ASSESSMENT AND EXERCISE SPECIFICATIONS

2006 NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS (NAEP) IN ECONOMICS

September 2002

Developed Under Contract Number ED01CO0130 for the National Assessment Governing Board by the American Institutes for Research, National Council on Economic Education, and the Council of Chief State School Officers

TABLE OF CONTENTS

PREFACE 1

CHAPTER ONE: INTRODUCTION 2

NAEP Administration and Student Samples 2

CHAPTER TWO: ASSESSMENT SPECIFICATIONS 4

Aligning the Assessment with the Framework 4
Accessibility 5
Assessment Accessibility Components 5
Reporting Requirements 6
Content Areas 7
Cognitive Categories 8
Item Formats 9
Contexts 10
Economic Terminology and Economic Tools 11

CHAPTER THREE: INCLUSION OF STUDENTS WITH DISABILITIES AND ENGLISH LANGUAGE LEARNERS 13

Accommodations 13
Item-Writing Considerations for English Language Learners 14
Item-Writing Considerations for Students with Disabilities 15
Scoring Responses from English Language Learners 15
Literacy Issues 15
Varied Background Experiences 16

CHAPTER FOUR: GENERAL ITEM SPECIFICATIONS 17

NAEP Item Development and Review Policy Statement 17
Item Characteristics 18
Cognitive Categories 18
Knowing 19
Applying 19
Reasoning 19
Principles of Good Item Writing 20
Item Formats 22
Stimulus Materials 23
Developing Multiple-Choice Items 24
Developing Constructed-Response Items and Scoring Rubrics 25
Short Constructed-Response Items 25
Extended Constructed-Response Items 26
Aligning Items and Rubrics 26
Item Tryouts and Reviews 28
Recruitment of Item Writers 29

Recruitment and Training of Item Scorers 29

CHAPTER FIVE: ITEM SPECIFICATIONS BY CONTENT AREA 31

The Market Economy 34
The National Economy 40
The International Economy 45

CHAPTER SIX: SAMPLE ITEMS 48

APPENDIX A: FRAMEWORK DEVELOPMENT COMMITTEES, MANAGEMENT TEAM, AND TASK LEADERS A-1

APPENDIX B: NAEP ECONOMICS PRELIMINARY ACHIEVEMENT LEVEL DESCRIPTIONS GRADE 12 ..... B-1

APPENDIX C: INAPPROPRIATE TERMINOLOGY FOR ITEM WRITERS C-1

APPENDIX D: NATIONAL ASSESSMENT GOVERNING BOARD NAEP ITEM DEVELOPMENT AND REVIEW POLICY STATEMENT D-1

APPENDIX E: ACKNOWLEDGMENTS E-1

PREFACE

The *Assessment and Item Specifications: NAEP 2006 Economics* translates the *Assessment Framework: 2006 National Assessment of Educational Progress in Economics* into guidelines for developing items and for developing the assessment as a whole. The primary purpose of the specifications is to provide the National Center for Education Statistics (NCES) and its assessment development contractor with information that will ensure that the NAEP Economics Assessment reflects the intent of the *Assessment Framework: 2006 National Assessment of Educational Progress in Economics* adopted by the National Assessment Governing Board (NAGB). The specifications are structured so that the assessment developer and item writers have a single document to refer to when they are working on the assessment. Therefore, portions of the framework that specify characteristics of the assessment or items are repeated in these specifications.

The document is divided into six chapters:

Chapter One: Introduction
This chapter provides background on the NAEP Economics Assessment and information about how and to whom the assessment is administered.

Chapter Two: Assessment Specifications
This chapter provides guidelines that apply to developing the assessment as a whole. It includes information about how the entire item pool should be structured so that it aligns with the *Assessment Framework: 2006 National Assessment of Educational Progress in Economics*.

Chapter Three: Inclusion of Students With Disabilities and English Language Learners
This chapter describes accommodations and other considerations for assessing students with disabilities and students who are English language learners. It includes item-writing considerations for these students that apply to all the items written for the assessment.

Chapter Four: General Item Specifications
This chapter focuses on item specifications. It includes full discussions of characteristics of NAEP economics items as well as descriptions of item formats and item-writing guidelines.

Chapter Five: Item Specifications by Content Area
This chapter contains descriptions of the benchmarks in each of the content areas covered by the assessment and gives further guidelines specific to the content areas and benchmarks.

Chapter Six: Sample Items
This chapter contains sample items illustrating the concepts and item-writing principles in Chapter Four and the specifications for content areas and benchmarks described in Chapter Five.

CHAPTER ONE
INTRODUCTION

The National Assessment of Educational Progress (NAEP), also known as "the Nation's Report Card," is the only nationally representative and continuing assessment of what America's students know and can do in various subject areas. Since 1969, assessments have been conducted periodically in reading, mathematics, science, writing, U.S. history, civics, geography, and the arts. The National Assessment Governing Board (NAGB), created by Congress in 1988, is responsible for formulating policy for NAEP. NAGB is specifically charged with developing assessment objectives and specifications, identifying appropriate achievement goals for each age and grade, and carrying out other NAEP policy responsibilities. In September 2001, NAGB awarded a contract to the American Institutes for Research, the Council of Chief State School Officers, and the National Council on Economic Education to develop recommendations to NAGB on the framework and specifications that will be used in the 2006 NAEP Economics Assessment.

Three project committees were involved in the NAEP 2006 framework development process:

1. The Steering Committee, consisting of representatives of the business community, economics and personal finance organizations, economics and social studies education, and media and government
2. The Planning Committee, consisting of economics educators in K–12 public and private schools and in colleges and universities, social studies consultants, and members of economics and personal finance education organizations
3. The Technical Advisory Panel, consisting of measurement experts from universities and education and research organizations

(See Appendix A for lists of committee members.)[1]

NAEP ADMINISTRATION AND STUDENT SAMPLES

As currently planned, the NAEP Economics Assessment will be administered in early 2006, with the administration conducted by trained field staff. Each economics assessment booklet will contain two separately timed, 25-minute sections of economics items. Three types of items will be used: multiple choice, short constructed response, and extended constructed response. The assessment is designed to have multiple forms of the assessment booklets. The items will be distributed across the booklets, using a matrix sampling design, so that students taking part in the assessment do not all receive the same items. In addition to the economics items, the assessment booklets will include background questionnaires, which take about 10 minutes to complete.[2]

[1] For more information about the development process and results, see the *Assessment Framework: 2006 National Assessment of Educational Progress in Economics*.

[2] See the NAEP website at http://nces.ed.gov/naep for more information about administrations, conditions, and procedures.

The details of the item development process—number of items, ratio of number of field-tested items to number of items in final pool, number of assessment blocks, and so on—will be determined by NCES in consultation with the assessment development contractor.

The assessment is designed to measure the economics achievement of students in the nation's schools in grade 12 and to report the results at the national level. To implement this goal, schools throughout the country are randomly selected to participate in the assessment. The sampling process is carefully planned to select schools that accurately represent the broad population of U.S. students.

The selection process is designed to include schools of various types and sizes, both public and private, from a variety of community and geographical regions. Their student populations represent different levels of economic status; racial, ethnic and cultural backgrounds; and instructional experiences. Students with disabilities and English language learners are included to the extent possible, with accommodations as necessary (see Chapter Three for more information about inclusion criteria and accommodations). The sophisticated sampling strategy helps ensure that the NAEP program can generalize the assessment findings to the diverse student populations in the nation. It allows the program to present information on the strengths and weaknesses in aggregate student understanding of economics, the analytical skills appropriate to economics, and the ability to apply that understanding and those skills in problem-solving situations; to provide comparative student data according to race/ethnicity, type of community, and geographic region; to describe trends in student performance over time; and to report relationships between student achievement and certain background variables.

CHAPTER TWO

ASSESSMENT SPECIFICATIONS

This chapter provides guidelines for developing the overall assessment. It includes information about how the entire item pool should be structured so that it aligns with the *Assessment Framework: 2006 National Assessment of Educational Progress in Economics.*

ALIGNING THE ASSESSMENT WITH THE FRAMEWORK

The assessment should be developed to align with the content expectations articulated by the standards, benchmarks, and cognitive categories defined by the *Assessment Framework: 2006 National Assessment of Educational Progress in Economics.* The work done in the areas of science and mathematics by Webb (1999) and others (e.g., Schmidt, 1999) became the basis for the five interrelated dimensions on which the NAEP assessment is structured so that it aligns with the framework.[3]

1. The match between the content of the assessment and the content of the framework: The assessment as a whole must reflect the breadth of knowledge and skills covered by the standards and benchmarks in the framework.

2. The match between the complexity of economics knowledge and skills on the assessment and in the framework: The assessment must represent the balance of cognitive categories as described in the framework.

3. The match between the emphasis of the assessment and the emphasis of the standards, and contextual requirements in the framework: The assessment must represent the balance of content and item formats specified in the framework and give appropriate emphasis to the conditions in which students are expected to demonstrate their achievement in economics.

4. The match between the assessment and the way scores are reported and interpreted: The assessment must be developed so that scores reflect both the framework and the performance described in the NAEP achievement levels.

5. The match between the assessment design and the characteristics of the targeted assessment population: The assessment must give all students tested a reasonable opportunity to demonstrate their knowledge and skills in the content and cognitive categories described in the framework.

These five dimensions are the foundation for the NAEP assessment and item specifications. The principles in these dimensions are used in each of the sections that follow.

---

[3] Webb, N.L. (1999). *Alignment of science and mathematics standards and assessments in four states.* Washington, DC: Council of Chief State School Officers; Schmidt, W. (1999, June). *The alignment of standards and assessments.* Paper presented at the CCSSO Annual Conference on Large-scale Assessment, Snowbird, UT.

ACCESSIBILITY

Accessibility in an educational assessment context refers to the degree to which the assessment provides all students in the targeted population with the opportunity to demonstrate their achievement in relation to the construct of interest, in this case the economics defined by the NAEP framework. The design for the NAEP Economics Assessment must address issues of student access—considerations that can either facilitate or block the goal of obtaining valid measurements of the assessed students' achievement in economics.

The NAEP Economics Assessment is designed to measure the achievement of students across the nation. Therefore, it should allow all students to demonstrate their content knowledge and skills—students who have learned economics in a variety of ways, following different curricula and using different instructional materials; students who have mastered the content to varying degrees; students who have learned economics outside of the classroom and have had no formal instruction in economics; students with disabilities; and students who are English language learners. The question to ask in developing the assessment is, What is a reasonable way to measure the *same intended constructs* for students who come to the assessment with different experiences, strengths, and challenges, who approach the constructs from different perspectives, and who have different ways of displaying their knowledge and skills?

The central requirement of such an assessment is that it must measure the same economic constructs across diverse groups of students. To this end, the assessment should maintain the rigor of the economics expectations in the framework while providing the means for all assessed students to demonstrate their levels of knowledge and skills.

The assessment and item specifications found in this chapter and in Chapters Three and Four incorporate methods for making the assessment accessible to most students, including those with disabilities and those learning English. The general item specifications in Chapter Four incorporate strategies for making the assessment accessible for all students.

Assessment Accessibility Components

The following appropriate actions will help ensure that the assessment is accessible to diverse population of students:

- Carefully present items to students in the assessment (e.g. plain language editing procedures, use of visuals, item format considerations).
- Allow multiple ways to respond to constructed-response items, as appropriate to the knowledge and skill assessed.
- Develop careful and detailed scoring rubrics so that the targeted knowledge and skills are evaluated at all score levels.
- Format the assessment booklet to allow enough space between items, and make judicious use of boxes and lines.
- Provide robust training procedures and materials for scorers.
- Provide administration accommodations for assessment takers with disabilities.

REPORTING REQUIREMENTS

The NAEP Economics Assessment reports results for the nation's students in grade 12 as well as for subgroups of the population defined by specific demographic characteristics, such as gender, type of school attended, and eligibility for free or reduced-price lunch (see *The Nation's Report Card*, available at http://nces.ed.gov/nationsreportcard for more information about subgroup scores). Scores are not reported for individual students or schools. Because the economics assessment is administered only to a national sample, it has no state-level results.

Results are reported in two ways: scale scores and achievement levels. Scale scores can range from 0 to 500. Reports identify the percentage of students who reach three achievement levels: *Basic*, *Proficient*, and *Advanced*. These achievement levels describe what students should know and be able to do in economics (see Appendix B for full descriptions).

- *Basic* denotes partial mastery of prerequisite knowledge and skills that are fundamental for proficient work.
- *Proficient* represents solid academic performance. Students reaching this level have demonstrated competency over challenging subject matter, including subject-matter knowledge, application of such knowledge to real-world situations, and analytical skills appropriate to the subject matter.
- *Advanced* represents superior performance.

The assessment should be designed so that results can be reliably and validly reported for the population and for subgroups by scale scores and achievement levels. Because the results are intended to describe the achievement of all students in the nation, results should provide an accurate picture of achievement across the entire scoring scale.

CONTENT AREAS

The framework for the 2006 Economics Assessment is anchored in three broad areas of economics content: the Market Economy, the National Economy, and the International Economy. These divisions are not intended to separate economics into discrete elements. Rather, they are intended to provide a helpful classification scheme that describes the full spectrum of economics content assessed by NAEP. Classifying items into one primary content area is not always clear cut, but doing so brings us closer to the goal of ensuring that important economic concepts and skills are assessed in a balanced way.

The assessment will include items from the benchmarks contained in the standards in each content area (see Chapter Five: Item Specifications by Content Area for detailed descriptions of the content areas). Table 1 shows the overall percentage of time that students should spend on items in the three content areas.

Table 1. Percent Distribution of Assessment Time by Content Area

Content Area	Percent
The Market Economy	45%
The National Economy	40%
The International Economy	15%

It is unlikely that every benchmark within a standard will have a unique item written to it. However, the item development managers should work to ensure that all benchmarks within a standard are assessed, even if that means that some benchmarks are assessed by items classified according to other benchmarks, but whose content addresses multiple benchmarks. The breadth and relative emphasis of economic knowledge and skills covered by the standards in each content area must be represented on the assessment as a whole.

- Each item on the assessment must measure part or all of one or more benchmarks.
- For multiple-choice items, incorrect options should be based on inadequate or incomplete knowledge of the benchmark(s) assessed.
- For constructed-response items, all criteria in the scoring rubrics should be based on the benchmark(s) assessed. On items for which partial credit may be given, students receive full credit if and only if they fulfill all the requirements of the item. Some responses may exceed the requirements of the items; however, *all* responses that meet all the items' requirements will receive full credit.
- The items and tasks should not require students to use knowledge and skills irrelevant to the benchmark(s) assessed. For example, care should be taken to word items clearly and simply, and appropriate accommodations should be available for students who require assistance in reading. (See Chapter Three for more discussion of this topic.)
- The contexts (e.g., *individual and household*, *business*, or *public*) in which items are set should be appropriate to the benchmark(s) being assessed.[4]

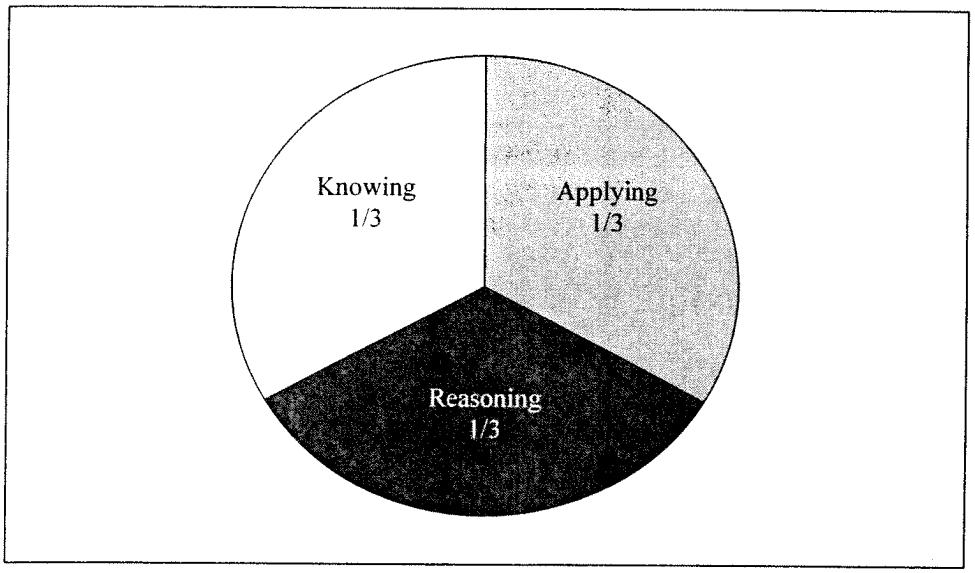
[4] See definitions of contexts later in this chapter.

Chapters Four and Five provide specifications for item development that expand on these topics. Specific recommendations for weighting the standards in each content area are presented in Chapter Five.

COGNITIVE CATEGORIES

The *Assessment Framework: 2006 National Assessment of Educational Progress in Economics* defines three cognitive categories: Knowing, Applying, and Reasoning. As shown in Figure 1, each category should represent one-third of the assessment in terms of assessment time.

Figure 1. Distribution of Assessment Time by Cognitive Category



The Knowing category measures students' abilities to identify and recall information and to recognize economic terms and concepts.

The Applying category measures students' abilities to describe or explain the relationship between information (data, summaries, headlines, problems, and scenarios) and economic concepts.

The Reasoning category measures students' abilities to use information and economic concepts accurately to solve problems, evaluate issues, and interpret solutions.

More detail on the cognitive categories can be found in Chapter Four. The three cognitive categories should be distributed

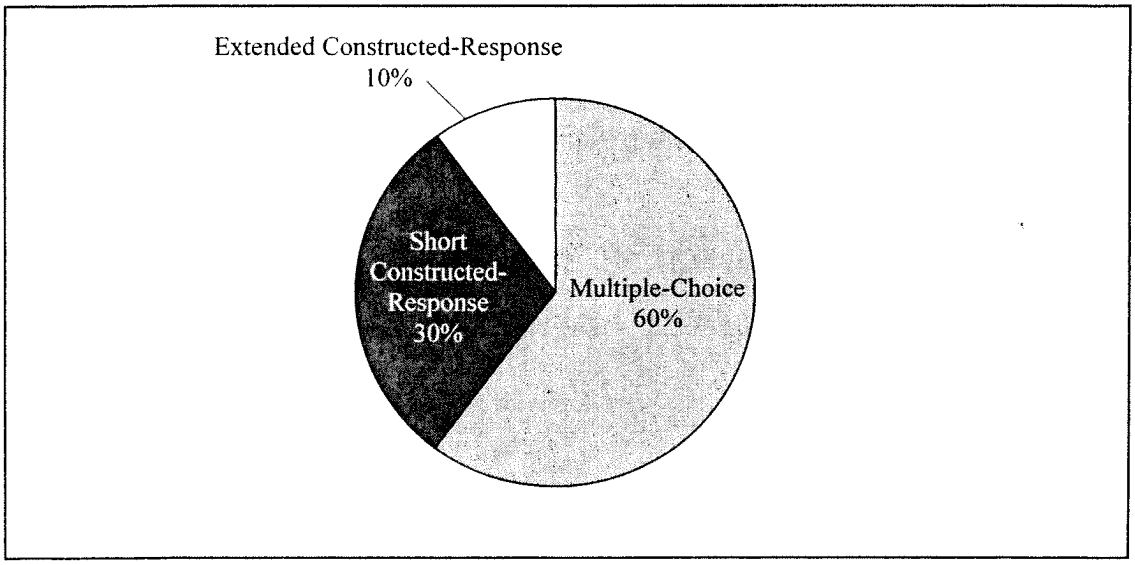
- across the three content areas as appropriate to the construct being assessed; and
- among the three item formats (multiple choice, short constructed response, and extended constructed response).

ITEM FORMATS

The assessment will use multiple-choice, short constructed-response, and extended constructed-response items. Multiple-choice items are scored right or wrong. Short constructed-response items are scored according to scoring rubrics with two or three score levels, and extended constructed-response items are scored according to scoring rubrics with four or five score levels. In both cases, a zero is the lowest score point. On multi-step items for which partial credit may be given, students receive full credit if they fulfill all the requirements of the item. In fact, they must receive full credit when they fulfill all requirements of the item. Chapters Four and Five contain guidelines for writing multiple-choice items and constructed-response items and scoring rubrics.

As shown in Figure 2, approximately 60 percent of student testing time should be allotted to multiple-choice items, approximately 30 percent to short constructed-response items, and approximately 10 percent to extended constructed-response items. These percentages represent the proportion of total assessment time, not the proportion of total items. The time students take to complete any item will depend on two factors in addition to format: the item's difficulty and the cognitive category. If the developers do not have data about how long students take to complete the items, they should use informed judgment based on each item's characteristics to assemble field-test blocks so that the blocks will fit time specifications. Once items have been field-tested, information about the actual time that students needed to complete the items should be used in developing the blocks.

Figure 2. Percent Distribution of Assessment Time by Response Item Type



CONTEXTS

Although a small number of items may not be set in a specific context, 60 to 90 percent should be, with 20 to 30 percent of the items set in each of the following three contextual areas:

- An *individual and household* context, including questions related to personal finance, that is, earning, spending, saving, borrowing, and investing;
- A *business* context, including questions related to entrepreneurs, workers, producers, and investors; and
- A *public* context, including questions related to government, policy, citizenship, and domestic and international organizations.

This requirement is intended to

- assess the extent to which students are able to demonstrate their understanding of economics in real-world situations; and
- make the assessment accessible to students whose understanding of economic concepts is derived from experience or familiarity with everyday economic decision-making rather than formal study.

The remainder of the items will be in other contexts, in multiple contexts, or context free. Context-based and context-free items will be distributed across the three content areas (the Market Economy, the National Economy, and the International Economy) and across the three cognitive categories (Knowing, Applying, and Reasoning). Examples of items and the types of answers expected in each of the different contexts follow:[5]

Example of a context-free item:

> Using demand and supply analysis, explain how the quantity of a product purchased could increase when price increases.
>
> *An increase in demand for a product will cause an increase in price and an increase in the amount produced and sold. The law of demand states that if everything else remains the same, the quantity of a product purchased will decrease when price increases. However, both price and quantity can increase if something else has changed and causes demand to increase.*

Example of an *individual- and household*-context item:

> Last year Marisol spent $500 per month to rent an apartment near State University. Many more students are looking for apartments this year than looked last year.

[5] The contextual examples could be written as multiple-choice or short constructed-response items.

What will most likely happen to the rent for a typical apartment near State University? Explain why.

How will Marisol respond? Explain why.

The rent will increase with an increase in demand for apartments. Marisol will have to reduce her consumption of other goods and services, rent a smaller apartment or move into a dorm room, spend more of her savings, and/or take a part-time job.

Example of a *business*-context item:

The price of cheese that restaurants use to make pizza has doubled during the past year.

What will most likely happen to the price of pizzas and the number of pizzas sold? Explain why.

The cost of making pizzas has increased because the price of cheese has doubled. As a result, the supply of pizzas will decrease, the average price will increase, and the number of pizzas sold will decrease.

Example of a *public*-context item:

Some high school graduates choose between going to college or entering the armed forces right after high school. Suppose Congress has recently passed a program that will significantly lower college tuition.

What will most likely happen to the number of high school graduates who go to college? Explain why.

What will most likely happen to the number of high school graduates who enter the armed forces? Explain why.

The program will increase the number of high school graduates who will choose to go to college because of the lower tuition. If more of these graduates go to college, fewer graduates will be available to enter the armed forces right after high school.

Note that although many questions on personal finance fall within the *individual and household context,* items with a personal finance perspective can be set in any context.

ECONOMIC TERMINOLOGY AND ECONOMIC TOOLS

Items should not include certain economic terms that are not accessible to all students, although the economic *concepts* named by these terms can and should be assessed when they are included in the standards and benchmarks. A list of terms that item writers should not use in the stems and

options of items is provided in Appendix C. In some cases, alternatives are suggested for testing the concepts that those terms represent.[6]

Also, calculation and manipulation of numbers should be minimal. Items will not require complex calculations that depend on a calculator and will not require the production of graphs, tables, and charts. However, graphs, tables, and charts may be included as stimulus materials for multiple-choice items in the assessment. Additionally, graphs may be used as options in a limited number of multiple-choice items. Finally, students may produce a chart or a graph as an appropriate response to a constructed-response item, if they so choose, but they will not be required to do so. Detailed specifications for stimulus materials can be found in Chapter Four.

[6] Students learn economics in a number of ways, and many may not know the terms used in standard economics courses and texts but may still understand the concept being assessed. The emphasis throughout the assessment should be on assessing how well students understand the meaning of concepts and can use application and reasoning skills, not how well they can define technical and economic-specific language. Item writers should be careful about the use of language and should write items, where possible, that assess understanding without using specific economic terminology. On the other hand, students who have had formal courses in economics may use economic terminology when responding to constructed-response items. Students who indicate equal understanding on a given constructed-response item should receive equal credit, regardless of whether they use economic terminology or not.

CHAPTER THREE

INCLUSION OF STUDENTS WITH DISABILITIES AND ENGLISH LANGUAGE LEARNERS

Most students with disabilities are eligible to be assessed in the NAEP program. Similarly, most students who are English language learners also are eligible to participate in NAEP. Since the assessment is designed to measure the academic achievement of all assessment takers at a given grade, students with a range of backgrounds and experiences, including those with disabilities and those who are mastering English as a second language, are included in the assessment. To make sure that the results provide a meaningful representation of all students' knowledge and skills, the assessment provides access for these students as well as others.

As noted in Chapter Two, NAEP addresses the issue of accessibility in two ways. One is to carefully follow item and assessment development procedures to build accessibility into the standard assessment. The other is to provide accommodations for students with disabilities and for English language learners.[7]

ACCOMMODATIONS

For many students with disabilities and students whose native language is not English, the standard administration of the NAEP assessment will be most appropriate. For some students with disabilities and some English language learners, the use of one or more administration accommodations will be more suitable. How to select and provide appropriate accommodations is an active area of research, and new insights are emerging on how best to apply accommodation guidelines to meet the needs of individual students. NAEP accommodations policy allows a variety of accommodations, depending on the needs of each student. Most accommodations that schools routinely provide in their own testing programs are allowed in the economics assessment, as long as they do not affect the construct tested. These accommodations include but are not limited to

- one-on-one testing,
- small group testing,
- extended time,
- oral reading of directions,
- large-print booklets,

[7] For more information about research that has guided the increased inclusion of students in NAEP, see Olson, J.F., & Goldstein, A.A. (1997). *The inclusion of students with disabilities and limited English proficient students in large-scale assessments: A summary of recent progress.* (NCES Publication No. 97– 482). Washington, DC: National Center for Education Statistics; Mazzeo, J., Carlson, J.E., Voelkl, K.E., & Lutkus, A.D. (1999). *Increasing the participation of special needs students in NAEP: A report on 1996 research activities.* (NCES Publication No. 2000– 473). Washington, DC: National Center for Education Statistics.

- bilingual dictionaries[8], and
- an aide to transcribe responses.

Accommodations are offered in combination, as needed; for example, students who receive one-on-one testing generally also use extended time.

In a very small number of cases, students will not be able to participate in the assessment, even with the accommodations offered by NAEP:

- Students with disabilities whose Individualized Education Plan (IEP) teams or equivalent groups have determined that they cannot participate, *or* whose cognitive functioning is so severely impaired that they cannot participate, *or* whose IEP requires an accommodation that NAEP does not allow
- Limited English proficient students who have received instruction primarily in English for less than three school years *and* who are judged by school staff as unable to participate in the assessment when it is administered in English, with or without an accommodation

ITEM-WRITING CONSIDERATIONS FOR ENGLISH LANGUAGE LEARNERS

Some students who are English language learners will take the standard, English-only version of the assessment. These students are diverse both across and within their language groups. This is particularly the case with Spanish language speakers who come from various countries in Latin American and the Caribbean. Among the Spanish-speaking population are linguistic differences (mainly in vocabulary), cultural differences, and differences in socio-economic backgrounds. English language learners may have trouble understanding what items are asking on assessment forms administered in English.[9]

Item development managers should address the following issues during the item writing and review process to minimize misinterpretation of items:

- It is not uncommon for the same word in English and in the student's native language to have different meanings; when possible, avoid using such words in items and related materials and in overall assessment instructions.
- English and other languages often have different rules of syntax, or word order. Although students may know the basic differences between their primary language and English, subtle differences that can lead to confusion based on word choice, syntax, or order should be avoided.

---

[8] It should be noted that, for the assessment of economics, problems might be introduced by the use of bilingual dictionaries. In particular, the use of bilingual dictionaries could lead students to answer questions incorrectly. Economic terms such as capital, scarcity, shortage, surplus, saving, and investment have meanings that are specific to economics and different from standard dictionary meanings.

[9] For more information about designing assessments that are accessible to English language learners, see Kopriva, R. (2000). *Ensuring accuracy in testing for English language learners*. Washington, DC: Council of Chief State School Officers.

- Students' informal translations of some words or phrases between other languages and English can lead to unintended interpretations in items. These can often be avoided by modifying the word choice or the structure of sentences.
- Dual meanings of words, which are frequent in the English language, can be unnecessarily confusing for English language learners and can be avoided by using a different word, defining the word, or providing a fuller context.

Chapter Four contains additional item-writing techniques that can increase accessibility for English language learners and other students.

ITEM-WRITING CONSIDERATIONS FOR STUDENTS WITH DISABILITIES

Most students with disabilities will take the standard assessment without accommodations, and those who take the assessment with accommodations also will use the standard version of the assessment. Item writers and the assessment developer should minimize item characteristics that could hinder accurately measuring the economics achievement of students with disabilities. In particular, they should do the following:

- Avoid layout and design features that could interfere with the ability of the student to understand the requirements and expectations of the item.
- Use plain language.
- Avoid item contexts that assume background experiences that may not be common to some students with sensory or physical disabilities.
- Develop items so that they can be used with allowed accommodations.

Chapter Four contains additional item-writing techniques that can increase accessibility for students with disabilities and for other students.

SCORING RESPONSES FROM ENGLISH LANGUAGE LEARNERS

Literacy issues and varied background experiences have an impact on how well scorers can properly read, understand, and evaluate the responses of English language learners to constructed-response items.[10] These are discussed below.

Literacy Issues

Responses sometimes can be difficult to read because language acquisition confusion arises between the students' native language and English. Although this situation is developmentally appropriate in terms of language acquisition, many scorers are not trained to interpret these types of systematic errors. The following procedures should be used to score these responses properly:

[10] For information about scoring responses from English language learners, see Kopriva, R., & Saez, S. (1997). *Guide to scoring LEP student responses to open-ended mathematics items*. Washington, DC: Council of Chief State School Officers.

- Scoring leaders should have additional training in recognizing and properly interpreting these responses.
- Experts in reading the responses of English language learners should be available to scorers throughout the scoring process.
- Examples of major systematic errors should be included in training materials for scorers so that they can more accurately score responses. Major systematic language errors include the following:

  - Intermittent use of the student's native language and English, called code switching.
  - Use of native language phonetics in attempting to write English or use of beginning-stage English phonetic spelling.
  - Use of writing conventions from the native language when the student is responding in English.
  - Inappropriate use of compound words (the condensing of words into one large word); transposition of words; and omissions of tense markers, articles, plurals, prepositions, or other words.
  - Substitution of common words for more precise terminology. For instance, it may be acceptable for students to substitute the word fattest for greatest when the intent of the item is not to evaluate students' understanding of economic terminology. However, if the intent is to measure the students' knowledge and ability to use such terminology in an application setting, then this substitution would be incorrect.
  - Confusion about the meaning of words (e.g., left as opposed to right vs. left defined as remaining).
  - Inappropriate use of unfamiliar words.
  - Novice sentence and paragraph structures.
  - Over-reliance on non-verbal forms of communication, such as charts or pictures.

Varied Background Experiences

Novel interpretations and responses are common for English language learners and often reflect background experiences quite different from those of most native English speakers. It is important for scorers to evaluate responses on the basis of the measurement intent of the item and recognize when an unusual response is actually addressing that intent.

It is not unusual for rubrics to implicitly or explicitly favor writing styles that mirror what is taught in language arts curricula in U.S. schools. However, circular, indirect, deductive, and abbreviated reasoning writing styles are encouraged by some cultures, and scorers should be trained to appropriately score these types of responses. When a specific writing style is not the measurement intent of the item, scorers need to understand the nature, conventions, and approaches of these kinds of styles and how to separate the structure and sophistication of the written response from the substantive content being evaluated.

CHAPTER FOUR

GENERAL ITEM SPECIFICATIONS

This chapter discusses specifications that apply to all items included in the NAEP Economics Assessment; Chapter Five contains specifications by content area; and Chapter Three includes a section on item writing considerations for students with disabilities and English language learners that apply to all the items written for the assessment. The guidelines in these three chapters focus on translating the intent of the *Assessment Framework: 2006 National Assessment of Educational Progress in Economics* into the items used on the assessment.

This chapter highlights only some of the central considerations in item development and concentrates on topics specific to the NAEP Economics Assessment. Item writers should refer to directions for developing items provided by the item development contractor, in addition to the information in Chapters Three, Four, and Five.

NAEP ITEM DEVELOPMENT AND REVIEW POLICY STATEMENT

Item writers, as well as item and assessment development managers, for the NAEP Economics Assessment will follow the NAGB policy for NAEP item development, which contains the following six guiding principles for item development and review.[11]

Principle 1
NAEP test questions selected for a given content area shall be representative of the content domain to which inferences will be made and shall match the NAEP assessment framework and specifications for a particular assessment.

Principle 2
The achievement level descriptions for basic, proficient, and advanced performance shall be an important consideration in all phases of NAEP development and review.

Principle 3
The Governing Board shall have final authority over all NAEP test questions. This authority includes, but is not limited to, the development of items, establishing the criteria for reviewing items, and the process for review.

Principle 4
The Governing Board shall review all NAEP test questions that are to be administered in conjunction with a pilot test, field test, operational assessment, or special study administered as part of NAEP.

[11] *NAEP Item Development and Review Policy Statement*, National Assessment Governing Board, adopted May 18, 2002. See Appendix D for the complete policy.

Principle 5
NAEP test questions will be accurate in their presentation, and free from error. Scoring criteria will be accurate, clear, and explicit.

Principle 6
All NAEP test questions will be free from racial, cultural, gender, or regional bias, and must be secular, neutral, and non-ideological. NAEP will not evaluate or assess personal or family beliefs, feelings, and attitudes, or publicly disclose personally identifiable information.

The NAGB policy is closely aligned with the No Child Left Behind Act of 2001 (P.L. 107-110), which contains a number of important provisions regarding item development and review for NAEP.

ITEM CHARACTERISTICS

Each item written for the NAEP Economics Assessment reflects two major dimensions:

- Economic content area (see Chapter Five); and
- Cognitive category (see next section).

Items also vary by

- Format (see Chapter Two, page 9, and this chapter, pages 22–23); and
- Context (see Chapter Two, pages 10–11).

Each of these characteristics is discussed in the specifications, and examples of items illustrating these characteristics are given in Chapter Six.

COGNITIVE CATEGORIES

Each NAEP item assesses one or more benchmarks that are associated with a standard within a single content area of economics. The item also makes certain cognitive demands on students. These demands define the cognitive category of the item. The demands that an item makes—what it asks the student to recall, understand, reason about, and do—are based on the assumption that the student is familiar with the economics of the task. If a student has not learned about that aspect of economics, the task is likely to make different and heavier cognitive demands, and the student may well not be successful on it.

The cognitive categories—Knowing, Applying, and Reasoning—form an ordered description of the demands an item may make on a student. However, there is not necessarily a perfect correlation between an item's cognitive category and its difficulty. Degrees of difficulty can be found in each category.

The cognitive category of an item is not related to its format (multiple choice, short constructed response, or extended constructed response). Items requiring the student to generate a response can make somewhat heavier demands on students than items requiring a choice among alternatives, but that is not always the case. Any type of item can deal with economics of greater or lesser depth and sophistication. Some multiple-choice items assess complex economics, and constructed-response items can be crafted to assess basic economic concepts.

Each cognitive category is described below, with reference to sample items. In addition, all the sample items in Chapter Six are labeled according to their cognitive categories. As noted earlier, approximately one-third of the total assessment time should be devoted to answering items from each cognitive category. These items should be distributed across the different content areas.

Knowing

Items in the Knowing category measure students' abilities to identify and recall information and to recognize economic terms and concepts. Items in the Knowing category should ask students to

- recognize and recall information and concepts; or
- interpret data and information to identify events or trends.

Item 1 in Chapter Six is an example of an item written in the Knowing category.

Applying

Items in the Applying category measure students' abilities to describe or explain the relationship between information (data, summaries, headlines, problems, and scenarios) and economic concepts. Items in the Applying category should ask students to

- restate an economic concept in their own words;
- interpret data and information to identify events or trends and explain cause;
- analyze a given scenario or event that requires only one step in the analysis; or
- apply or use a concept when the concept is specified.

Items 2 and 6 in Chapter Six are examples of items written in the Applying category.

Reasoning

Items in the Reasoning category measure students' abilities to use information and economic concepts accurately to solve problems, evaluate issues, and interpret situations. Items in the Reasoning category should ask students to

- interpret data to identify an event or a trend, explain the cause, and recommend policy;

- apply or use a concept when the concept is not specified;
- apply more than one concept when one or more concepts are specified; or
- analyze a given scenario or event that requires multiple steps in the analysis.

Item 3 in Chapter Six is an example of an item written in the Reasoning category.

PRINCIPLES OF GOOD ITEM WRITING

The following principles of good item writing should be used as appropriate, depending on the measurement intent of the item. If the intent of the item is to have students recognize the importance of certain information and use that information in their solutions, then it might not be appropriate to include the information in an accompanying graphic. However, these guidelines should be followed unless the targeted construct precludes doing so.

1. Clear Measurement Intent: An essential step in good item writing is making sure that the measurement intent of the item is clear and that students understand what is being measured and what type of response is expected.

   a) Clear Intent in Development

   - Item writers should provide a clear description of what each item is intended to measure. This will help developers classify items according to assessment specifications and develop clear scoring rubrics and scoring materials. It will also reduce confusion in reviews and provide evidence of the degree to which the assessment aligns with the framework.
   - Item writers should ensure that the response to one item does not depend on the response to another. For example, students should not be asked in one item to determine the change in demand for a product and then be asked in another item to use that change to determine the relative price change for a different product. Items can be related to one another by being part of a set of items based on the same stimulus, such as an excerpt from a newspaper article, a chart or graph, cartoon, and so on, but each question in a set must function independently of the others. This restriction is not intended to refer to extended constructed-response items, which may have multiple parts for which credit can be given.

b) Clear Intent for Test Takers

- It should be clear to students what each item is measuring. Item writers should be careful not to make assumptions about how students will interpret an item's implicit requirements.
- Constructed-response items should contain clear directions to students about how they can respond. For example, directions should indicate whether students can use visuals, such as tables, graphs or charts, or equations to respond, or whether students are required to produce verbal descriptions. Directions also should indicate whether more than one type of response is appropriate.
- Constructed-response items should *not* offer students a choice of topics on which to respond.

2. Plain Language: Plain language is a writing and editing tool designed to convey meaning clearly without altering what items are intended to measure. All items should use plain language. Even when the intent of the item is for students to define, recognize, or use economics vocabulary correctly, the surrounding text should be in plain language. Plain language guidelines often increase access and minimize confusion for students.

   - Write items using brief, "simple" sentences or stems.
   - Use the same structure for paragraphs throughout the assessment as much as possible (e.g., topic sentence, supporting sentences, and conclusion sentence).
   - Use present tense and active voice.
   - Minimize paraphrasing.
   - Avoid using pronouns.
   - Use high-frequency words as much as possible.
   - Avoid colloquialisms.
   - When using words with multiple meanings, make sure the intended meaning is clear.
   - Avoid using unnecessary descriptive information.
   - Use format to clarify text; for example, use bullets, allow space between pieces of text, and use boxes and lines judiciously.

3. Visuals: Visuals, such as pictures, charts, and diagrams, can be the basis for the task(s) posed by an item or a set of items. Visuals can also effectively support text, illustrate economic concepts in the text, and increase item access, although their use for these purposes must be weighed against the potential liability of providing redundant information. However valid the intended use of visuals, care must be taken to avoid visuals that are confusing or that distract the students from what the items are asking them to do. Proper use of visuals requires that the visuals:

   - mirror and parallel the wording and expectations of the item text; and
   - contain sufficient, but not extraneous or distracting, information.

Detailed specifications on the use of visuals and other stimulus materials in the economics assessment are given later in this chapter.

Items 4 and 5 in Chapter Six show the effective use of visuals, including their use as response options.

4. Using Contextual Information Appropriately: Approximately 60 to 90 percent of the items will be designed to measure economics in the following contexts, 20 to 30 percent in each: individual and household, business, and public. Using contextualized information places economic concepts in fuller, often more realistic, conditions and provides background information students may not have. However, the contextual information should not interfere with the economics being assessed or become a barrier to a student's ability to demonstrate his or her economics knowledge. Guidelines for using contextual information appropriately include the following:

    - Use plain language as much as possible.
    - Use charts, tables, or other visuals to increase item clarity.
    - Spread the directions and explanations throughout the item or block.
    - Use contexts that are meaningful to the economics being assessed.
    - Use contexts that are appropriate for grade 12.
    - Use familiar contexts; avoid contexts that may confuse or be unfamiliar to some students taking the assessment.
    - Avoid contextual information that could corrupt the measurement of the intended skill.

The items in Chapter Two on pages 10-11 provide examples of items in these contexts.

5. Multiple Response Opportunities: Where possible, constructed-response items should be written to allow students to respond in multiple formats. For example, students might be allowed to show their answers through charts, tables, graphs, formulas, and/or text. Constructed-response items that can be answered *only* by drawing a chart, table, or graph or by providing a formula cannot be used, however. Items and scoring rubrics should be written to measure the economics knowledge and skills from students with a range of achievement. This can be accomplished by developing a variety of items that have different formats and contexts and that assess different cognitive abilities.

ITEM FORMATS

The NAEP Economics Assessment uses three types of items: multiple choice, short constructed response, and extended constructed response.

- *Multiple-choice* items require students to select the correct or best answer to a given item. These items have one correct or best answer and are scored as either correct or incorrect.

- *Short constructed-response* items require students to respond in short answers that may vary from one or two words or phrases to several sentences. Short constructed-response items are scored according to scoring rubrics with two or three categories.
- *Extended constructed-response* items require students to consider a situation that demands more than a short response and allows a number of gradations of correctness. Extended constructed-response items may require the application of an economics concept such as supply and demand, a detailed analysis, the synthesis or interpretation of data, and/or the projection of a trend. Extended constructed-response items are scored according to scoring rubrics with five categories.[12]

Item writers should carefully consider the content areas and cognitive categories they intend to assess when deciding whether to write a multiple-choice or a constructed-response item. Each content area includes knowledge and skills that can be measured with each of the three item formats. As noted earlier, each cognitive category can be measured by any of the item formats. Although a particular cognitive category may seem to lend itself more readily to one item format, each type of item—multiple choice, short constructed response, and extended constructed response—can deal with economics of greater or less depth and sophistication. Guidelines for scoring rubrics are provided later in this chapter.

STIMULUS MATERIALS

In the development of all types of exercises, it is important to control the stimulus tightly. The item writer must formulate clearly in his or her own mind the nature of the task the student is to perform (i.e., the kind of operation the student is intended to carry out); define the materials on which the operation is to be carried out; and incorporate into the item a clear statement of the task to be performed. The statement must include the situation or setting the examinees are to be given and what they are to do.

Stimulus material must be varied and can include both written and non-textual types of materials. Written materials may include information passages, excerpts from newspaper and magazine articles, op-ed columns, headlines, quotations, speeches, interviews, created dialogues, and materials and archives from the Internet. Non-textual stimuli accompanying these written materials may include photos, cartoons (including op-ed cartoons), and Internet postings, as well as diagrams, graphs, tables, and charts.

The use of visual stimuli should be judicious, however. A limited use of data in table form can enhance the assessment of important concepts such as equilibrium price and quantity in a supply and demand problem. Tables, charts, and time-series graphs that summarize levels and trends in data may be used as stimulus material in multiple-choice,

[12] In some cases, it may be appropriate to have four scoring categories for an extended constructed-response item, depending on the construct assessed and the nature of expected student responses to the item.

short constructed-response, and extended constructed-response items to assess students' understanding of such concepts as unemployment, inflation, and economic growth. However, two-variable graphs that depict relationships such as supply and demand, cost and investment functions, and aggregate supply and demand should *not* be used in constructed response items, either as stimulus materials or in the stems of the items.[13]

Although a variety of types of materials should be used, the complexity of the materials should be appropriate for a range of grade 12 students. In addition, careful consideration should be given to the amount and level of reading material presented in the textual materials. The reading difficulty should be below the grade 12 level to ensure that the assessment does not become a test of reading skills. Finally, both textual and non-textual materials must be free from bias and be secular, neutral, and non-ideological (see Principle 6 on page 19).

Developing Multiple-Choice Items

Multiple-choice items are an efficient way to assess knowledge and skills, and they can be developed to require the cognitive processing associated with any of the three cognitive categories. In a well-designed multiple-choice item, the stem clearly presents the problem to the student. The stem may be a question or a phrase to be completed by selecting the correct option. The stem is followed by four answer choices (options), labeled A–D, only one of which is correct. Good multiple-choice items have the following characteristics:

1. The stem includes only information needed to make the student's task clear or to set the item in an appropriate context.
2. Options are as short as possible.
3. Options are parallel in structure, syntax, and complexity.
4. Options contain only one correct answer (the key).
5. Options do not contain inadvertent cues to the correct answer, such as repeating a word from the stem in the key or using so-called "specific determiners" (e.g., all, never) in the distractors (incorrect options) to make them incorrect.
6. Options are free of repetitive wording that adds to the reading load.
7. Options are grammatically consistent with the stem.
8. Options are free from overlap.
9. Options are free of language such as "all of the above" and "none of the above."
10. Options are arranged in a logical order (numerically, chronologically).
11. Distractors are plausible, but not so plausible that they could be possible correct answers.

[13] Students without a formal economics courses, but who understand economic concepts and their application, may be able to answer a challenging extended-response item that requires them to explain the effects on price and quantity in a market of a change in costs of production. Those same students, however, are likely to be *unable* to answer a question that asks them to interpret the effects of a shift of the supply curve to the right and will certainly be unable to draw such a graph. However, students who have had a course in economics may answer the first question by drawing a set of curves.

12. Distractors are not only wrong but are based on a misunderstanding of the content. That is, they are designed to reflect the measurement intent of the item, not to trick students into choices that are not central to the economic idea being assessed.

Developing Constructed-Response Items and Scoring Rubrics

The type of constructed-response item–short or extended–that is written should depend on the economic construct that is being assessed. Item writers should draft the scoring rubric as they develop the item so that both the item and the rubric reflect the construct being measured.

In developing the scoring rubric for an item, writers should think about what kind of student responses would show increasing degrees of knowledge and understanding. Writers should sketch condensed sample responses for each score category. They also should include a justification or an explanation for each rubric category description. Doing so will enable development of a clear scoring rubric and will provide guidance for scoring the item.

Short Constructed-Response Items

Some short constructed-response items are written to be scored dichotomously. Short constructed-response items with two scoring categories (i.e., 0 or 1) should measure knowledge and skills in a way that multiple-choice items cannot or should require greater evidence of the depth of students' understanding. Such short constructed-response items are useful when more than one correct answer is possible, when different ways to display an answer are available, or when a brief explanation is required. Item writers should take care that short constructed-response items would not be better or more efficiently structured as multiple-choice items—they should not be simply multiple-choice items without the response options.

Some short constructed-response items are written to be scored on a three-category scale (i.e., 0, 1, or 2). Short constructed-response items with three scoring categories should measure knowledge and skills that require students to go beyond recalling an acceptable answer. These items allow gradations of correctness in a response so that a student can receive some credit for demonstrating partial understanding of a concept or skill.

Item writers must draft a scoring rubric for each short constructed-response item. For dichotomous items, the rubrics should define the following two categories:

0 = Incorrect
1 = Correct

For items with three score categories, the rubrics should define the following categories:

0 = Incorrect
1 = Partially correct
2 = Correct

Items 7 and 8 in Chapter Six are examples of constructed response items with three scoring categories.

Extended Constructed-Response Items

In general, extended constructed-response items ask students to solve a problem by applying and integrating economic concepts, or they require students to analyze an economic situation and explain a concept, or both. Extended constructed-response items usually have five scoring categories:

0 = Incorrect
1 = Minimal
2 = Partial
3 = Satisfactory
4 = Superior

Item writers must develop a draft-scoring rubric that is specific for each extended constructed-response item. The rubric should clearly reflect the measurement intent of the item. The next section describes some requirements for writing scoring rubrics.

Item 9 in Chapter Six is an example of an extended constructed-response with five scoring categories.

Aligning Items and Rubrics

Item writers should refer to the measurement intent of the item when they are developing its scoring rubric. The number of score points in the rubric should be based on the complexity of the item and thus the possible gradations of correctness.

1. Defining the Score Categories: Each score category must be distinct from the others; descriptions of the score categories should clearly reflect increasing understanding and skill in the economic constructs of interest. Distinctions among the categories should suggest the differences in student responses that would fall into each category; the definitions must be clear enough to use for training scorers. Each score level should be supported by the economic intent of the item. Factors unrelated to the measurement intent of the item should not be evaluated in the rubric. For example, if an item is not meant to measure writing skills, then the scoring rubric should be clear that the demonstration of economic understanding in the response does not need to be tied to how well the response is written. However, if an explanation is part of the item

requirement, the rubric should reflect that such explanations should be clear and understandable.

2. Measuring More Than One Concept: If an item is measuring more than one skill or concept, the description of the score categories in the rubric should clearly reflect increasing understanding and achievement in each area. For instance, if the item is measuring students' understanding of a topic both in the national economy and in the international economy, the description of each category in the rubric should explain how students' understanding is evaluated in terms of both topics. If an item requires both an acceptable solution and an explanation, the rubric should show how these two requirements are addressed in each score category.

3. Specifying Response Formats: Unless the item is measuring whether or not a student can use a specified approach to a problem, each score category should allow various approaches to the item. It should be clear in the rubric that different approaches to the item are allowed. Varied approaches may include

    - different economic solutions and/or
    - different formats (e.g., using a graph, table, or chart to solve or explain a solution may be appropriate).

4. Partial Credit: Constructed-response items should be accompanied by detailed rubrics that clearly establish distinguishable levels of performance and describe the performance expectation at each level. A response need not exceed the requirements to get full credit. The rubrics should include examples of specific elements of responses that are likely to appear at each level, but should also describe ways of looking for evidence of reasoning that is correct but still results in an incorrect answer. For example, in a Reasoning item, the student may make an error in the analysis of data but reason to a conclusion in a sound manner. This would be evidence of proper economic reasoning, even though the final answer may be incorrect. The rubric should direct scorers to look at responses in this way and to award an appropriate amount of partial credit. An example follows:

    Suppose the costs of producing a product increase. What will happen in a market and why?

    Correct answer:

4 points	The supply of the product will decrease. As a result of the decrease in supply, there will be a shortage in the market at the existing price. Prices will begin to increase, which in turn causes the quantity demanded to decrease and the quantity supplied to increase. The process will continue until the quantity demanded equals the quantity supplied. Less will be produced in the market at a higher equilibrium price.

3 points	The supply of the product will decrease. As a result of the decrease in supply, there will be a shortage in the market at the existing price. Less will be produced in the market at a higher equilibrium price.
2 points	The supply of the product will decrease. Less will be produced in the market at a higher equilibrium price.
1 point	The supply of the product will decrease.
0 points	No appropriate response

Possible *incorrect* answers that should receive partial credit follow. Although every part of these answers is incorrect (that is, they are opposites of the correct answers), the parts after the first statement are logical consequences of that statement, and the student should receive partial credit for the correct logic.

3 points	The supply of the product will **increase**. (This is incorrect.) As a result of the **increase** in supply, there will be a **surplus** in the market at the existing price. (This is correct based on the first part of the answer.) Prices will begin to **decrease**. (This is correct based on the first part of the answer.) As a result the quantity demanded will **increase** and the quantity supplied will **decrease**. (This is correct based on the first part of the answer.) The process will continue until the quantity demanded equals the quantity supplied. **More** will be produced in the market at a **lower** equilibrium price. (This is correct based on the first part of the answer.)
2 points	The supply of the product will **increase**. As a result of the increase in supply, there will be a surplus in the market at the existing price. More will be produced in the market at a lower equilibrium price.
1 point	The supply of the product will **increase**. More will be produced in the market at a lower equilibrium price.
0 points	The supply of the product will increase or no appropriate response.

ITEM TRYOUTS AND REVIEWS

The assessment development contractor should build careful review and quality control procedures into the assessment development process. Items will go through a large-scale field test to provide critical item-level information for test development. Before field testing, items and scoring rubrics should be reviewed by experts in economics and measurement, including economics teachers and representatives of state education agencies, and by reviewers trained in bias and sensitivity review procedures. After field

testing, the items and the assessment as a whole should be reviewed to make sure that they are as free as possible from irrelevant variables that could interfere with students' demonstrating their economics knowledge and skills. In addition, any items that statistical evaluation reveals to be technically flawed should be eliminated.

Bias and sensitivity reviews are a particularly important part of the assessment development process. Reviewers, including educators and community members who are experts in the schooling or cultural backgrounds of the primary minority and special needs students who will be taking the assessment, evaluate assessment items and associated materials. The reviewers should focus on identifying offensive or stereotypical subject matter and other construct-irrelevant factors in the assessment. They should provide valuable guidance about the context, wording, and structure of items and identify flaws in the items that confound the validity of the inferences for the groups of students they represent. In addition to applying the industry standards for bias and sensitivity, the item development contractors must ensure that "all items selected for use in the National Assessment are free from racial, cultural, gender, or regional bias and are secular, neutral, and non-ideological" (see Principal 6 on page 19). When specific item revisions are recommended to avoid such unwanted characteristics, the item development contractors must also ensure that the revisions do not introduce inaccuracies in the content of the item.

Recruitment of Item Writers

Many of the item writers will be from the Planning and Steering Committees. To maintain continuity of the development process, NAGB policy requires that a minimum of 20 percent of the membership of the Planning and Steering Committees serve on the item writing and review groups. Other members of the item-writing team must have a demonstrable claim to expertise in economics and the teaching of economics to grade 12 students. Writers will be trained on these specifications, on these principles of item writing and review, and on bias and sensitivity issues.

Recruitment and Training of Item Scorers

Scorers of constructed-response items should have background and training in economics. They should have experience with high school students and understand the various ways students learn economics and what the range of grade 12 students know about economics and are able to do with that knowledge.

Training is important for scorers of the constructed-response items. The training program for scorers (raters) of constructed-response items should include an orientation to the NAEP economics assessment process, a review of the various constructed-response items that they will be scoring, and a discussion of each rating scale they will use. When raters are being trained to apply a rating scale, they should examine samples of student work chosen to demonstrate various levels of performance (anchor papers for written response) so that they can understand the range of performance they are likely to encounter. They

should have the opportunity to discuss the samples to help them arrive at a common understanding of the meaning of each point on a rating scale. Raters should practice applying the rating scales, and their ratings should be checked to make certain that they can apply the scoring criteria reliably and validly. The scoring contractor will need to conduct inter-rater reliability checks immediately after training and throughout the scoring process. Procedures for monitoring rater performance are necessary to minimize rater drift and to ensure rater reliability and scoring validity.

Training in cultural sensitivity and instruction not to score performance on the basis of personal bias must be explicit. Specific instructions for scoring responses from English language learners can be found on pages 15-16 in Chapter Three.

CHAPTER FIVE

ITEM SPECIFICATIONS BY CONTENT AREA

This chapter describes the standards and benchmarks in each content area covered by the assessment: the Market Economy, the National Economy, and the International Economy. Although the Framework has organized economics into these component parts, it is likely that the benchmarks and the assessment items built on them will, in many cases, cross these content areas. All assessment items, however, must be assigned the primary classification that seems most relevant, given the standards and benchmarks presented in this chapter.

Sample items are presented in Chapter Six. The sample items illustrate different content areas and cognitive categories. They are not intended to unduly restrict the types of items that can be written, but are meant to illustrate, in a concrete way, the range of skills and knowledge in each content area and provide guidance to item writers. As the specifications are used, the assessment development contractor should augment item writer training with sample items that illustrate particular characteristics of the content areas or show a different way of assessing a cognitive category.

Table 2 shows how the standards are distributed across the three content areas.

Table 2. Standards Distribution Across Content Areas

Standards	The Market Economy	The National Economy	The International Economy
1. Choices and costs	×		
2. Effective decision making	×		
3. Resource allocation methods		×	
4. Incentives	×		
5. Voluntary exchange			×
6. Benefits of trade			×
7. Markets	×		×
8. Prices	×		
9. Competition	×		
10. Institutions	×		
11. Money		×	
12. Interest rates		×	
13. Income	×		
14. Entrepreneurs	×		
15. Investment, productivity, and growth	×	×	×
16. Economic role for government	×	×	
17. Government decision making	×	×	×
18. Gross Domestic Product (GDP)		×	
19. Unemployment and inflation		×	
20. Fiscal and monetary policies		×	

As Table 2 illustrates, most of the standards have been assigned to only one of the three content areas, although a few have been assigned to more than one. It is important for the item writers to note that different benchmarks within these standards are being measured in each of the different content areas. Table 3 illustrates how the different standards within each content area should be distributed. The recommended percentages indicate the proportion of total assessment time, rather than the proportion of total items, that should be allocated to each standard.

Table 3. Percent Distribution of Standards within Content Areas[14]

Standard	Recommended Percent
The Market Economy	
1	6%
2	7%
4	4%
7*	5%
8	4%
9	3%
10	4%
13	4%
14	3%
15*	2%
16*	3%
17*	1%
Total	*45%*
The National Economy	
3	5%
11	5%
12	5%
15*	3%
16*	4%
17*	2%
18	5%
19	5%
20	6%
Total	*40%*
The International Economy	
5	2%
6	6%
7*	2%
15*	3%
17*	1%
Total	*15%*
Grand Total	***100%***

*An asterisk means that a standard is included in more than one content area

[14] Note that the totals for each content area may not sum to the overall percentages due to rounding.

THE MARKET ECONOMY

The core content in this area includes the relevance of limited resources, how individuals and institutions make and evaluate decisions, the role of incentives, how buyers and sellers interact to create markets, how markets allocate resources, and the economic role of government in a market economy.

Economically literate grade 12 students are able to identify what trade-offs they face—what they gain and what they give up when they consider alternatives and make choices. Doing so informs their decisions, helping them choose alternatives that promote their goals as consumers, producers, savers, investors, and citizens. They can identify incentives that affect people's behavior and explain how incentives affect their own behavior.

They are able to describe how the interaction of buyers and sellers in markets influences prices and output levels. They are able to predict how prices change when there is either a shortage or a surplus of the product. They can explain how changes in the supply and demand conditions and in the level of competition in different markets can affect prices and output levels.

Students who understand and are able to use economics can describe the likely effects of plans for education, training, and career options on future earnings and can identify the risks, returns, and other characteristics of entrepreneurship that bear on its attractiveness as a career. They can predict the consequences, the risks, and the potential returns of investment decisions made by individuals, businesses, and governments.

Students are able to describe the roles of economic institutions, such as legal systems, private property, labor unions, and corporations. They can identify and evaluate the benefits and the costs of alternative public policies and assess who enjoys the benefits and who bears the costs and explain why government policies exist.

Content in the Market Economy includes much of what is traditionally described as microeconomics.

Standard 1 Choices and costs

Productive resources are limited. Therefore, people cannot have all the goods and services they want; as a result, they must choose some things and give up others.

1.1.1 The opportunity cost of a choice is the value of the best alternative given up.

1.1.2 Scarcity is the condition of not being able to have all the goods and services one wants. It exists because human wants for goods and services exceed the quantity of goods and services that can be produced from all available resources.

1.1.3 Choices involve trading off the expected value of one opportunity against the expected value of its best alternative.

1.1.4 Choices made by individuals, firms, or government officials often have long-run unintended consequences that can partially or entirely offset the initial effects of their decisions.

1.1.5 Productive resources are the natural resources, human resources, and capital goods available to make goods and services.

Standard 2 Effective decision making

Effective decision making requires comparing the additional costs of alternatives with the additional benefits. Most choices involve doing a little more or a little less of something: few choices are "all or nothing" decisions.

1.2.1 Marginal benefit is the change in total benefit resulting from an action. Marginal cost is the change in total cost resulting from an action.

1.2.2 As long as the marginal benefit of an activity exceeds the marginal cost, people are better off doing more of it; when the marginal cost exceeds the marginal benefit, they are better off doing less of it.

1.2.3 To determine the best level of consumption of a product, people must compare the additional benefits with the additional costs of consuming a little more or a little less.

1.2.4 To produce the profit-maximizing level of output and hire the optimal number of workers and other resources, producers must compare the marginal benefits and marginal costs of producing a little more with the marginal benefits and marginal costs of producing a little less.

1.2.5 To determine the optimal level of a public policy program, voters and government officials must compare the marginal benefit and marginal cost of providing a little more or a little less of the program's services.

1.2.6 The time value of money refers to the relationship between the length of time money is invested and its growth due to the compounding of gains.

Standard 4 Incentives

People respond to positive and negative incentives.

1.4.1 Changes in incentives cause people to change their behavior in predictable ways.

1.4.2 Acting as consumers, producers, workers, savers, investors, and citizens, people respond to incentives in order to allocate their scarce resources in ways that provide the highest possible returns to them.

Standard 7 Markets

Markets exist when buyers and sellers interact. This interaction determines market prices and thereby allocates scarce goods and services.

1.7.1 A market exists whenever buyers and sellers exchange goods and services.

1.7.2 Market prices are determined through the buying and selling decisions made by buyers and sellers.

1.7.3 The equilibrium price of a good or a service is the one price at which quantity supplied equals quantity demanded.

1.7.4 If a price is above the equilibrium price, it will fall, causing sellers to produce less and buyers to purchase more; if it is below the equilibrium price, the price will rise, causing sellers to produce more and buyers to purchase less.

1.7.5 Shortages of a product usually result in price increases in a market economy; surpluses usually result in price decreases.

Standard 8 Prices

Prices send signals and provide incentives to buyers and sellers. When supply or demand changes, market prices adjust, affecting incentives.

1.8.1 An increase in the price of a good or a service encourages people to look for substitutes, causing the quantity demanded to decrease, and vice versa. This relationship between price and quantity demanded, known as the law of demand, exists as long as other factors influencing demand do not change.

1.8.2 An increase in the price of a good or a service enables producers to cover higher per-unit costs, causing the quantity supplied to increase, and vice versa. This relationship between price and quantity supplied is normally true as long as other factors influencing the costs of production and supply do not change.

1.8.3 Demand for a product changes when there is a change in consumers' incomes or preferences, in the prices of related goods or services, or in the number of consumers in a market.

1.8.4 Supply of a product changes when there are changes in the prices of the productive resources used to make the good or the service, the technology used to make the good or the service, the profit opportunities available to producers by selling other goods or services, or the number of sellers in a market.

1.8.5 Elasticity describes the degree to which buyers and sellers respond to price changes.

1.8.6 Changes in supply or demand cause relative prices to change; in turn, buyers and sellers adjust their purchase and sales decisions.

1.8.7 Government-enforced price ceilings set below the equilibrium price and government-enforced price floors set above the equilibrium price distort price signals and incentives to producers and consumers. The price ceilings cause persistent shortages, whereas the price floors cause persistent surpluses.

Standard 9 Competition

Competition among sellers lowers costs and prices and encourages producers to produce more of what consumers are willing and able to buy. Competition among buyers increases prices and allocates goods and services to those people who are willing and able to pay the most for them.

1.9.1 The level of competition in an industry is affected by the ease with which new producers can enter the industry and by consumers' information about the availability, price, and quantity of substitute goods and services.

1.9.2 The pursuit of self-interest in competitive markets generally leads to choices and behavior that also promote the national level of economic well-being.

1.9.3 When competition is limited, producers are able to gain more control of the market and the prices they set.

1.9.4 The introduction of new products and production methods by entrepreneurs is an important form of competition and is a source of technological progress and economic growth.

Standard 10 Institutions

Institutions evolve in market economies to help individuals and groups accomplish their goals. Banks, labor unions, corporations, legal systems, and not-for-profit organizations are examples of important institutions. Another institution, clearly defined and well-enforced property rights, is essential to a market economy.

1.10.1 Through the process of collective bargaining with employers, labor unions represent some workers in negotiations involving wages, fringe benefits, and work rules.

1.10.2 Incorporation allows firms to accumulate sufficient financial capital to make large-scale investments and achieve economies of scale. Incorporation also reduces the risk to investors by limiting stockholders' liability to their share of ownership of the corporation.

1.10.3 Banks and other financial institutions channel funds from savers to borrowers and investors.

1.10.4 Property rights, contract enforcement, standards for weights and measures, and liability rules affect incentives for people to produce and exchange goods and services.

Standard 13 Income

Income for most people is determined by the market value of the productive resources they sell. What workers earn depends, primarily, on the market value of what they produce and how much they add to the production of it.

1.13.1 Employers are willing to pay wages and salaries to workers because they expect to sell the goods and services those workers produce at prices high enough to cover the wages and salaries and all the other costs of production.

1.13.2 More productive workers are likely to be of greater value to employers and earn higher wages than less productive workers.

1.13.3 People's incomes, in part, reflect choices they have made about education, training, skill development, and careers. People with few marketable skills are more likely to be poor.

1.13.4 Changes in the prices for productive resources affect the incomes of the owners of those productive resources and the combination of those resources used by firms.

1.13.5 Changes in demand for specific goods and services often affect the incomes of the workers who make those goods and services.

Standard 14 Entrepreneurs

Entrepreneurs are people who take calculated risks in organizing productive resources to make goods and services. Profit is an important incentive that leads entrepreneurs to accept the risks of business failure.

1.14.1 Entrepreneurs are individuals who take calculated risks in order to start new businesses and develop innovative products and processes.

1.14.2 Entrepreneurs accept the risk of organizing resources to produce goods and services, and they hope to earn profits.

1.14.3 Entrepreneurs and other sellers earn profits when buyers purchase the products they sell at prices high enough to cover the costs of production; they incur losses when buyers do not purchase the products they sell at prices high enough to cover the costs of production.

Standard 15 Investment, productivity, and growth

Investment in factories, machinery, and new technology and in the health, education, and training of people can raise future standards of living.

1.15.1 Investments in physical and human capital can increase productivity, but such investments entail opportunity costs and economic risks. Investing in new physical or human capital involves a trade-off of lower current consumption in anticipation of greater future production and consumption.

1.15.2 Workers can improve their productivity by improving their human capital and by using physical capital such as tools and machinery.

Standard 16 Economic role for government

Government has an economic role in a market economy whenever the benefits of a government policy outweigh its costs. Governments often provide national defense, address environmental concerns, define and protect property rights, and through regulation attempt to make markets more competitive. Most government policies also redistribute income.

1.16.1 Markets do not allocate resources effectively if (1) property rights are not clearly defined or enforced, (2) externalities (spillover effects) affecting large numbers of people are associated with the production or consumption of a product, or (3) markets are not competitive.

1.16.2 An important role for government in the economy is to define, establish, and enforce property rights. A property right to a good or a service includes the right to exclude others from using the good or the service and the right to transfer the ownership or the use of the resource to others.

1.16.3 When a price fails to reflect all the benefits of a product, too little of the product is produced and consumed. When a price fails to reflect all the costs of a product, too much of it is produced and consumed. Government can use subsidies to help correct for insufficient output; it can use taxes to help correct for excessive output; or it can regulate output directly to correct for over- or underproduction or for over- or underconsumption of a product.

1.16.4 Externalities exist when some of the costs and the benefits associated with production and consumption fall on someone other than the producers or the consumers of the product.

1.16.5 Governments provide an alternative method to markets for supplying goods and services when markets fail and when it appears that the benefits to society of doing so outweigh the costs to society.

Standard 17 Government decision making

The costs of government policies may exceed the benefits. This may occur because social goals other than economic efficiency are being pursued; because of incentives facing voters, government officials, and government employees; or because of actions pursued through government and legal channels by special-interest groups that can impose costs on the general public.

1.17.1 Price controls are often advocated by special-interest groups. Price controls reduce the quantity of goods and services consumed or produced, thus depriving consumers of some goods and services whose value would exceed their cost.

THE NATIONAL ECONOMY

The National Economy content area includes the concepts, terminology, and data used to identify and describe inflation, unemployment, output, and growth; the factors that cause changes in those conditions; the role of money and interest rates in an economy; and the mechanics and the appropriate uses of Federal Reserve monetary policies and federal government fiscal policies.

Economically literate grade 12 students are able to describe how economies use different systems of allocating goods and services and can compare the benefits and the costs of different methods. Students can identify the various economic roles that governments play as providers of goods and services.

Students can explain the role of money in an economy and identify interest rates as the prices of borrowing or lending money. They can give examples of situations in which they might pay or receive interest and how they would react to changes in interest rates.

Students can identify the effects of technological change and investment on gross domestic product. They are able to explain the function of taxes and how taxes may redistribute income. They can interpret media reports about current economic conditions and explain how these conditions can influence decisions made by consumers, producers, and governments.

Students are able to make informed decisions by anticipating the consequences of inflation and unemployment. They can explain the macroeconomic policies of the federal government and the Federal Reserve System, under what conditions the policy decisions are likely to change, and the effects of those changes on themselves and others.

Content in the National Economy includes much of what is traditionally described as macroeconomics.

Standard 3 Resource allocation methods

Different methods can be used to allocate goods and services. People acting individually or collectively through government must choose which methods to use to allocate different kinds of goods and services.

2.3.1 People in all economies must answer three basic questions: What goods and services will be produced? How will these goods and services be produced? Who will consume them?

2.3.2 National economies vary in the extent to which they rely on government directives (central planning) and signals from private markets to allocate scarce goods, services, and productive resources.

2.3.3 A comparison of the benefits and the costs of different allocation methods in order to choose the method that is most appropriate for a specific problem can result in more effective allocations and a more effective overall allocation system.

Standard 11 Money

Money makes it easier to trade, borrow, save, invest, and compare the value of goods and services.

2.11.1 Money is anything widely accepted as final payment for goods and services.

2.11.2 The basic money supply in the United States consists of currency, coins, and checking account deposits.

2.11.3 In many economies, when banks make loans, the money supply increases; when loans are paid off, the money supply decreases.

Standard 12 Interest rates

Interest rates, adjusted for inflation, rise and fall to balance the amount saved with the amount borrowed, thus affecting the allocation of scarce resources between present and future uses.

2.12.1 An interest rate is a price of money that is borrowed or saved.

2.12.2 Like other prices, interest rates are determined by the forces of supply and demand.

2.12.3 The real interest rate is the nominal or current market interest rate minus the expected rate of inflation.

2.12.4 Higher real interest rates provide incentives for people to save more and to borrow less. Lower real interest rates provide incentives for people to save less and to borrow more.

2.12.5 Real interest rates usually are positive because people must be compensated for deferring the use of resources from the present into the future.

2.12.6 Riskier loans command higher interest rates than do safer loans because of the greater chance of default on the repayment of risky loans.

Standard 15 Investment, productivity, and growth

Investment in factories, machinery, and new technology and in the health, education, and training of people can raise future standards of living.

2.15.1 Productivity is measured by dividing output (goods and services) by the number of inputs used to produce the output. A change in productivity is a change in output relative to input.

2.15.2 The rate of productivity increase in an economy is strongly affected by the incentives that reward successful innovation and investments in research and development and in physical and human capital.

2.15.3 Increases in productivity result from advances in technology and increases in physical and human capital.

2.15.4 Economic growth is a sustained rise in a nation's production of goods and services. It results from investments in human and physical capital, research and development, technological change, and improved institutional arrangements and incentives.

2.15.5 Economic growth creates new employment and profit opportunities in some industries, but growth reduces opportunity in others.

Standard 16 Economic role for government

Government has an economic role in a market economy whenever the benefits of a government policy outweigh its costs. Governments often provide national defense, address environmental concerns, define and protect property rights, and through regulation attempt to make markets more competitive. Most government policies also redistribute income.

2.16.1 Governments pay for the goods and services they use or provide by taxing or borrowing from people.

2.16.2 Governments often redistribute income directly in response to individuals or interest groups who are not satisfied with the income distribution resulting from markets; governments also redistribute income indirectly as side effects of other government actions that affect prices or output levels for various goods and services.

2.16.3 Most federal tax revenue comes from personal income and payroll taxes. Payments to social security recipients, the costs of national defense, medical expenditures, and interest payments on the national debt constitute the bulk of federal government spending.

2.16.4 Different tax structures affect consumers and producers differently.

Standard 17 Government decision making

The costs of government policies may exceed the benefits. This may occur because social goals other than economic efficiency are being pursued; because of incentives facing voters, government officials, and government employees; or because of actions pursued through government and legal channels by special-interest groups that can impose costs on the general public.

2.17.1 Incentives exist for political leaders to favor programs that entail immediate benefits and future costs; few incentives favor programs promising immediate costs and future benefits, even though the latter programs are sometimes economically more effective than the former programs.

Standard 18 Gross Domestic Product (GDP)

A nation's overall levels of income, employment, and prices are determined by the interaction of spending and production decisions made by all households, firms, government agencies, and others in the economy.

2.18.1 One person's spending is other people's income. When consumers make purchases, goods and services are transferred from businesses to households in exchange for money payments. That money is used in turn by businesses to pay for natural resources, human resources, and capital goods and to pay taxes.

2.18.2 Gross Domestic Product (GDP) is a basic measure of a nation's economic output and income. It is the total market value, measured in dollars, of all final goods and services produced in the economy in one year.

2.18.3 Nominal GDP is measured in current dollars; thus, an increase in GDP may reflect not only increases in the production of goods and services, but also increases in prices. GDP adjusted for price changes is called real GDP. Real GDP per capita is a measure that permits comparisons of material living standards over time and among people in different nations.

2.18.4 The potential level of real GDP for a nation is determined by the quantity and quality of its natural resources, the size and skills of its labor force, and the size and quality of its stock of capital resources.

2.18.5 When desired expenditures for consumption, investment, government spending, and net exports are greater than the value of a nation's output of final goods and services, GDP rises and inflation occurs and/or employment rises. When desired expenditures for consumption, investment, government spending, and net exports are less than the value of a nation's output of final goods and services, GDP decreases and inflation and/or employment decreases.

Standard 19 Unemployment and inflation

Unemployment imposes costs on individuals and on nations. Unexpected inflation imposes costs on many people and benefits some others because it arbitrarily redistributes purchasing power. Inflation can reduce the rate of growth of national living standards because individuals and organizations use resources to protect themselves against the uncertainty of future prices.

2.19.1 The unemployment rate is the percentage of the labor force that is willing and able to work, does not currently have a job, and is actively looking for work.

2.19.2 Unemployment rates differ for people of different ages, races, and gender. This reflects differences in work experience, education, training, and skills, as well as discrimination.

2.19.3 Unemployment can be caused by people changing jobs, by seasonal fluctuations in demand, by changes in the skills needed by employers, or by cyclical fluctuations in the level of national spending.

2.19.4 Unemployment has costs for society as well as for individuals. When unemployment is substantial, the economy will not produce as much as it could.

2.19.5 Inflation is an increase in the general level of prices. It reduces the value of money.

2.19.6 When people's incomes increase more slowly than the inflation rate, their purchasing power declines.

2.19.7 The consumer price index (CPI) is the most commonly used measure of price-level changes. It can be used to compare the price level in one year with price levels in earlier or later periods.

2.19.8 The costs of inflation are different for different groups of people. Unexpected inflation hurts savers and people on fixed incomes; it helps people who have borrowed money at fixed rates of interest.

Standard 20 Fiscal and monetary policies

Federal government budgetary policy and the Federal Reserve System's monetary policy influence the overall levels of employment, output, and prices.

2.20.1 Fiscal policies are decisions to change spending and tax levels by the federal government. These decisions are adopted to influence national levels of output, employment, and prices.

2.20.2 In the short run, increasing federal spending and/or reducing taxes can promote more employment and output, but these policies also put upward pressure on the price level and interest rates. Decreased federal spending and/or increased taxes tend to lower price levels and interest rates, but they reduce employment and output levels in the short run.

2.20.3 In the long run, the interest-rate effects of fiscal policies lead to changes in private investment spending by businesses and individuals that partially, if not entirely, offset the output and employment effects of fiscal policy.

2.20.4 The federal government's annual budget is balanced when its revenues from taxes and user fees equal its expenditures. The government runs a budget deficit when its expenditures exceed its revenues. The government runs a surplus when its revenues exceed its expenditures.

2.20.5 When the government runs a budget deficit, it must borrow from individuals, corporations, or financial institutions to finance that deficit.

2.20.6 The national debt is the total amount of money the federal government owes. This is the sum of all its past annual deficits and surpluses. The government pays interest on the money it borrows to finance the national debt.

2.20.7 In the long run, inflation results from increases in a nation's money supply that exceed increases in its output of goods and services.

2.20.8 Monetary policies are decisions by the Federal Reserve System that lead to changes in the supply of money and the availability of credit. Changes in the money supply can influence overall levels of spending, employment, and prices in the economy by inducing changes in interest rates charged for credit and by affecting the levels of personal and business investment spending.

2.20.9 The major monetary policy tool that the Federal Reserve System uses is open market purchases or sales of government securities. Other policy tools used by the Federal Reserve System include increasing or decreasing the discount rate charged on loans it makes to banks (and other depository institutions) and raising or lowering reserve requirements for those same financial institutions.

THE INTERNATIONAL ECONOMY

The content in this area includes the reasons for individuals and businesses to specialize and trade; the rationale for specialization and trade across international borders; and the comparison of the benefits and costs of that specialization and resulting trade for consumers, producers, and governments.

Economically literate grade 12 students are able to explain how voluntary exchange, whether in a domestic or an international market, is undertaken because both parties in the exchange expect to benefit. Students can explain how they benefit themselves and others by developing special skills and strengths. They are able to negotiate exchanges and identify the gains to themselves and others.

They can compare the benefits and costs of policies that alter trade barriers between nations, such as tariffs and quotas. In addition, they are able to identify who bears those costs and who receives the benefits.

They are able to explain why exchange rates change and can predict the effects of those changes on themselves and others. They can explain how investment, technological change, education, and incentive structures contribute to differences in economic growth and standards of living among countries.

Standard 5 Voluntary exchange

Voluntary exchange occurs only when all participating parties expect to gain. This is true for trade among individuals or organizations within a nation and among individuals or organizations in different nations.

3.5.1 Voluntary exchange of goods and services takes place internationally because people or organizations expect to be better off.

3.5.2 When imports are restricted by public policies, consumers pay higher prices and job opportunities and profits in exporting firms decrease.

Standard 6 Benefits of trade

When individuals, regions, and nations specialize in what they can produce at the lowest cost and then trade with others, both production and consumption increase.

3.6.1 Like trade among individuals within one country, international trade promotes specialization and division of labor and increases the productivity of labor, output, and consumption.

3.6.2 Greater specialization leads to increased interdependence among producers and consumers. As a result of growing international economic interdependence, economic conditions and policies in one nation increasingly affect economic conditions and policies in other nations.

3.6.3 Comparative advantage is the primary motivating factor driving international trade.

3.6.4 Individuals and nations have a comparative advantage in the production of goods or services if they can produce a product at a lower opportunity cost than other individuals or nations.

3.6.5 Comparative advantages change over time due to changes in factors such as abundance of resources, resource prices, and international institutions.

Standard 7 Markets

Markets exist when buyers and sellers interact. This interaction determines market prices and thereby allocates scarce goods and services.

3.7.1 An exchange rate is the price of one nation's currency in terms of another nation's currency. Like other prices, exchange rates are determined by the interactions of supply and demand. Foreign exchange markets allocate international currencies.

3.7.2 When exchange rates fluctuate, the prices of exports and imports change, and some groups gain while others lose in each country.

Standard 15 Investment, productivity, and growth

Investment in factories, machinery, and new technology and in the health, education, and training of people can raise future standards of living.

3.15.1 Economic growth varies across countries because of differences in human and physical capital investments, technologies, and institutional arrangements and incentives.

3.15.2 Economic growth has been the primary vehicle for alleviating poverty and raising standards of living.

Standard 17 Government decision making

The costs of government policies may exceed the benefits. This may occur because social goals other than economic efficiency are being pursued; because of incentives facing voters, government officials, and government employees; or because of actions pursued through government and legal channels by special-interest groups that can impose costs on the general public.

3.17.1 Although barriers to international trade usually impose more costs than benefits, they are often advocated by people and groups who expect to gain substantially from them. Because the costs of these barriers are typically spread over a large number of people, each of whom pays only a little and may not recognize the cost, policies supporting trade barriers are often adopted through the political process.

CHAPTER SIX

SAMPLE ITEMS

This chapter contains the items to which reference is made in Chapter Four.

Item Number	Content Area	Benchmark	Cognitive Category	Context	Format	Key
1	National	2.19.3	Knowing	None	MC	D

Which of the following is an example of structural unemployment?

A. A computer programmer who quits her job to go back school
B. A construction worker who loses his job in the winter
C. A tour bus driver who loses her job during a recession
D. A factory worker who is replaced by a robot

Item Number	Content Area	Benchmark	Cognitive Category	Context	Format	Key
2	Market	1.8.4	Applying	Business	MC	C

If farmers grow twice as much lettuce this year as last year, what is most likely to happen?

A. More farmers will plant lettuce next year.
B. People will demand less lettuce.
C. The price of lettuce will decrease.
D. Farmers' incomes will increase.

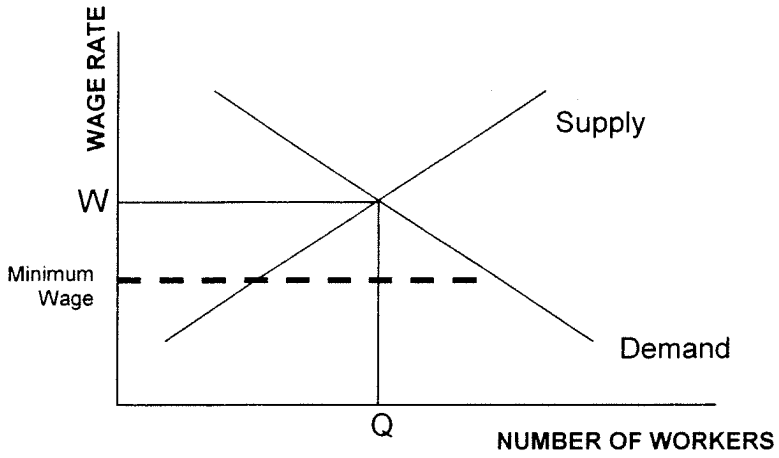
Item Number	Content Area	Benchmark	Cognitive Category	Context	Format	Key
3	International	3.7.1	Reasoning	None	MC	C

Which of the following is most likely to happen if Canadians increase their imports from Mexico?

A. The gross domestic product of Mexico will decrease.
B. The gross domestic product of Canada will increase.
C. The demand for Mexico's currency will increase.
D. The demand for Mexico's currency will decrease.

Item Number	Content Area	Benchmark	Cognitive Category	Context	Format	Key
4	Market	1.8.7	Applying	Public	MC	C

THE MARKET FOR LOW-SKILL LABOR



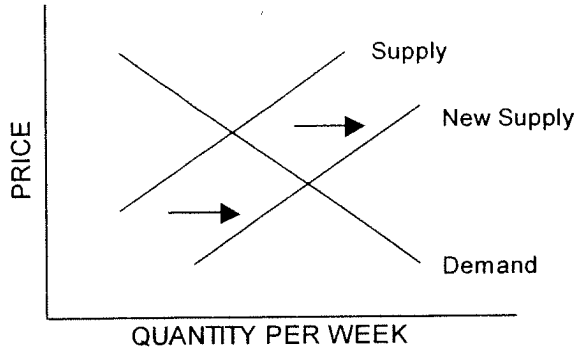
Which of the following will occur as a result of the minimum wage being set as indicated on the graph?

A. There will be a shortage of low-skilled labor.
B. There will be a surplus of low-skilled labor.
C. The number of low-skilled workers hired will be Q.
D. Unemployment will increase.

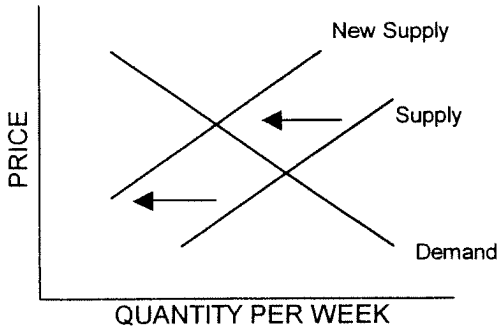
Item Number	Content Area	Benchmark	Cognitive Category	Context	Format	Key
5	Market	1.8.4	Knowing	Business	MC	B

Suppose that an industry experiences significant increases in the prices of the inputs used in its manufacturing processes. Which of the following graphs represents the effects in the market for the industry's products?

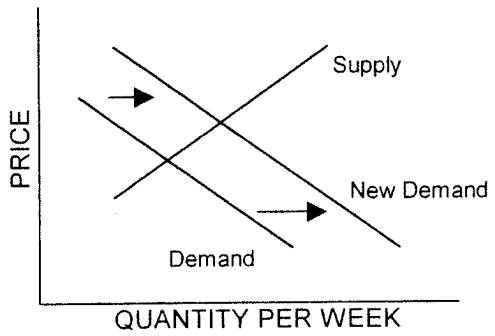
A.



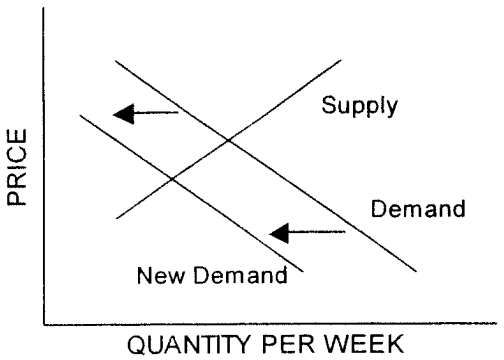
B.



C.



D.



Item Number	Content Area	Benchmark	Cognitive Category	Context	Format	Key
6	Market	1.1.1	Applying	Individual and household	MC	A

The Bodor family wants to rent a larger apartment and would like to buy new car. They found two apartments that would meet their needs but one costs more and has several more desirable features. The Bodors' current financial condition would enable them to rent either the more expensive apartment and not buy a new car or to rent the less expensive apartment and buy a new car. The Bodors decided to rent the more expensive apartment. What is the opportunity cost of the decision?

A. Both the new car and the less expensive apartment
B. The new car only
C. The less expensive apartment only
D. The more expensive apartment only

Item Number	Content Area	Benchmark	Cognitive Category	Context	Format	Key
7	Market	1.8.6	Applying	Individual and household	SCR	Rubric

Five years ago, Juan purchased his first personal computer for $3,000. Juan recently purchased a new computer for $2,000 that is 150 times more powerful than his original computer. He also learned that many more people are buying computers now. Assume that either supply or demand, but not both, has changed. Which one most likely changed and did it increase or decrease? Explain your answer.

2 points — The response states that it is supply that has changed and that it has increased. An increase in supply will result in an increase in the quantity sold and a decrease in the price of the good which is what occurred here.

1 point — The response states that it is supply that has changed and that it has increased.

0 points — No appropriate response

Item Number	Content Area	Benchmark	Cognitive Category	Context	Format	Key
8	Market	1.16.3	Reasoning	Public	SCR	Rubric

Imagine that the following news headline appears in your local newspaper:

> Hazardous Waste Discovered Behind Waterfront Chemical Factory:
> City Council to Discuss Removal

Explain how economic analysis can be used to help the city decide whether to remove the waste or leave it behind the factory.

2 points — The response states that if the benefits of removing the hazardous waste outweigh the costs, it should be removed, and if the benefits of removing the waste are less than the costs then it should not be removed.

1 point — The response states that the decision should be based on a comparison of the benefits and costs of removing the hazardous waste.

0 points — No appropriate response

Item Number	Content Area	Benchmark	Cognitive Category	Context	Format	Key
9	Market	1.8.6	Reasoning	Business	ECR	Rubric

Insects reduce the corn crop in Kansas one year. At the same time, health experts report that eating corn lowers the risk of certain common diseases. Explain how and why the equilibrium price and quantity of corn might change.

(The purpose of the scoring rubrics in the Specifications document is to illustrate the type of response needed to obtain a given number of points. Examples of alternative rubrics for 3 and 2 points are listed below. Actual rubrics used in scoring are more elaborate and will include a number of possible options for each of the point totals.)

4 points — The response states that the reduction in supply causes the price to increase and the quantity to decrease and that the increase in demand causes the price to increase and the quantity to increase. The price must increase. The quantity may either increase or decrease depending upon whether the change in supply or demand is larger. Or the response states

that one cannot tell about the change in quantity as the two events have opposite effects on quantity.

3 points — The response states that the reduction in supply causes the price to increase and the quantity to decrease and that the increase in demand causes the price to increase and the quantity to increase. The price must increase.

OR

The response states that the reduction in supply causes the price to increase and the quantity to decrease and that the increase in demand causes the price to increase and the quantity to increase. The quantity may either increase or decrease depending upon whether the change in supply or demand is larger.

2 points — The response states that the reduction in supply causes the price to increase and the quantity to decrease or that an increase in demand will cause the price to increase and the quantity to increase.

OR

The response states that there is a decrease in supply and an increase in demand. The response does not make a final conclusion about changes in price and quantity.

1 point — The response states that there is a decrease in supply or an increase in demand.

0 points — No appropriate response

APPENDIX A

FRAMEWORK DEVELOPMENT COMMITTEES, MANAGEMENT TEAM, AND TASK LEADERS

STEERING COMMITTEE

Robin Bartlett
Professor of Economics
Denison University
Granville, OH

Patricia Concannon
Social Studies Consultant
New Mexico State Department of Education
Santa Fe, NM

Robert Costrell
Director of Research and Development
Executive Office for Administration and Finance
The Commonwealth of Massachusetts
Boston, MA
(on leave from the University of Massachusetts-Amherst)

Dara Duguay
Executive Director
Jump$tart Coalition for Personal Financial Literacy
Washington, DC

Robert Duvall
President and Chief Executive Officer
National Council on Economic Education
New York, NY

Pete Harder
Consultant
(Former Senior Vice President—Education)
Junior Achievement
Colorado Springs, CO

Betty Lin-Fisher
Consumer Reporter and Columnist
Akron Beacon Journal
Akron, OH

Sally Meek
Teacher
Plano West Senior High School
Plano, TX

Diane Oakley
Vice President
TIAA-CREF
Washington, DC

Bill Odom
Former Chairman and Chief Executive Officer
Ford Motor Credit Company
Lake City, SC

Dennis Placone
Associate Professor of Economics
Clemson University
Center for Economics Education
Clemson, SC

Gary Stern
President and Chief Executive Officer
Federal Reserve Bank of Minneapolis
Minneapolis, MN

Bob Troyer
Assistant Principal
West Lafayette Junior-Senior High School
West Lafayette, IN

Gary Walton
President
Foundation for Teaching Economics
Professor of Economics
University of California, Davis
Davis, CA

Bob Wynn
Financial Education Officer
Wisconsin Department of Financial Institutions
Madison, WI

PLANNING COMMITTEE

Rosella Bannister
Principal Consultant
Bannister Financial Education Services
Ann Arbor, MI

Nancy I. Brown
Principal Consultant
Brown & Associates, LLC
Colorado Springs, CO

Sterlind S. Burke
Principal
Patuxent Valley Middle School
Jessup, MD

Fredrick Czarra
Consultant
Council of Chief State School Officers
Washington, DC

Rae Jean B. Goodman
Professor of Economics
United States Naval Academy
Annapolis, MD

Daniel Gregg
Social Studies Consultant
Connecticut State Department of Education
Bureau of Curriculum and Instruction
Hartford, CT

Gail Mitchell Hoyt
Associate Professor of Economics
University of Kentucky
College of Business and Economics
Lexington, KY

Joy Joyce
Teacher
Willowbrook High School
Villa Park, IL

Don R. Leet
Professor of Economics and
Director of the Center for Economic Education
California State University, Fresno
Fresno, CA

Darrell Luzzo
Senior Vice President—Education
(Former Vice President for Education
Research, Evaluation, and Outreach)
Junior Achievement
Colorado Springs, CO

Richard MacDonald
Assistant Professor of Economics and Assistant
Director of the Center for Economic
Education
St. Cloud State University
St. Cloud, MN

Sarapage McCorkle
Director
Center for Entrepreneurship and Economic
Education
University of Missouri - St. Louis
St. Louis, MO

Karl Ochi
Teacher
George Washington High School
San Francisco, CA

Timothy O'Driscoll
Teacher
Arrowhead High School
Hartland, WI

Joann Prewitt
Social Studies Assessment Specialist
Delaware Department of Education
Dover, DE

Kathryn Ratté
Consultant in Economic Education
Foundation for Teaching Economics
Littleton, CO

Robert Strom
Teaching and Research Fellow
Ewing Marion Kauffman Foundation
Kansas City, MO

George M. Vredeveld
Director
Greater Cincinnati Center for Economic Education
University of Cincinnati
Cincinnati, OH

Scott Oppler
Liaison from the Technical Advisory Panel to the Planning Committee

William (Bill) Walstad
Liaison from the Technical Advisory Panel to the Planning Committee

TECHNICAL ADVISORY PANEL

Ruth Childs
Assistant Professor
Ontario Institute for Studies in Education
Of the University of Toronto (OISE/UT)
Toronto, Ontario
Canada

Huynh Huynh
Professor of Education and Affiliated
Professor of Statistics
College of Education
University of South Carolina
Columbia, SC

John Olson
Assistant Director
State Education Assessment Center
Council of Chief State School Officers
Washington DC

William (Bill) Walstad
Professor of Economics
University of Nebraska-Lincoln
Department of Economics
Lincoln, NE

MANAGEMENT TEAM

Stephen Buckles
Project Associate
Department of Economics
Vanderbilt University
Nashville, TN

Mary Crovo
Project Officer
National Assessment Governing Board
Washington, DC

Stephen Klein
Co-project Director
Co-chair of Steering Committee
American Institutes for Research
Washington, DC

Wayne Martin
Co-chair of Steering Committee
Council of Chief State School Officers
Washington, DC

Claire Melican
Co-chair of Planning Committee
National Council on Economic Education
New York, NY

Julia (Judy) Mitchell
Co-project Director
Co-chair of Planning Committee
American Institutes for Research
Palo Alto, CA

Scott Oppler
Chair of Technical Advisory Panel
Liaison to the Planning Committee
American Institutes for Research
Washington, DC

Task Leaders

Daniel Conrad
Web Site Development Task Leader and
Project Associate
American Institutes for Research
Washington, DC

Elizabeth Greenberg
Background Questionnaires Task Leader
American Institutes for Research
Washington, DC

APPENDIX B

NAEP ECONOMICS PRELIMINARY ACHIEVEMENT LEVEL DESCRIPTIONS GRADE 12

Basic

Students performing at the Basic level of achievement should be able to identify, recall, and recognize economic concepts such as scarcity, choices, price, supply and demand, competition, inflation, unemployment, imports and exports, and trade.

They should be able to describe and explain the relationship between economic concepts. Examples include relationships between inflation and purchasing power, taxes and government spending, unemployment and consumption, price and quantity in supply and demand, trade and specialization, interest rates and loan payment levels, and human capital investment and income.

Students at the Basic level should be able to use data and information to identify an economic outcome. For example, students should be able to predict the effects of a natural disaster on the quantity or price of a product or the effect of a factory closing on a local economy. Students should be able to identify the potential for greater return when given the descriptions of two financial assets.

Proficient

Students performing at the Proficient level should be able to identify, recall, and recognize economic concepts and terms such as costs and benefits in decision-making, responses to incentives, the mechanics of monetary and fiscal policy, trade barriers, exchange rates, and factors that influence economic growth.

Students should be able to demonstrate their understanding of economic ideas and terms by explaining the relationship between a real-world economic situation and its underlying economic concepts. For example, given a description of two industries, students should be able to explain the relevant characteristics that distinguish them. Students should be able to compare and contrast the effects of monetary and fiscal policy on the price level and output. They also should be able to identify the effects of changes in interest rates on individuals' decisions to finance the purchase of a car or home.

At the Proficient level, students should be able to use economic data, information, and concepts to solve problems, evaluate issues, and interpret situations. For example, students should be able to explain the appropriate monetary and fiscal policies for a given set of economic data. Or, they should be able to explain the cause and effect of an increase in taxes on the consumption of a specific product. In the international arena, students should be able to interpret the effect of a new tariff on employment in a domestic industry.

Advanced

Students at the Advanced level should be able to identify, recall, and recognize economic terms such as real interest rate, elasticity, property rights, and comparative advantage; and concepts such as present and future values of money, market structure, and real gross domestic product in the long run.

At the Advanced level, students should be able to analyze economic data and information, and to apply the economic concepts to real-world situations. For example, by applying tools such as aggregate supply and aggregate demand analysis, students should be able to explain what happens to real gross domestic product during a business cycle. Students should be able to interpret economic trends and to apply data to future personal investment options.

Students at the Advanced level should be able to analyze data, determine trends, and make economic projections. Students should be able to reason economically using a variety of tools including charts and graphs, computations, and written explanations. For example, students should be able to analyze the rationale for an entrepreneur to start a new business and the subsequent changes in the market, including the number of firms, prices, profit, and output. Students should be able to extend the analysis of monetary and fiscal policy options to include the effects on exchange rates and international trade.

APPENDIX C

INAPPROPRIATE TERMINOLOGY FOR ITEM WRITERS

The following terms should *not* be used in assessment items. In some cases there are suggested alternatives for testing the concepts that those terms represent.

Terms NOT to be used	**Alternatives**
Absolute advantage	
Balance of trade	
Circular flow	
Double coincidence of wants	
Expected Value	
Externalities and spillover effects	
Macroeconomics	
Marginal analysis	
Marginal benefit/cost	Use "additional" or "extra" or "changes in"
Market-clearing price	Use "equilibrium price"
Microeconomics	
Monopolistic competition	
NAIRU	
Natural monopoly	
Natural rate of unemployment	
Net exports	
Nominal	
Oligopoly	
Potential GDP	Use "real gross domestic product in the long run"
Price elasticity of demand	
Production possibility frontier or curve	
Time value of money	
Transaction costs	

Additionally, item writers also should not test specifically the differences between "quantity demanded" and "demand," and between "quantity supplied" and "supply." The use of the term "open market operations" is permissible, but the preference is that item writers use "purchase or sales of bonds." "Gross domestic product" should be used rather than "GDP."

APPENDIX D

NATIONAL ASSESSMENT GOVERNING BOARD NAEP ITEM DEVELOPMENT AND REVIEW POLICY STATEMENT

It is the policy of the National Assessment Governing Board to require the highest standards of fairness, accuracy, and technical quality in the design, construction, and final approval of all test questions and assessments developed and administered under the National Assessment of Educational Progress (NAEP). All NAEP test questions or items must be designed and constructed to reflect carefully the assessment objectives approved by the National Assessment Governing Board. The final assessments shall adhere to the requirements outlined in the following Guiding Principles, Policies and Procedures for NAEP Item Development and Review.

The Governing Board's Assessment Development Committee, with assistance from other Board members as needed, shall be responsible for reviewing and approving NAEP test questions at several stages during the development cycle. In so doing, the Guiding Principles, Policies and Procedures must be adhered to rigorously.

INTRODUCTION

The No Child Left Behind Act of 2001 (P.L. 107-110) contains a number of important provisions regarding item development and review for the National Assessment of Educational Progress (NAEP). The legislation requires that:

- "the purpose [of NAEP] is to provide...a fair and accurate measurement of student academic achievement"
- "[NAEP shall]...use widely accepted professional testing standards, objectively measure academic achievement, knowledge, and skills, and ensure that any academic assessment authorized....be tests that do not evaluate or assess personal or family beliefs and attitudes or publicly disclose personally identifiable information;"
- "[NAEP shall]...only collect information that is directly related to the appraisal of academic achievement, and to the fair and accurate presentation of such information;"
- "the Board shall develop assessment objectives consistent with the requirements of this section and test specifications that produce an assessment that is valid and reliable, and are based on relevant widely accepted professional standards;"
- "the Board shall have final authority on the appropriateness of all assessment items;"

- "the Board shall take steps to ensure that all items selected for use in the National Assessment are free from racial, cultural, gender, or regional bias and are secular, neutral, and non-ideological;" and
- "the Board shall develop a process for review of the assessment which includes the active participation of teachers, curriculum specialists, local school administrators, parents, and concerned members of the public."

Given the importance of these mandates, it is incumbent upon the Board to ensure that the highest standards of test fairness and technical quality are employed in the design, construction, and final approval of all test questions for the National Assessment. The validity of educational inferences made using NAEP data could be seriously impaired without high standards and rigorous procedures for test item development, review, and selection.

Test questions used in the National Assessment must yield assessment data that are both valid and reliable in order to be appropriate. Consequently, technical acceptability is a necessary, but not a sufficient condition, for judging the appropriateness of items. In addition, the process for item development must be thorough and accurate, with sufficient reviews and checkpoints to ensure that accuracy. The Guiding Principles, Policies, and Procedures governing item development, if fully implemented throughout the development cycle, will result in items that are fair and of the highest technical quality, and which will yield valid and reliable assessment data.

Each of the following Guiding Principles is accompanied by Policies and Procedures. Full implementation of this policy will require supporting documentation from the National Center for Education Statistics (NCES) regarding all aspects of the Policies and Procedures for which they are responsible.

This policy complies with the documents listed below which express acceptable technical and professional standards for item development and use. These standards reflect the current agreement of recognized experts in the field, as well as the policy positions of major professional and technical associations concerned with educational testing.

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

Code of fair testing practices in education. (1988). Washington, DC: Joint Committee on Testing Practices.

National Center for Education Statistics (NCES) Statistical Standards, DRAFT, February 2002.

Guiding Principles – Item Development and Review Policy

Principle 1

NAEP test questions selected for a given content area shall be representative of the content domain to which inferences will be made and shall match the NAEP assessment framework and specifications for a particular assessment.

Principle 2

The achievement level descriptions for basic, proficient, and advanced performance shall be an important consideration in all phases of NAEP development and review.

Principle 3

The Governing Board shall have final authority over all NAEP test questions. This authority includes, but is not limited to, the development of items, establishing the criteria for reviewing items, and the process for review.

Principle 4

The Governing Board shall review all NAEP test questions that are to be administered in conjunction with a pilot test, field test, operational assessment, or special study administered as part of NAEP.

Principle 5

NAEP test questions will be accurate in their presentation and free from error. Scoring criteria will be accurate, clear, and explicit.

Principle 6

All NAEP test questions will be free from racial, cultural, gender, or regional bias, and must be secular, neutral, and non-ideological. NAEP will not evaluate or assess personal or family beliefs, feelings, and attitudes, or publicly disclose personally identifiable information.

Policies and Procedures for Guiding Principles

Principle 1

NAEP test questions selected for a given content area shall be representative of the content domain to which inferences will be made and shall match the NAEP assessment framework and specifications for a particular assessment.

POLICIES AND PROCEDURES

1. Under the direction of the Board, the framework for each assessment will be developed in a manner that defines the content to be assessed, consistent with NAEP's purpose and the context of a large-scale assessment. The framework development process shall result in a rationale for each NAEP assessment, which delineates the scope of the assessment relative to the content domain. The framework will consist of a statement of purpose, assessment objectives, format requirements, and other guidelines for developing the assessment and items.

2. In addition to the framework, the Board shall develop assessment and item specifications to define the: a) content and process dimensions for the assessment; b) distribution of items across content and process dimensions at each grade level; c) stimulus and response attributes (or what the test question provides to students and the format for answering the item); d) types of scoring procedures; e) test administration conditions; and f) other specifications pertaining to the particular subject area assessment.

3. The Board will forward the framework and specifications to NCES, in accordance with an appropriate timeline, so that NCES may carry out its responsibilities for assessment development and administration.

4. In order to ensure that valid inferences can be made from the assessment, it is critical that the pool of test questions measures the construct as defined in the framework. Demonstrating that the items selected for the assessment are representative of the subject matter to which inferences will be made is a major type of validity evidence needed to establish the appropriateness of items.

5. A second type of validity evidence is needed to ensure that NAEP test items match the specific objectives of a given assessment. The items must reflect the objectives, and the item pool must match the percentage distribution for the content and cognitive dimensions at each grade level, as stated in the framework. Minor deviations, if any, from the content domain as defined by the framework will be explained in supporting materials.

6. Supporting material submitted with the NAEP items will provide a description of procedures followed by item writers during development of NAEP test questions. This description will include the expertise, training, and demographic characteristics of the

groups. This supporting material must show that all item writing and review groups have the required expertise and training in the subject matter, bias, fairness, and assessment development.

7. In submitting items for review by the Board, NCES will provide information on the relationship of the specifications and the content/process elements of the pool of NAEP items. This will include procedures used in classifying each item.

8. The item types used in an assessment must match the content requirements as stated in the framework and specifications, to the extent possible. The match between an objective and the item format must be informed by specifications pertaining to the content, knowledge or skill to be measured, cognitive complexity, overall appropriateness, and efficiency of the item type. NAEP assessments shall use a variety of item types as best fit the requirements stated in the framework and specifications.

9. In order to ensure consistency between the framework and specifications documents and the item pools, NCES will ensure that the development contractor engages a minimum of 20% of the membership of the framework project committees in each subject area to serve on the item writing and review groups as the NAEP test questions are being developed. This overlap between the framework development committees and the item developers will provide stability throughout the NAEP development process, and ensure that the framework and specifications approved by the Board have been faithfully executed in developing NAEP test questions.

Principle 2

The achievement level descriptions for basic, proficient, and advanced performance shall be an important consideration in all phases of NAEP development and review.

POLICIES AND PROCEDURES

1. During the framework development process, the project committees shall draft preliminary descriptions of the achievement levels for each grade to be assessed. These preliminary descriptions will define what students should know and be able to do at each grade, in terms of the content and process dimensions of the framework at the basic, proficient, and advanced levels. Subsequent to Board adoption, the final achievement level descriptions shall be an important consideration in all future test item development for a given subject area framework.

2. The achievement level descriptions will be used to ensure a match between the descriptions and the resulting NAEP items. The achievement level descriptions will be examined, and appropriate instruction provided to item writers to ensure that the items represent the stated descriptions, while adhering to the content and process requirements of the framework and specifications. The descriptions will be used to evaluate the test questions to make certain that the pool of questions encompasses the range of content and

process demands specified in the achievement level descriptions, including items within each achievement level interval, and items that scale below basic.

3. As the NAEP item pool is being constructed, additional questions may need to be written for certain content/skill areas if there appear to be any gaps in the pool, relative to the achievement level descriptions.

4. Supporting materials will show the relationship between the achievement levels descriptions and the pool of NAEP test questions.

Principle 3

The Governing Board shall have final authority over all NAEP test questions. This authority includes, but is not limited to, the development of items, establishing the criteria for reviewing items, and the process for review.

POLICIES AND PROCEDURES

1. Under the No Child Left Behind Act, a primary duty of the Governing Board pertains to "All Cognitive and Noncognitive Assessment Items." Specifically, the statute states that, "The Board shall have final authority on the appropriateness of all assessment items." Under the law, the Board is therefore responsible for all NAEP test questions as well as all NAEP background questions administered as part of the assessment.

2. To meet this statutory requirement, the Board's Policy on NAEP Item Development and Review shall be adhered to during all phases of NAEP item writing, reviewing, editing, and assessment construction. The National Center for Education Statistic (NCES), which oversees the operational aspects of NAEP, shall ensure that all internal and external groups involved in NAEP item development activities follow the Guiding Principles, Policies and Procedures as set forth in this Board policy.

3. Final review of all NAEP test questions for bias and appropriateness shall be performed by the Board, after all other review procedures have been completed, and prior to administration of the items to students.

Principle 4

The Governing Board shall review all NAEP test questions that are to be administered in conjunction with a pilot test, field test, operational assessment, or special study administered as part of NAEP.

POLICIES AND PROCEDURES

1. To fulfill its statutory responsibility for NAEP item review, the Board shall receive, in a timely manner and with appropriate documentation, all test questions that

will be administered to students under the auspices of a NAEP assessment. These items include those slated for pilot testing, field testing, and operational administration.

2. The Board shall review all test items developed for special studies, where the purpose of the special study is to investigate alternate item formats or new technologies for possible future inclusion as part of main NAEP, or as part of a special study to augment main NAEP data collection.

3. The Board shall not review items being administered as part of test development activities, such as small-scale, informal try-outs with limited groups of students designed to refine items prior to large-scale pilot, field, or operational assessment.

4. NCES shall submit NAEP items to the Board for review in accordance with a mutually agreeable timeline. Items will be accompanied by appropriate documentation as required in this policy. Such information shall consist of procedures and personnel involved in item development and review, the match between the item pool and the framework content and process dimensions, and other related information.

5. For its first review, the Board will examine all items prior to the pilot test or field test stage. In the case of the NAEP reading assessment, all reading passages will be reviewed by the Board prior to item development. For each reading passage, NCES will provide the source, author, publication date, passage length, rationale for minor editing to the passage (if any), and notation of such editing applied to the original passage. NCES will provide information and explanatory material on passages deleted in its fairness review procedures.

6. For its second review, the Board will examine items following pilot or field testing. The items will be accompanied by statistics obtained during the pilot test or field test stage. These statistics shall be provided in a clear format, with definitions for each item analysis statistic collected. Such statistics shall include, but shall not be limited to: p-values for multiple-choice items, number and percentage of students selecting each option for a multiple-choice item, number and percentage not reaching or omitting the item (for multiple-choice and open-ended), number and percentage of students receiving various score points for open-ended questions, mean score point value for open-ended items, appropriate biserial statistics, and other relevant data.

7. At a third stage, for some assessments, the Board will receive a report from the calibration field test stage, which occurs prior to the operational administration. This "exceptions report" will contain information pertaining to any items that were dropped due to differential item functioning (DIF) analysis for bias, other items to be deleted from the operational assessment and the rationale for this decision, and the final match between the framework distribution and the item pool. If the technology becomes available to perform statistically sound item-level substitutions at this point in the cycle (from the initial field test pool), the Board shall be informed of this process as well.

8. All NAEP test items will be reviewed by the Board in a secure manner via in-person meetings, teleconference or videoconference settings, or on-line via a password-protected Internet site. The Board's Assessment Development Committee shall have primary responsibility for item review and approval. However, the Assessment Development Committee, in consultation with the Board Chair, may involve other NAGB members in the item review process on an *ad hoc* basis. The Board may also submit items to external experts, identified by the Board for their subject area expertise, to assist in various duties related to item review. Such experts will follow strict procedures to maintain item security, including signing a Nondisclosure Agreement.

9. Items that are edited between assessments by NCES and/or its item review committees, for potential use in a subsequent assessment, shall be re-examined by the Board prior to a second round of pilot or field testing.

10. Documentation of the Board's final written decision on editing and deleting NAEP items shall be provided to NCES within 10 business days following completion of Board review at each stage in the process.

Principle 5

NAEP test questions will be accurate in their presentation, and free from error. Scoring criteria will be accurate, clear, and explicit.

POLICIES AND PROCEDURES

1. NCES, through its subject area content experts, trained item writers, and item review panels, will examine each item carefully to ensure its accuracy. All materials taken from published sources must be carefully documented by the item writer. Graphics that accompany test items must be clear, correctly labeled, and include the data source where appropriate. Items will be clear, grammatically correct, succinct, and unambiguous, using language appropriate to the grade level being assessed. Item writers will adhere to the specifications document regarding appropriate and inappropriate stimulus materials, terminology, answer choices or distractors, and other requirements for a given subject area. Items will not contain extraneous or irrelevant information that may differentially distract or disadvantage various subgroups of students from the main task of the item.

2. Scoring criteria will accompany each constructed-response item. Such criteria will be clear, accurate, and explicit. Carefully constructed scoring criteria will ensure valid and reliable use of those criteria to evaluate student responses to maximize the accuracy and efficiency of scoring.

3. Constructed-response scoring criteria will be developed initially by the item writers, refined during item review, and finalized during pilot or field test scoring. During pilot or field test scoring, the scoring guides will be expanded to include examples of

actual student responses to illustrate each score point. Actual student responses will be used as well, to inform scorers of unacceptable answers.

4. Procedures used to train scorers and to conduct scoring of constructed-response items must be provided to the Board, along with information regarding the reliability and validity of such scoring. If the technology becomes available to score student responses electronically, the Board must be informed of the reliability and validity of such scoring protocol, as compared to human scoring.

Principle 6

All NAEP test questions will be free from racial, cultural, gender, or regional bias, and must be secular, neutral, and non-ideological. NAEP will not evaluate or assess personal or family beliefs, feelings, and attitudes, or publicly disclose personally identifiable information.

POLICIES AND PROCEDURES

1. An item is considered biased if it unfairly disadvantages a particular subgroup of students by requiring knowledge of obscure information unrelated to the construct being assessed. A test question or passage is biased if it contains material derisive or derogatory toward a particular group. For example, a geometry item requiring prior knowledge of the specific dimensions of a basketball court would result in lower scores for students unfamiliar with that sport, even if those students know the geometric concept being measured. Use of a regional term for a soft drink in an item context may provide an unfair advantage to students from that area of the country. Also, an item that refers to a low-achieving student as "slow" would be unacceptable.

2. In conducting bias reviews, steps should be taken to rid the item pool of questions that, because of their content or format, either appear biased on their face, or yield biased estimates of performance for certain subpopulations based on gender, race, ethnicity, or regional culture. A statistical finding of differential item functioning (DIF) will result in a review aimed at identifying possible explanations for the finding. However, such an item will not automatically be deleted if it is deemed valid for measuring what was intended, based on the NAEP assessment framework. Items in which clear bias is found will be eliminated. This policy acknowledges that there may be real and substantial differences in performance among subgroups of students. Learning about such differences, so that performance may be improved, is part of the value of the National Assessment.

3. Items shall be secular, neutral, and non-ideological. Neither NAEP nor its questions shall advocate a particular religious belief or political stance. Where appropriate, NAEP questions may deal with religious and political issues in a fair and objective way.

The following definitions shall apply to the review of all NAEP test questions, reading passages, and supplementary materials used in the assessment of various subject areas:

Secular – NAEP questions will not contain language that advocates or opposes any particular religious views or beliefs, nor will items compare one religion unfavorably to another. However, items may contain references to religions, religious symbolism, or members of religious groups where appropriate.

Examples: The following phrases would be acceptable: "shaped like a Christmas tree", "religious tolerance is one of the key aspects of a free society," "Dr. Martin Luther King, Jr. was a Baptist minister," or "Hinduism is the predominant religion in India."

- Neutral and Non-ideological - Items will not advocate for a particular political party or partisan issue, for any specific legislative or electoral result, or for a single perspective on a controversial issue. An item may ask students to explain both sides of a debate, or it may ask them to analyze an issue, or to explain the arguments of proponents or opponents, without requiring students to endorse personally the position they are describing. Item writers should have the flexibility to develop questions that measure important knowledge and skills without requiring both pro and con responses to every item.

  Examples: Students may be asked to compare and contrast positions on states rights, based on excerpts from speeches by X and Y; to analyze the themes of Franklin D. Roosevelt's first and second inaugural addresses; to identify the purpose of the Monroe Doctrine; or to select a position on the issue of suburban growth and cite evidence to support this position. Or, students may be asked to provide arguments either for or against Woodrow Wilson's decision to enter World War I. A NAEP question could ask students to summarize the dissenting opinion in a landmark Supreme Court case.

  The criteria of neutral and non-ideological also pertain to decisions about the pool of test questions in a subject area, taken as a whole. The Board shall review the entire item pool for a subject area to ensure that it is balanced in terms of the perspectives and issues presented.

4. The Board shall review both stimulus materials and test items to ensure adherence to the NAEP statute and the polices in this statement. Stimulus materials include reading passages, articles, documents, graphs, maps, photographs, quotations, and all other information provided to students in a NAEP test question.

5. NAEP questions will not ask a student to reveal personal or family beliefs, feelings, or attitudes, or publicly disclose personally identifiable information.

APPENDIX E

ACKNOWLEDGMENTS

Julia Mitchell and Daniel Conrad, of the American Institutes for Research, prepared this report, with major contributions from the members of the Planning Committee, the Technical Advisory Panel, and the 2006 NAEP Economics Assessment Management Team.