g., *Ms., Mrs.*).

6. The achievements of men and women throughout history should be portrayed accurately and positively. The contributions of women in history should be portrayed as active rather than passive contributions. When historical inequities are portrayed, there should be an accompanying context for the inequity.

Applicable primarily to sets of materials in the aggregate:

7. Women and men should be portrayed primarily in active roles.

8. Women and men should be portrayed as expressing a wide range of emotions. For example, women should not always be portrayed as nurturing and self-sacrificing, nor should men always be portrayed as wise, self-confident, powerful, and productive.

9. Women and men can be portrayed in various types of occupations and professions. Gender-neutral occupational titles should generally be used. Use of pronouns to denote that certain occupations are perceived as being male- or female-dominated should generally be avoided (*doctor, he; nurse, she; scientist, he*).

10. Women and men should not be shown merely in traditional roles (woman as wife and mother, man as breadwinner). Both sexes should exhibit a wide range of activities. The changing nature of families (single parent, same-sex parents) and family roles (father doing domestic chores, working mother) should be respectfully acknowledged.