Smarter Balanced Assessment Consortium:

Bias and Sensitivity Guidelines

Developed by ETS
April 16, 2012

Smarter Balanced Bias and Sensitivity Guidelines

Table of Contents

Introduction 4
- Purpose 4
- Uses 4
- Link to ECD 5

Defining Validity, Bias, Sensitivity, and Fairness 5
- Validity 5
- Bias and sensitivity 5
- Fairness 5

Judgmental and Empirical Evaluations of Fairness 6
- Judgmental evaluations 6
- Empirical evaluations 6
- Combinations of evaluations 7

Proper Balance in Interpreting the *Guidelines* 7

Issues Often Confused with Fairness 7
- Difficulty of items 7
- Overextending guidelines 7

Content and Language that Is Fair to Include in Smarter Balanced Assessments 7
- Fairness of content 8
- Exposure to information 8
- Information in the stimulus 8
- Students with disabilities 8
- English language learners 9

Barriers to Fairness 9
- Barriers related to invalid knowledge 9
- Barriers related to emotional reactions 9
- Barriers related to physical abilities 9

Avoidance of Invalid Knowledge 11
- Regionalisms 11
- Religion 11
- Occupational and technical information 11
- Idioms 11

Topics to Avoid 12

Topics to Be Treated with Care 13
Accidents and natural disasters 13
Advocacy 13
Alcohol, tobacco, and illegal drugs 13
Animals that are frightening to children 13
Antisocial, criminal, or inappropriate behaviors 13
Biographical materials 14
Dancing 14
Dangerous activities 14
Death and dying 14
Evolution 14
Family problems 14
Gambling 14
Holidays and birthdays 15
Homelessness and evictions 15
Immigration 15
Junk food 15
Luxuries 15
Medicines, including diet supplements 15
Obesity and body-image problems 15
Personal questions 15
Religion 16
Serious illnesses 16
Slavery 16
Terrorism, wars, violence, suffering 16
Avoidance of Stereotypes 16
Stereotyped language 16
Stereotyped images 17
Stereotyped social/occupational roles 17
Stereotyped behaviors and characteristics 17
Appropriate Labels for Groups 17
African American people 17
Asian American people 17

People with disabilities 17
Latino/Latina American people 18
Native American people 18
Older people 18
Women and men 18
Representation of Diversity 18
A Final Word 18
References 20
Other Useful References for Fairness in Assessment 21
Appendix 22
Examples of Acceptable and Unacceptable Test Materials 22
Math problems 22
Excerpts from stimuli for English language arts 24
English language arts items 27

Smarter Balanced Assessment Consortium: Bias and Sensitivity Guidelines

Introduction

Purpose

The purpose of the *Smarter Balanced Assessment Consortium Bias and Sensitivity Guidelines* (hereafter referred to as "the *Guidelines*") is to help ensure that the Smarter Balanced assessments are fair for all groups of test takers, despite differences in characteristics including, but not limited to, disability status, ethnic group, gender, regional background, native language, race, religion, sexual orientation, and socioeconomic status.

The goal of fairness in assessment can be approached by ensuring that test materials are as free as possible of unnecessary barriers to the success of diverse groups of test takers. Those unnecessary barriers can be reduced by following some fundamental rules:

- Do not measure irrelevant knowledge or skill.
- Do not anger, offend, upset, or otherwise distract test takers.
- Treat all groups of people with appropriate respect in test materials.

This document describes in detail how to follow these rules for the Smarter Balanced assessments of the Common Core State Standards in English language arts and mathematics. Some aspects of the *Guidelines*, such as the rule dealing with evolution, are not appropriate for tests of specific subjects such as biology or psychology.

Uses

Though many people think of bias and sensitivity guidelines as applying primarily to the review of test items after they have been written, fairness must be considered in all phases of test development and use.

The intended use of the *Guidelines* is in the development of the Smarter Balanced assessments, particularly in item writing and review. This document describes the rules agreed upon by the Smarter Balanced Assessment Consortium states for achieving fairness in test content and will help reduce subjectivity in evaluating test items for fairness. Only items that are in compliance with the *Guidelines* will be included in the Smarter Balanced assessments. The *Guidelines* will help ensure

that the test content is fair for test takers as well as acceptable to the many stakeholders and constituent groups within the Smarter Balanced states.[1]

Use of the *Guidelines* will help the Smarter Balanced assessments comply with Standard 7.4 of the *Standards for Educational and Psychological Testing*:

> Test developers should strive to identify and eliminate language, symbols, words, phrases, and contents that are generally regarded as offensive by members of racial, ethnic, gender, or other groups except when judged to be necessary for adequate representation of the domain (AERA, APA, & NCME, 1999, p. 82).

Link to ECD

The Smarter Balanced assessments are developed using the principles of Evidence-Centered Design (ECD). Three basic elements of ECD are: 1) stating the claims to be made about test takers, 2) deciding what evidence is required to support the claims, and 3) administering test items that provide the required evidence (Mislevy, Steinberg, & Almond, 1999). ECD provides a chain of evidence-based reasoning linking test performance to the claims to be made about test takers. Fair assessments are essential to the use of ECD. If the items are not fair, then the evidence they provide means different things for different groups of test takers. Under those circumstances, the claims cannot be equally well-supported for all test takers. Appropriate use of the *Guidelines* helps to ensure that the evidence provided by the items means the same thing for various groups of test takers and allows ECD to work as intended.

Defining Validity, Bias, Sensitivity, and Fairness

Validity

To define "fairness" and "bias" for the purposes of the *Guidelines*, it is necessary to understand the meaning of "validity." Validity is the extent to which the inferences and actions made on the basis of test scores are appropriate and backed by evidence (Messick, 1989). More simply, validity can be thought of as the extent to which test scores accurately reflect the relevant knowledge and skills of test takers. For the Smarter Balanced assessments, the relevant knowledge and skills are defined by the Common Core State Standards.

Bias and sensitivity

According to the *Standards for Educational and Psychological Testing*, "bias in tests and testing refers to construct-irrelevant [i.e., invalid] components that result in systematically lower or higher scores for identifiable groups of examinees" (AERA, APA, & NCME, 1999, p. 76). "Sensitivity" is used to refer to an awareness of the need to avoid bias in assessment. In common usage, reviews of tests for bias and sensitivity are reviews to help ensure that the test items and stimuli are fair for various groups of test takers.

Fairness

"Fairness" is a more difficult word to define because, as indicated in the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999, p. 74), "*fairness* is used in many different

[1] The Consortium is committed to addressing concerns that have been raised regarding some special populations (including indigenous populations and Native Americans) and will continue to collaborate on finding practicable resolutions to the issues.

ways and has no single technical meaning." For an extensive discussion of the meanings of "fairness" in assessment, see Camilli, 2006. A useful definition of fairness for the purposes of the *Guidelines* is the extent to which the test scores are valid for different groups of test takers. For example, a math item may contain difficult language unrelated to mathematics. If the language interfered about equally with all test takers, validity would be reduced for all test takers, but the item would not necessarily be unfair. If, however, the language were a bigger barrier for students who are not native speakers of English than for other students, then the item would be unfair.

Even if the items are more difficult for some groups of students than for other groups of students, the items may not necessarily be unfair. For example, if an item were intended to measure the ability to comprehend a reading passage in English, score differences between groups based on real differences in comprehension of English would be valid and, therefore, fair. As Cole and Zieky (2001, p. 375) noted, "If the members of the measurement community currently agree on any aspect of fairness, it is that score differences alone are not proof of bias."

Fairness does not require that all groups have the same average scores. Fairness requires any existing differences in scores to be valid. An item would be unfair if the source of the difficulty were not a valid aspect of the item. For example, an item would be unfair if members of a group of test takers were distracted by an aspect of the item that they found highly offensive. But if the difference in difficulty reflected real and relevant differences in the group's level of mastery of the tested Common Core State Standards, the item would be fair.

Judgmental and Empirical Evaluations of Fairness

Judgmental evaluations

No quantitative indicator of fairness can replace human judgment in evaluations of fairness. Issues that may affect fairness are often too subtle to be captured by any statistic. Therefore, the first line of defense against the inclusion of unfair materials in the Smarter Balanced assessments is the judgment of trained test developers who follow these *Guidelines.* Judges may, however, miss potential fairness issues that sophisticated statistical analyses later find (Bond, 1993). Therefore, both judgmental and statistical evaluations of fairness are required.

Empirical evaluations

In addition to judgmental reviews for fairness, items in the Smarter Balanced assessments receive an empirical check for fairness. Items in the assessments are field tested to see how well they work before they are used to evaluate students. At that stage, a statistic called Differential Item Functioning (DIF) is used as a statistical indicator of fairness. DIF studies are required by Standard 7.3 of the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999, p. 81).

Merely calculating differences between groups in the difficulty of items is not a useful indicator of fairness because the differences may be valid. That is, the groups may actually differ on relevant knowledge that is supposed to be measured by the item. Though the statistics involved may be considered complicated, DIF is based on the straightforward concept that people who know the same amount about a subject should perform similarly on test items concerning that subject, regardless of differences in such factors as gender or race. Deciding which people know the same amount is most often done on the basis of test scores. People with the same or very similar scores are considered to be "matched." Significant differences in the difficulties of test items for matched people in different groups result in elevated values of the DIF statistic.

DIF alone, however, is not proof of bias. No test is perfect. Therefore, no matching of test takers on the basis of test scores can be perfect. A fair item may show DIF merely because the test scores

have not matched people well on the particular knowledge or skill validly measured by the item. (See Holland & Thayer, 1988; Dorans, 1989; and Zieky, 1993, for more information about DIF and its uses in test development.)

Combinations of evaluations

Neither DIF nor any other statistic can be considered proof that an item is either fair or biased, but appropriate statistics can help identify any potentially unfair items. Using a combination of judgmental reviews and statistical data on the performances of matched test takers in different groups is the best method available to help ensure the fairness of the Smarter Balanced assessments.

Proper Balance in Interpreting the *Guidelines*

An overly zealous interpretation of the *Guidelines* can be as harmful as an overly lax interpretation. Critics have responded very negatively to what they consider the excesses of some bias and sensitivity reviews. For example, with respect to what she considered rampant over-generalizations of bias and sensitivity guidelines, Diane Ravitch (2003, p. 4) wrote, "What began with admirable intentions has evolved into an increasingly broad and increasingly bizarre policy of censorship that has gone far beyond its original scope."

The position of the Consortium and other professionals in the field of measurement is that, when done appropriately, bias and sensitivity reviews are not a "bizarre policy of censorship" but are instead a sincere attempt to make fair and valid tests. It is not true that anything that anybody might possibly object to in test materials is necessarily judged to be unfair. In fact, one purpose of using written guidelines such as the ones in this document is to limit overly broad, idiosyncratic judgments about the fairness of test materials. There must be a balance between striving to ensure fairness and the ability to measure the full range of the Common Core State Standards with authentic and interesting materials. Proper use of the *Guidelines* will help maintain that balance.

Issues Often Confused with Fairness

Difficulty of items

The difficulty of items should not be confused with fairness. A difficult item is not necessarily unfair if the sources of difficulty are valid. For example, the Common Core State Standards call for students to read documents important in the history of the United States. Some of those documents are difficult to read, and valid items will appropriately reflect that difficulty.

Overextending guidelines

Reviewers should avoid overextending the *Guidelines* to contrive situations in which an innocuous topic is judged to be unfair. That practice inappropriately limits test content because any topic can be judged to be potentially upsetting in some set of circumstances for some test takers. For example, a reviewer might say that an innocuous depiction of a mother with her child might upset a test taker who is an orphan. A topic that is upsetting on its face is probably unfair, but an innocuous topic that might possibly be upsetting for some atypical test taker(s) under some particular set of circumstances is not necessarily unfair.

Content and Language that Is Fair to Include in Smarter Balanced Assessments

Fairness of content

With respect to the validity and fairness of the Smarter Balanced assessments, any content that is required by the Common Core State Standards, consistent with the Consortium's item specifications guidelines and reviewed through the processes described by the Consortium's item review procedures, may be included in the Smarter Balanced assessments. In addition, state laws and state policies in one or more of the states may also affect the content of the assessments that states present to their students. For example, a state law may require the inclusion of content based on the achievements of specific groups. Any content required by state law or state policy may be administered during the same test event as a Smarter Balanced assessment but will need to conform to the Consortium's policies regarding the addition of state-specific content. These policies may include, at a minimum, the ability to derive a score that excludes the additional state content such that a comparable score may be reported for all students taking a Smarter Balanced assessment.

Exposure to information

Stimuli for English language arts items have to be about some topic. Mathematics problems are often placed in real-world contexts. The Common Core State Standards cover only mathematics and English language arts. They do not include all of the content areas from which topics and contexts must be drawn. Which topics and contexts are fair to include in the Smarter Balanced assessments? One fairness concern is that students differ in exposure to information through their life experiences outside of school. For example, some students experience snow every winter, and some have never experienced snow. Some students swim in the ocean every summer, and some have never seen an ocean. Some students live in houses, some live in apartments, some live in mobile homes, and some are homeless.

Even though curricula differ, the concepts to which students are exposed in school tend to be much more similar than are their life experiences outside of school. If students have become familiar with concepts through exposure to them in the classroom, the use of those concepts as topics and contexts in test materials is fair, even if some students have not been exposed to the concepts through their life experiences. For example, a student in grade 4 should know what an ocean is through classroom exposure to the concept, even if he or she has never actually seen an ocean. A student does not have to live in a house to know what a house is, if there has been classroom exposure to the term. Similarly, a student does not have to be able to run in a race to know what a race is. Mention of snow does not make an item unacceptable for students living in warmer parts of the country if they have been exposed to the concept of snow in school.

Information in the stimulus

A major purpose of reading is to learn about new things. Therefore, it is fair to include material that may be unfamiliar to students if the information necessary to answer the items is included in the tested material. For example, it is fair to test the ability of a student who has never been in a desert to comprehend an appropriate reading passage about a desert, as long as the information about deserts needed to respond to the items is found in the passage.

Students with disabilities

A similar issue arises for students with disabilities. Is it fair to include material about the visual arts or music for students who cannot experience them directly? Is it fair to include passages about physical activities for students who cannot participate in them? As noted above, it is acceptable to include material that may be unfamiliar to some students through life experiences, as long as the information necessary to answer the items is included in the stimuli or is part of the information

expected from classroom exposure. For students with certain disabilities, it is necessary to add the provision that the information necessary to answer the items does not need to be obtained through direct, personal experience. For example, a high school student who is deaf could fairly be expected to know what a bell is, but could not fairly be expected to know what a bell sounds like.

English language learners

Unnecessarily difficult language can be a source of unfairness when the language itself is not the focus of measurement. This is particularly true for English language learners. The language in mathematics items should not be a barrier to a correct answer for people who could do the required mathematics. For mathematics assessments, therefore, nonmathematical language should be targeted no higher than the grade level below the tested grade level. For English language arts (ELA) assessments, the focus is on the use of language. Valid assessment of the ELA Common Core State Standards requires the use of language targeted at the tested grade level. In general, when the language itself is not being tested, the clearest language consistent with validity should be used. For more detail on appropriate language for English language learners, see the *Smarter Balanced Assessment Consortium Accessibility Guidelines for English Language Learners* (2012).

Barriers to Fairness

Fairness review is intended to remove barriers to valid measurement that may affect different groups of test takers in different ways. There are three major types of barriers related to fairness.

Barriers related to invalid knowledge

Barriers related to invalid knowledge occur when uncommon information—not reasonably expected of some group(s) of students and not related to the Common Core State Standards—is required to answer a test item. For example, assuming a student knows what a “foyer” is would be unfair because the term: 1) is more likely to be known by some groups of students than by other groups of students, 2) is not required by the Common Core State Standards, and 3) is not likely to have been routinely used in the classroom.

Barriers related to emotional reactions

Barriers related to emotional reactions may occur if language or images cause strong emotional reactions among members of some groups of test takers and those reactions potentially interfere with test performance. For example, if a passage advocates for one position of a controversial issue such as gun control, a student who is a strong supporter of the opposite position may be disadvantaged by having to put his or her beliefs aside to respond correctly to items.

Even if the performances of students are not directly affected, the presence of offensive, inflammatory, controversial, upsetting, or disrespectful material in tests will lower the confidence of students, parents, politicians, educators, and other community members in the test.

Barriers related to physical abilities

Barriers related to physical abilities occur if test takers have difficulty seeing or hearing the test materials or have physical difficulty responding to the test under standard conditions (e.g., manipulating a computer mouse). The Consortium uses the principles of Universal Design (Thompson, Johnstone, & Thurlow, 2002) to help reduce physical barriers to valid measurement, but some accommodations for students with disabilities and modifications to test materials will still be necessary.

The Consortium has decided that accessibility concerns are best conveyed in a separate document. See the *Smarter Balanced Assessment Consortium General Accessibility Guidelines* (2012) for information about how to ensure that the Smarter Balanced assessments are as accessible as possible for all students, including students with disabilities.

Avoidance of Invalid Knowledge

It is necessary to avoid unfair barriers to success based on group differences in knowledge unrelated to the purpose of the test. Requiring specialized invalid knowledge to answer a test item is unfair. For example, requiring prior knowledge of the number of people on a soccer team to answer an item in the Smarter Balanced assessments is unfair because people who know the valid content may not have the invalid knowledge of soccer needed to answer the item. (Note that testing specialized knowledge is appropriate when that knowledge is valid. Requiring knowledge about the number of people on a soccer team may be perfectly appropriate on a physical education test.) This guideline prohibits the testing of specialized knowledge when that knowledge is not relevant to the purpose of the test. Specialized knowledge that is explained in the stimulus material or can be inferred by contextual clues is acceptable, however, if the understanding of such explanations or the making of such inferences is supposed to be tested.

The following topics have been common sources of specialized invalid knowledge in tests like the Smarter Balanced assessments. Familiarity with these topics and other similarly specialized knowledge—when unrelated to the purpose of the test—should not be required to answer items, unless the necessary information is provided in the stimulus material.

Regionalisms

Avoid requiring knowledge of words and phenomena limited to a region or certain regions of the country and words that carry different meanings in different regions (e.g., “hero” for “sandwich,” “snow days” at school, “tonic” or “pop” for “soda,” “muffler” as an article of clothing, “bubbler” for “water fountain.”)

Religion

Avoid requiring knowledge of any particular religion. For example, to say that something is “as colorful as an Easter egg” may be an unfamiliar comparison for some students.

Occupational and technical information

Avoid requiring knowledge of specialized information and terminology—not related to the purpose of the test—that is associated with a particular occupation or field of knowledge such as agriculture, law, mechanics, military, science, sports, technology, transportation, or weapons. For example, avoid requiring irrelevant knowledge of the purpose of a silo, the less common names for tools, the chain of command in military organizations, the functions of parts of weapons, the scoring systems or rules of play in various sports, the uses of a flange, or the meaning of “lumen.”

The point at which words become too specialized is a matter of judgment and will vary with the grade level of the students who are being tested. The best judges of the appropriateness of words associated with a particular field of knowledge are experienced teachers of students at the tested grade level.

Idioms

Avoid requiring understanding of idioms and figures of speech unless understanding them is called for by the Common Core State Standards, as in some English language arts items (e.g., spill the beans, hit the hay, fly in the ointment, flash in the pan).

Topics to Avoid

Certain topics are extremely controversial, upsetting, inflammatory, and often judged by parents and communities to be inappropriate for children. Such topics should be excluded from the Smarter Balanced assessments unless required to measure the Common Core State Standards. The goal is to avoid material that may cause extreme negative emotions in test takers because such emotions have the potential to interfere with test performance. It is best not to include materials that may cause strong negative emotions such as anger, disgust, fear, hatred, or sadness.

The following list is intended to indicate the nature of topics that should be excluded from Smarter Balanced assessments, but the list is not exhaustive. Current events may add topics that are so problematic that they should be excluded from the assessments. Topics to be avoided include, for example:

- abortion
- abuse of people or animals
- contraception
- deportation of immigrants
- experimentation on people or animals that is dangerous or painful
- killing of animals for sport
- the occult, witches, ghosts, vampires
- pregnancy of human beings
- rape
- sexual behavior or sexual innuendo
- suicide
- torture

Other issues have become so sensitive that the topics are difficult to treat in a way that does not cause fairness problems. It is safest not to include topics such as the following in Smarter Balanced assessments:

- euthanasia
- gun control
- climate change caused by human behavior
- prayer in school
- current or recent partisan political issues, ethnic conflicts, and religious disputes

Topics to Be Treated with Care

Other sensitive but less upsetting topics may be included in Smarter Balanced assessments. Such topics must, however, be treated carefully to minimize potential fairness issues. Guidelines that forbid a topic are easy to apply. Guidelines that require treating a topic with care are more difficult to apply because different people will have different opinions about what is acceptable.

When making judgments about the suitability of materials on topics such as those listed below, it is important to keep in mind that the Smarter Balanced assessments must not only be fair and valid for test takers, they must also appear to be fair and valid in the opinions of various constituencies within the Consortium. It is counterproductive to use test materials that various groups within the Consortium will consider inappropriate for their children.

Accidents and natural disasters

Mention of these topics or general, objective discussions may be acceptable, but avoid a focus on suffering, destruction, or graphic, gruesome details that may upset or frighten students.

Advocacy

The Smarter Balanced assessments should not support one side on a controversial issue unless it is necessary to do so for validity, as in presenting an argument for test takers to evaluate. Avoid advocacy when possible because test takers with opposite views may be disadvantaged. If, however, advocacy is required to measure a Common Core State Standard, indicate that the material does not necessarily represent the views of the Consortium. Do not use brand names to avoid the impression that the Consortium is advocating use of a particular brand. Avoid advocating for or against a political party unless doing so is important to measure a Common Core State Standard.

Alcohol, tobacco, and illegal drugs

The goal is to avoid giving any impression of approval of these substances. It is best to avoid depictions of people using these substances. In particular, do not depict use of these substances as pleasurable, alluring, or as signs of sophistication and maturity. Warnings against the use of these substances may be acceptable for students in middle or high school.

Animals that are frightening to children

Younger students are more likely to be upset by certain dangerous animals than are older students. Depictions of spiders and poisonous snakes have been cited as causing problems for some children and are best avoided. Objective depictions of a food chain or nonthreatening descriptions of animals are acceptable, but avoid depicting predators engaged in violent, threatening behavior. For example, a discussion of how members of a wolf pack interact with each other is likely to be acceptable. A depiction of a wolf ripping the entrails from a fawn or attacking a child should be avoided.

Antisocial, criminal, or inappropriate behaviors

(e.g., bullying, cheating, cutting school, joining gangs, fighting, lying, stealing). One goal is to avoid modeling inappropriate or bad behavior for students. It is particularly important to avoid making such behavior appear to be attractive, fun, glamorous, sophisticated, or something to be emulated. Another goal is to avoid upsetting students who may have been the victims of such behavior by others.

Biographical materials

Take care in selecting biographical materials. Some biographical materials may be controversial because different groups of people may view the individuals depicted very differently. For example, one group's heroic freedom fighter is another group's cowardly terrorist. A possible concern with the use of biographical material about living people is that a person who is widely admired at the time he or she is included in test materials may become involved in a highly publicized scandal before the test is administered.

Dancing

Allow all forms of dance except couples social dancing, which is the type most likely to draw criticism from some groups.

Dangerous activities

The goal is to avoid modeling behaviors that are inherently dangerous and making dangerous behaviors appear to be attractive, fun, glamorous, or something to be emulated. Particularly for younger children, avoid showing potentially dangerous behavior such as running away from home, going with strangers, or using dangerous tools or weapons without supervision, even if all turns out well. Common actions that are dangerous if done improperly (such as crossing the street, riding a bicycle, hiking, or swimming) are acceptable if depicted as being done properly. Describing dangerous substances or devices such as weapons, poisons, or explosives in ways that make them appear attractive or safe is not acceptable.

Death and dying

Detailed depictions of the death of parents, siblings, contemporaries, and family pets should be avoided unless necessary to measure a Common Core State standard. It is acceptable to mention death (e.g., Dr. Martin Luther King, Jr., died in 1968), but it is not acceptable to depict gruesome details.

Evolution

Evolution of human beings or similarity of human beings to other primates should be avoided as highly controversial for some groups. Evolution within a species (such as evolution of bacteria to withstand antibiotics) is much less problematic and could be allowed if treated with care. Fossils and the age of Earth are acceptable if not linked to evolution of human beings. (In tests intended to measure knowledge of science, any aspect of evolution required for validity is acceptable.)

Family problems

The goal is to avoid upsetting test takers with detailed descriptions of serious family problems such as the loss of a job, divorce, or serious illness of a parent or sibling, except as needed in historical or literary materials to measure Common Core State Standards.

Gambling

Instruments used for gambling such as playing cards and dice may be used as required in math problems, but do not assume that all students will be familiar with them and will know such things as the number of cards in a deck or the maximum number obtainable on a pair of dice. Depictions of people gambling for fun or profit should be avoided.

Holidays and birthdays

Not all test takers will be familiar with every religious or quasi-religious holiday (e.g., Halloween). Not all test takers celebrate birthdays. Mention of holidays and birthdays is acceptable as long as all of the information necessary to answer items is included in the stimulus material. The general need to avoid religious materials would argue against extended discussion of religious holidays.

Homelessness and evictions

These topics may be upsetting to students, particularly those who have direct experience with them or fear having a future experience with them. If the inclusion of any of these topics is important to measure a Common Core State Standard, the topic must be treated factually rather than emotionally and must not focus on anguish and distress.

Immigration

Immigration has become very controversial. If the inclusion of the topic is important to measure a Common Core State Standard, the topic must be treated factually and objectively.

Junk food

The goal is to avoid modeling unhealthy behavior by showing excessive consumption of junk food or the selection of junk food in preference to more healthful food. However, it is acceptable to mention eating a cookie, for example, or to use the sharing of a pie to illustrate a fraction.

Luxuries

The goal is to avoid elitism and the impression that ordinary people are excluded from the test materials. However, test materials do not have to be limited to what is affordable by the least affluent of families. Luxuries such as servants, mansions, and yachts should be avoided except as needed in literary or historical materials to measure the Common Core State Standards. Avoid more common luxuries such as ski trips and private tennis lessons. Avoid depicting expenditures that most people would consider excessive. For example, in a math item, do not have a man purchase three suits at $1500 per suit.

Medicines, including diet supplements

Treatments for serious illnesses may be upsetting to some students and should be avoided. Do not model the use of drugs, even prescription drugs, as a way to solve problems. Some groups are opposed to medical treatment, so it is best to avoid the topic unless it is important for the measurement of a Common Core State Standard.

Obesity and body-image problems

The goal is to avoid upsetting children who depart from the norm in height, weight, or other physical attributes with negative depictions of people who depart from the physical norm. A wide range of body types should be represented in any pictorial material, but stereotypes and negative depictions of people with atypical body shapes should be avoided.

Personal questions

Items must not invade the privacy of students by asking them to divulge personal or family issues such as religion, political preference, or antisocial or criminal behavior. For example, do not use an item that asks a test taker to describe a time when he or she was caught doing something wrong. It

is best to avoid constructed-response items that require students to reveal how they would act in situations contrary to their beliefs about appropriate behavior.

Religion

Religion was cited previously as a source of information that is not common to all students. Religion is cited here as a topic best treated with great care in Smarter Balanced assessments. Some people will see even an objective description of a religion as proselytizing. However, it is acceptable to mention religion. For example, mentioning that Buddhism is one of the main religions in Singapore is acceptable. Going into detail about the practices of adherents of Buddhism is not acceptable. In particular, avoid praising or criticizing the practices of a religion. Also avoid references to God, euphemisms for God, or Creationism except in historical or literary documents important for the measurement of the Common Core State Standards.

Serious illnesses

Serious illnesses include mental as well as physical illnesses. Illnesses that primarily affect certain groups, such as some genetic diseases, may be particularly problematic. Mention of serious illnesses may be acceptable, but avoid a focus on suffering or graphic, gruesome details that may upset students.

Slavery

This topic may be included in historical or literary documents if important for the measurement of the Common Core State Standards. A focus on graphic, upsetting aspects of slavery should be avoided.

Terrorism, wars, violence, suffering

These topics may be included in historical or literary documents if important to measure the Common Core State Standards. A focus on graphic, upsetting, or frightening aspects of the topics should be avoided.

Avoidance of Stereotypes

Materials in Smarter Balanced assessments should not reinforce stereotypes. It is acceptable to show traditional behavior (e.g., a woman caring for children), but traditional behaviors must be balanced by depictions of nontraditional behaviors to avoid reinforcing stereotypes. For adaptive tests (assembled by computer as they are administered to a student), balance is best handled at the level of the item pool. To help ensure that the pool is balanced, item writers should produce items showing nontraditional behaviors whenever they produce items showing traditional behaviors that could be considered stereotyped.

All types of stereotypes should be avoided, but the following types have been particularly problematic:

Stereotyped language

Representations of dialect are not acceptable unless in historical or literary material important for the measurement of a Common Core State Standard. Phrases such as “man-sized job,” or “Dutch uncle” should be avoided. Language that uses different terms for the same characteristic in men and women is not acceptable. For example, it is not appropriate to label a man as “forceful” or “assertive” and a woman as “pushy” or “controlling” for exhibiting the same behavior. Language that assumes all members of a profession are one gender is unacceptable (e.g., use “sales representative” instead of “salesman,” “firefighter” instead of “fireman,” “mail carrier” instead of

"mailman"). Some stereotyped language may be acceptable in literary or historical material important for the measurement of a Common Core State Standard.

Stereotyped images

Avoid stereotyped images. For example, do not show all girls in frilly dresses and all boys in jeans. Do not show all White men in suits and ties and all Black men dressed as laborers. If it is impossible to show diversity in a single image, diversity should be shown across images.

Stereotyped social/occupational roles

There should be a mix of genders and races shown in any social or occupational role. For example, do not depict all male doctors with all female nurses. Do not show all Black workers with all White bosses. If it is impossible to show diversity in a single item, diversity should be shown across items.

Stereotyped behaviors and characteristics

Do not treat all members of a gender, sexual orientation, racial, ethnic, national, or other such group as though they all share the same characteristic. For example, do not depict all Native American people as close to nature or all Asian American students as smart. It is particularly important to avoid offensive stereotypes of any such group. For example, do not portray any such group as more (or less) lazy, immoral, primitive, ignorant, prone to crime, gullible, violent, miserly, arrogant, or dirty than any other such group.

Appropriate Labels for Groups

It is very important to avoid derogatory labels for any group. Whenever possible, use the label that the group prefers. History has shown that the following groups have often been mislabeled.

African American people

Use "Black" or "African American." Do not use "Negro" or "Colored" except in the names of institutions, or in historical or literary material important for the measurement of the Common Core State Standards.

Asian American people

When possible, use specific terms such as "Japanese American" or "Chinese American." Terms such as "Pacific Island American," "Native Hawaiian," and "Asian/Pacific Island American" should be used as appropriate. Do not use the word "Oriental" to refer to people except in historical or literary material important for the measurement of the Common Core State Standards.

People with disabilities

Put the person before the disability. For example, use "a person who is blind" rather than "a blind person." In general, avoid using adjectives as nouns for people with disabilities (e.g., "the blind" or "the deaf") except in the names of organizations or in literary or historical material important for measurement of the Common Core State Standards. Avoid euphemisms such as "challenged." Use objective language rather than emotionally loaded terms (e.g., "uses a wheelchair" rather than "confined to a wheelchair"). Do not depict people with disabilities, including people with learning disabilities and people with developmental disabilities, as helpless victims. Terms to be avoided include "dumb" for a person who is mute, "handicapped" for a person with a disability, and "retarded" for a person with a cognitive disability.

Latino/Latina American people

The terms “Latino American” (for men) and “Latina American” (for women) are acceptable. The term “Hispanic American” is also acceptable. When appropriate for the context, it is preferable to use specific group names such as “Cuban American,” “Dominican American,” or “Mexican American.”

Native American people

“Native American” and “American Indian” are both acceptable. When possible, use specific names for peoples such as “Pequot” or “Mohegan.” Some Native Americans prefer the words “nation” or “people” to the word “tribe.”

Older people

It is best to refer to older people by specific ages or age ranges. Minimize the use of euphemisms such as “seniors.”

Women and men

The primary rule is to use parallel terms for men and women. For example, do not use titles for men and first names for women, as in “Dr. Sanchez and his wife, Juanita.” Do not refer to women as “wives” unless men are referred to as “husbands” in the same context. Do not refer to women by physical attributes and to men by accomplishments as in “the successful lawyer and his beautiful wife.” Do not refer to males as “men” and women of similar ages as “girls.” Do not use the generic “he” or “man” to refer to all human beings. Historical or literary material important for the measurement of the Common Core State Standards is acceptable even if it uses outmoded terms and nonparallel language for women and men.

Representation of Diversity

There should be representations of different groups in the pool of items so tests built from the pool will, on average, be appropriately balanced. In items and stimuli that mention people, the following conditions are required in the pool of items and should be reflected in assignments to item writers:

- Males and females should be approximately equally represented.
- People who are members of what are traditionally considered to be minority groups must be represented.
- People of different ages, physical abilities, and social classes should be depicted.
- A wide variety of life situations, living conditions, types of housing, types of families (including single-parent families), regions, and the like should be depicted.

A Final Word

Neither this nor any other set of guidelines can cover all of the possible variations in content that will have to be evaluated for fairness in the Smarter Balanced assessments. Current events (e.g., natural disasters, issues raised during political campaigns, terrorist attacks) can add new topics that may cause fairness problems. Issues that were neutral may become controversial. If the specific guidelines do not offer sufficient guidance in some particular situation, the best practice is to turn to the fundamental rules and ask:

- Do the items measure any irrelevant knowledge or skill? If so, will some group(s) be more greatly affected than others?

- Will any aspect of the test materials anger, offend, upset, or otherwise distract test takers? If so, will some group(s) be more greatly affected than others?
- Do the test materials treat all groups of people with respect? If not, will some group(s) be more greatly offended than others?

If some group(s) will be more greatly affected than others, a potential fairness problem exists. The next step is to determine whether or not the potential problem is a real one. Has difficulty been confused with fairness? Has a guideline been overgeneralized? Has a "treat carefully" guideline been interpreted as a "must avoid" guideline? Has a situation been contrived to make innocuous content seem unfair? Is the material important for valid measurement of the Common Core State Standards?

Finally, it is important to keep in mind that the intent of using the *Guidelines* is to remove unnecessary barriers to the success of diverse groups of test takers. Some potential barriers, such as difficult language in an ELA stimulus based on historical documents, may be necessary to allow valid measurement of the Common Core State Standards and are acceptable.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing.* Washington, DC: American Psychological Association.

Bond, L. (1993). Comments on the O'Neill & McPeek paper. In P. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 277–280). Hillsdale, NJ: Lawrence Erlbaum Associates.

Camilli, G. (2006). Test fairness. In R. L. Brennan (Ed.), *Educational measurement* (pp. 221–256). Washington, DC: American Council on Education/Praeger.

Cole, N. S., & Zieky, M. J. (2001). The new faces of fairness. *Journal of Educational Measurement. 38,* 4.

Dorans, N. (1989). Two new approaches to assessing differential item functioning: standardization and the Mantel-Haenszel method. *Applied Measurement in Education. 2,* 3, pp. 217–233.

Holland, P., & Thayer, D. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer & H. Braun (Eds.), *Test validity.* Hillsdale, NJ: Lawrence Erlbaum.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement.* Washington, DC: American Council on Education.

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (1999). *Evidence-centered assessment design.* Princeton, NJ: Educational Testing Service.

Ravitch, D. (2003). *The language police: How pressure groups restrict what students learn.* New York: Knopf.

Smarter Balanced Assessment Consortium. (2012). *Smarter Balanced Assessment Consortium general accessibility guidelines.* Olympia, WA: Author.

Smarter Balanced Assessment Consortium. (2012). *Smarter Balanced Assessment Consortium accessibility guidelines for English language learners.* Olympia, WA: Author.

Thompson, S. J., Johnstone, C. J., & Thurlow, M. L. (2002). *Universal design applied to large scale assessments.* Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Zieky, M. J. (1993). Practical questions in the use of DIF statistics in test development. In P. Holland & H. Wainer (Eds.), *Differential item functioning.* Hillsdale, NJ: Lawrence Erlbaum.

Other Useful References for Fairness in Assessment

ACT. (2011). *Fairness report for the ACT tests.* Iowa City, IA: Author.

American Institutes for Research. (n.d.). *Standards for language accessibility, bias, and sensitivity*. Washington, DC: Author.

American Psychological Association. (2001). *Publication manual of the American Psychological Association.* Washington, DC: Author.

Data Recognition Corporation. (2003). *Fairness in testing: Guidelines for training bias, fairness and sensitivity issues.* Maple Grove, MI: Author.

ETS. (2009). *ETS guidelines for fairness review of assessments.* Princeton, NJ: Author.

Ramsey, P. (1993). Sensitivity review: The ETS experience as a case study. In P. Holland, & H. Wainer (Eds.), *Differential item functioning.* Hillsdale, NJ: Lawrence Erlbaum.

Zieky, M. J. (2006). Fairness review. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of test development* (pp. 359–376). Mahwah, NJ: Lawrence Erlbaum Associates.

Appendix

Examples of Acceptable and Unacceptable Test Materials

Following are excerpts from test items and stimuli. Some of the excerpts illustrate various violations of the guidelines. Others illustrate items that are acceptable in terms of fairness.

Math problems

The first set of examples consists of math problems or excerpts from math problems. Note the tension between adding realistic context to a math problem and avoiding linguistic complexity and irrelevant knowledge requirements.

1. The drawing below shows a roof truss. Highlight the two triangles that are congruent with each other.
    a. **Unacceptable.** Few children are likely to be familiar with a roof truss, and knowledge of a roof truss is not related to the purpose of the question.
2. In the drawing below, highlight the two triangles that are congruent with each other.
    a. **Acceptable.** The reading load is reduced, and there is no unfamiliar context.
3. Shaquan helps assemble food packages for poor people at Christmas. Each box holds 6 cans in a row. There is room for 4 rows in a box. Write the expression that best describes the number of cans in one full box.
    a. **Unacceptable.** The first sentence adds to the reading load of a math question but adds no useful information. The references to “Christmas” and “poor people” are inappropriate and unnecessary.
4. Two people who were conversing at a street corner parted and moved away from the corner in straight lines that are perpendicular to each other. If one person walked at 3 miles per hour and the second person jogged at 4 miles per hour, how far apart would they be after one hour?
    a. **Unacceptable.** The linguistic load is high for a math question. The sentences are long, and the syntax is complicated. “Conversing” is a difficult synonym for “talking,” and the people’s actions before they started to move are not relevant in any case. (“Perpendicular” is acceptable as a valid mathematical term.)
5. Two people stood next to each other. They started walking in straight lines that are perpendicular to each other. One person walked at 3 miles per hour. The other person walked at 4 miles per hour. How far apart are they after one hour?
    a. **Acceptable.** Unnecessary information has been deleted. Two long sentences have been replaced by five shorter sentences. The conditional syntax (“If one person. . .”) has been replaced by brief statements of fact.
6. A modem can send X bits per second. Write the expression that shows how many seconds it would take to send Y bits.

a. **Unacceptable.** The mention of a "modem" and "bits" is irrelevant and is likely to be unfamiliar. A student might skip the item or waste time wondering what "bits" are or what a "modem" is.

7. Lee can walk X miles an hour. Write the expression that shows how many hours it would take Lee to walk Y miles.

   a. **Acceptable.** The context is familiar.

8. If one card is taken at random from a deck of playing cards, what is the probability that the card will be an ace?

   a. **Unacceptable.** The question assumes knowledge of the number of aces and the total number of cards in a deck of playing cards. It is acceptable to ask about probability, and it is acceptable to use playing cards in math problems. According to the guideline about gambling, however, it is not acceptable to assume that test takers have knowledge of the characteristics of a deck of playing cards.

9. There are 4 aces in a deck of 52 playing cards. If one card is taken at random from the deck, what is the probability that the card will be an ace?

   a. **Acceptable.** No knowledge of the characteristics of a deck of cards is required to answer the item.

10. When Ms. Luna pulled her car into the parking garage, the machine at the gate issued a ticket stamped with the time, 11:30 a.m. When she left the garage that afternoon, her ticket was stamped with the time she left, 12:15 p.m. What was the total length of time that Ms. Luna's car was in the parking garage?

    a. **Unacceptable.** The question is very wordy and uses an unfamiliar context for many children. In addition, "pulled her car" is an idiom that children may not know.

11. Sandip went to the library at 11:30 in the morning. He left at 12:15 that afternoon. How long did Sandip stay in the library?

    a. **Acceptable.** The reading load is reduced, and the context is familiar.

12. It takes Sarah an average of 30 minutes to clean her bedroom. She cleans her bedroom once a week. How many hours would Sarah spend cleaning her bedroom in one year? Acceptability depends on the mix of items in the test.

    a. **Unacceptable** if many questions in the test had girls cleaning rooms or doing what was traditionally considered "woman's work," because the test would reinforce a stereotype and be unfair.

    b. **Acceptable** if combined with questions showing women doing nontraditional work. Not all children have their own bedrooms, but the concept that some children have individual bedrooms should be neither strange nor upsetting. Whether or not the required knowledge of the number of minutes in an hour and the number of weeks in a year is fair depends on the grade level of the test takers.

13. According to the graph, the number of unemployed workers was highest in which year?

    a. **Acceptable.** The mere mention of unemployed workers is acceptable.

14. Marisa hit the bull's-eye with her rifle 7 times out of 9 shots. What percent of the time did Marisa hit the bull's-eye?
    a. **Unacceptable.** Students who are not familiar with the phrase "bull's-eye" in the context of a target will have a rather gruesome mental picture of Marisa's shooting. The use of guns tends to be controversial in any case.
15. The data tables below show how long a driver will be impaired based on the consumption of 1, 2, or 3 ounces of alcohol within one hour. Use the data to predict the amount of time a driver will be impaired after consuming 4 or 5 ounces of alcohol in one hour. Explain your reasoning for obtaining the predicted values.
    a. **Unacceptable**. A brief item concerning alcohol might be acceptable in the higher grades in the context of showing impairment, but basing an entire performance item on the topic is excessive. Also, showing consumption of more than 1 or 2 drinks per hour models inappropriate or even dangerous behavior.
16. A pizza is cut into 8 slices. If 5 girls eat one slice each, how many slices will be left?
    a. **Acceptable.** One slice of pizza is not excessive consumption of junk food.

Excerpts from stimuli for English language arts

The next set of examples consists of brief excerpts from ELA stimuli.

1. Wagner used the orchestra to achieve certain effects in much the same way that other composers of operas used the singers.
    a. **Acceptable** if the knowledge needed to answer the questions was included in the passage. The mere mention of opera or a composer does not make the excerpt unfair.
    b. **Unacceptable** if understanding the passage required knowledge of opera and how composers "used" the orchestra or "used" singers.
2. The African Americans living in Middletown tended to be part of households consisting of extended families living together.
    a. **Acceptable.** The statement of fact about a particular group of African American people is acceptable and does not reinforce a stereotype.
3. Cyanide is one of the fastest-acting poisons known to science.
    a. **Unacceptable.** The excerpt violates the guideline about avoiding dangerous actions and substances. Parents are likely to oppose including information about lethal substances in the test.
4. The AIDS epidemic, which has devastated some countries in sub-Saharan Africa, has affected children as well as adults, leaving many children not only orphaned and uncared for, but also malnourished, diseased, and close to death.
    a. **Unacceptable.** Excessive detail about the suffering of children makes the excerpt unacceptable.

5. Harlow was best known for the experiment in which he separated infant monkeys from their mothers shortly after the infants were born.
    a. **Unacceptable.** The excerpt violates the guideline that prohibits inclusion of painful or harmful experimentation. The excerpt would be acceptable in a psychology test, however.
6. I love to make videos! I use the camera in my phone to capture my friends having a good time with their dates at parties and at school dances.
    a. **Unacceptable.** Owning a cell phone with video capabilities is currently a luxury beyond the reach of many test takers. The references to "dates" and "dances" are not in compliance with the guideline concerning social dancing.
7. An ancestor of the modern horse the size of a dog gave rise to progressively larger species.
    a. **Acceptable.** The passage concerns the evolution of horses, which is in compliance with the guideline that identifies the evolution of human beings as the aspect of evolution to avoid.
8. The Japanese immigrants enrolled in Ms. Kubota's class worked very hard.
    a. **Acceptable.** The reference is to a particular group of Japanese immigrants, so it does not stereotype all Japanese immigrants.
9. The amount of caffeine in a cup of coffee can still affect the human body more than three hours after it has been ingested.
    a. **Acceptable.** The mention of caffeine appears to be in an objective discussion of the effects of drinking coffee and would be in compliance with the guideline on harmful substances, if the passage did not encourage the drinking of coffee.
10. People who drive gas-guzzling SUVs contribute to global warming.
    a. **Unacceptable.** This excerpt is a clear violation of the guideline against advocating for one side in a controversial situation.
11. In the 17th century, many convicted criminals were hanged, but some were slowly crushed to death.
    a. **Unacceptable.** Death by slow crushing is clearly out of compliance with the guidelines about death and suffering.
12. A large influx of immigrants will destroy the equilibrium of a neighborhood.
    a. **Unacceptable.** The negative view of immigrants in the excerpt makes it out of compliance with the guideline forbidding offensive stereotypes of any group. The verb "destroy" is particularly harsh in that context.
13. There has been an increase in the number of people who identify themselves as American Indians.
    a. **Acceptable.** Either "American Indian" or "Native American" is appropriate. The fact that more people than before identify themselves as American Indians is not a fairness problem.
14. Surprisingly, a girl won the math contest.

a. **Unacceptable.** By expressing surprise that a girl won the math contest, the excerpt reinforces the stereotype that girls have less quantitative ability than boys.

15. The soldiers and their wives attended the ceremony.

    a. **Unacceptable.** Unless the reference is to a previously specified group of all male soldiers, refer to "the soldiers and their spouses" to avoid the implication that only males are soldiers.

16. …that all men are created equal, that they are endowed by their Creator with certain unalienable Rights…

    a. **Acceptable.** In spite of the use of "men" to refer to all people and in spite of the reference to God, the excerpt is acceptable because it is from an important historical document of the type required by the Common Core State Standards.

17. Bridges with steel frames are more likely to survive an earthquake than are stone bridges.

    a. **Acceptable.** The mention of a natural disaster is acceptable. There is no focus on death and destruction.

18. Lee's father and Juan's father are both policemen.

    a. **Unacceptable.** Even though both officers are male, "police officers" is preferred to "policemen" to avoid the impression that only men are police officers.

19. The ancient Romans played handball and engaged in other sports while nude in the public baths.

    a. **Unacceptable.** Though unintended, "engaged in other sports while nude" could be taken as sexual innuendo.

20. Many of the people in the United States who speak Spanish come from Mexico.

    a. **Acceptable.** The excerpt is a statement of fact and is not a violation of any guideline.

21. He be at work…

    a. **Unacceptable.** The use of dialect is stereotyped language and is in violation of the guideline unless it is important to measure a Common Core State Standard in literary or historical material.

22. The men's room is on the right; the girls' room is on the left.

    a. **Unacceptable.** Parallel language would call for "women" to match "men" or "boy" to match "girl."

23. Some Native Americans claim to be members of the Algonquian tribe, but according to anthropologists, "Algonquian" is a general term applied to many Native American peoples who speak related languages, not the name of any particular tribe.

    a. **Unacceptable.** There is a problem in that the academic definition of "Algonquian" is taken as correct, but the usage of Native Americans about themselves is taken as incorrect. The excerpt is out of compliance with the guideline to call people what they prefer to be called.

24. Frederick Douglass, the great African American abolitionist, was said to be born on Valentine's Day.
    a. **Acceptable.** The excerpt requires no knowledge of how or why Valentine's Day is celebrated, nor any agreement that it should be celebrated.
25. Edward Said and Daniel Barenboim cofounded a children's orchestra.
    a. **Unacceptable.** Though there is nothing overtly problematic about the excerpt, Edward Said was famous as a Palestinian activist and remains a highly controversial figure. He is viewed very positively by some groups and very negatively by other groups. Reviewers who are not familiar with the people depicted in test materials should check reference sources to avoid the inadvertent inclusion of controversial figures.
26. That on the 1st day of January, A.D. 1863, all persons held as slaves within any State or designated part of a State the people whereof shall then be in rebellion against the United States shall be then, thenceforward, and forever free.
    a. **Acceptable.** Mention of slavery is acceptable, and the Common Core State Standards call for the inclusion of documents important in American history, such as the Emancipation Proclamation.

English language arts items

The next set of examples consists of items or parts of English language arts items.

1. According to the passage, how long ago did Homo sapiens evolve into a distinct species?
    a. **Unacceptable.** To answer the question about when human beings evolved implies that human beings evolved from other species. That implication is not in compliance with the guideline regarding evolution.
2. The character delivering the monologue attributes the arrogance of the French to which of the following?
    a. **Unacceptable.** Describing all of the people in a nation as "arrogant" is a clear case of offensive stereotyping. The question is not in compliance with the guideline concerning stereotypes.
3. Describe the changes within the ecosystem portrayed in the video, including the impact of man's activities on weather patterns, and possible solutions to correct ecological problems.
    a. **Unacceptable.** The question uses "man" to refer to all people, which is not in compliance with the guideline on appropriate terminology for men and women. The influence of people on climate change is highly controversial and is out of compliance with the guideline on the avoidance of advocacy.
4. The author compares the artist's use of color to which of the following?
    a. **Acceptable** if direct experience of color is not required to understand the passage and answer the items.

b. **Unacceptable** if direct experience of color is required. The material would be unfair for students who are blind.

5. Our society stereotypes old people as weak, uninformed, forgetful, and foolish. Discuss the extent to which you agree or disagree with this stereotype.

   a. **Unacceptable.** The question blatantly reinforces stereotypes of a group and invites test takers to agree with the offensive stereotypes.

6. Isaiah wrote, "Woe unto them that are wise in their own eyes." Describe the meaning of that quotation and give two examples of people who are "wise in their own eyes" from your reading or from your personal experience. Explain your choices.

   a. **Unacceptable.** The excerpt violates guidelines about the avoidance of religious material, even though the students are not asked to write directly about religion.

7. In the play, Luz was restricted to a wheelchair for which of the following reasons?

   a. **Unacceptable.** The phrase "was restricted to a wheelchair" should be replaced with more objective terminology such as "began using a wheelchair."

8. According to the newspaper article, Robert died how many years after his brother John?

   a. **Acceptable.** According to the *Guidelines*, it is acceptable to mention death as long as gruesome details are not depicted.

9. It can be inferred from the passage that the spinnaker is most effective during a race when the wind is in which position relative to the boat?

   a. **Unacceptable.** Using sailboats for racing is out of compliance with the prohibition against luxuries. Also, unless "spinnaker" and its use are explained clearly in the passage, the item would depend on invalid specialized knowledge.

10. Based on information in the documentary, which of the following people is most likely to carry the sickle cell trait but show no symptoms of the sickle cell disease?

    a. **Unacceptable.** Diseases that affect particular groups of people are likely to be problematic in terms of fairness. This topic is best avoided.

11. The lecturer stated that among spiders found in many houses in the United States, the bite of which of the following is most likely to cause painful, deep wounds?

    a. **Unacceptable.** The focus on "painful, deep wounds" from spiders "found in many houses" makes the item out of compliance with the guideline regarding animals that are frightening to children.

12. The video excerpt of Baryshnikov dancing in *The Nutcracker* best illustrates which of the following aspects of his work described in the magazine article?

    a. **Acceptable** if all of the information needed to respond to the item is included in the video excerpt and the magazine article.

    b. **Unacceptable** if knowledge of ballet is required to answer the item. Only social dancing of couples is prohibited by the guidelines.

13. Read the excerpt from the diary of a ship captain engaged in transporting slaves and watch the video dealing with the history of slavery in the United States. Imagine that you are a newly captured slave. Describe your experiences on land and on the sea during your journey from Africa to the United States. Use information from both the diary and the video in your description.

    a. **Unacceptable.** Mention of slavery as a topic is acceptable, but forcing test takers to imagine that they personally experienced the transatlantic journey, during which many captives are known to have suffered and died, will be upsetting to some students.